Mathematics: a Minimal Introduction (Updated Version)

Alexandru Buium

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA
E-mail address: buium@math.unm.edu

INTRODUCTION

This aim of this course is to introduce Mathematics from scratch, with no other prerequisites except Language. In order to follow this approach the students will have to pretend they have never been exposed to Mathematics before. What will have been achieved through this exercise is a reevaluation of the basic concepts and methods of Mathematics. We will mostly ignore the historical and motivational aspects of Mathematics and we will concentrate instead on Mathematical Discourse. Before we embark onto this project, however, we consider, in this Introduction, a few remarks on the history and philosophy of Mathematics. After that we will give a rough outline of the course.

Remarks on the history of Mathematics. Throughout its history Mathematics has evolved as a growing body of statements about numbers, figures, and functions; the problems involving these concepts were studied using methods of Algebra, Geometry, and Calculus, respectively. These chapters of Mathematics are closely intertwined and their development was also closely intertwined with that of the sciences. However, the development of Mathematics had, in addition, its own internal logic and motivations; in particular, mathematical discoveries have often anticipated the motivations coming from the sciences. In the 20th century Mathematics eventually came to be seen as part of Logic; and Logic, in its turn, came to be seen as a body of statements about language. So Mathematics, viewed as a theory in the sense of Logic, only needs language as its prerequisite. Nevertheless, a historical perspective on Mathematics greatly enriches its appreciation and motivates new developments. The history of Mathematics and Logic may be roughly summarized as follows.

1. Algebra and Geometry were established fields of inquiry around the 4th century BC in ancient Greece. Due to the earlier discovery (by the Pythagoreans) that integer numbers cannot account for relations among certain geometric quantities (such as the side and the diagonal of the square) Algebra and Geometry were essentially viewed, up until modern times (17th century), as irreducible to each other.

2. Algebra and Geometry were unified (in that Geometry was "reduced" to Algebra) through the work of Descartes and Fermat (17th century) who invented analytic geometry and through subsequent work of Euler, Langrange (18th century), and Plücker, Möbius, Riemann (19th century).

3. Calculus was invented by Leibniz and Newton in the late 17th century. In the beginning it relied on intuitive (and hence imprecise) concepts.

4. Calculus was reduced to Algebra (which was viewed as relying on more precise concepts) through the work of Cauchy, Weierstrass, etc., (19th century).

5. The concepts of Algebra were reduced to the (naive) concept of set through the work of Cantor, Dedekind, Russell (late 19th and early 20th century). Cantor's Set Theory was not a theory in the technical sense of Logic and his concepts were soon discovered to lead to contradictions (as in Russell's paradox).

6. Logic was first developed by Aristotle (4th century BC). It essentially stayed unchanged for the next two thousand years until the work of Boole (19th century) and especially Frege (19th century) and Russell (early 20th century).

7. The concept of set was reduced to Logic first through the work of Frege, Russell, and Whitehead and eventually by Zermelo and Fraenkel (early 20th century) who put forward a system of axioms (the ZFC axioms) for Set Theory which made the latter a theory in the technical sense of Logic; in this theory paradoxes such as Russell's do not directly arise anymore (although it is possible that similar paradoes or other types of paradoxes might occur in ways that have not yet been discovered; if this is the case new foundations of Mathematics will need to be found!). Other systems of axioms (cf. von Neumann, Gödel, etc.) were proposed which we will ignore here.

8. An approximate mirror of Logic itself was constructed within Set Theory in the early 20th century; this mirror is referred to as Mathematical Logic (not to be confused with Logic). The major theorems in this area belong to Gödel, Turing, Tarski, Cohen, Robinson, Matjasevich, etc.

Remarks on the philosophy of Mathematics. Although the above historical account of Mathematics is generally accepted today there are different ways of interpreting this account from the viewpoint of the philosophy of Mathematics and Logic. Roughly speaking philosophy asks questions about existence (ontology), knowledge (epistemology), and value (axiology). We will ignore, in this course, the epistemological and axiological problems related to Mathematics but we will implicitly make a statement about the ontological status of Mathematics. Ontology distinguishes between particulars (or concrete entities) and universals (or abstract entities). Examples of particulars are the (possible) referents of words such as: planet Earth, Caesar, this reading of this Geiger counter, Hamlet, the Holy Grail, etc. Examples of universals are the (possible) referents of words such as: whiteness, dog, set, number, space, time, movement, causality, natural law, justice, the Good, the Beautiful, etc. One of the problems of ontology is the existence of universals. There are roughly three possible viewpoints on universals, already present in the medieval philosophical debates: realism, conceptualism, and nominalism. Realism holds that universals exist and their existence is independent of the mind. Conceptualism holds that universals exist but their existence is in the mind only. And nominalism declares that universals do not exist in any form. Corresponding to realism, conceptualism, and nominalism there are three doctrines in the modern philosophy of Mathematics: Platonism, intuitionism, and formalism, respectively. In this course we adopt a nominalist (hence formalist) point of view on Logic and Mathematics. It is the minimalist position from the ontological point of view so it requires the least amount of background "faith" in the existence of its entities: only words and the rules governing their use are being taken for granted. Since this position requires the 'least amount of existence" it is the "safest" position from a logical standpoint. The theorems of formalist Mathematics are the same as the theorems of Platonist Mathematics; what is different in these two approaches is not the mathematical content but the interpretation of this content. The theorems of intuitionist Mathematics, on the other hand, are different from those of formalist or Platonist Mathematics; arguably, this places intuitionism outside today's mainstream mathematics.

Platonism has been the standard position among working mathematicians of all times and some of the great logicians (e.g., Russell, Gödel) are associated with it. Formalism and intuitionism came late on the scene; they were the result of a critical reevaluation of the foundations of mathematics that took place in the late 19th and early 20th century. Formalism was initiated by Hilbert (who named Kant as a precursor) while intuitionism was initiated by Brouwer (and also invoked the Kantian tradition). There is more than one way to present the formalist view. In particular the presentation in this course differs from the classical viewpoint of Hilbert in a number of key aspects (e.g., Hilbert's formalism has a Platonic "background" while our approach seeks to eliminate any form of realist underlying assumptions). We will not go into explaining the details of these differences here. We will also say nothing more here about intuitionism. Suffices to note that what mainly distinguishes formalism from both Platonism and intuitionism is the desire to eliminate the "human element" from Mathematics: for the formalist, Mathematics must not invoke "intuitions" (instinctive beliefs, to use a phrase of Russell's) but instead it should be checkable by a machine. (Even if human brains are considered machines the above distinction is still meaningful because we ask that "checking" be done by "following explicitly stated rules" and hence should not involve "intuitions" which may be based on "unconscious rules".)

One should note here an important limitation of formalism. To explain it one needs to distinguish between Mathematical Discourse (which here we always refer to as Mathematics and by which we always understand Mathematics viewed as a theory in the sense of Logic) and the Mathematical Practice (by which we understand the invention/discovery and application of Mathematics). The formalist concentrates on Mathematical Discourse but is forced to ignore Mathematical Practice because the latter involves intuition in an essential way. (The word "discovery" applied to Mathematics does not necessarily imply Platonism; it can be viewed as standing for "uncovering" statements via Logic guided by intuition.) Accordingly, this course will entirely ignore Mathematical Practice and will concentrate on the exposition of Mathematical Discourse.

The modern philosophy of Mathematics (with its Platonist, intuitionist, and formalist variants) originates in the work of Frege and Russell who initiated what is referred to as logicism. The logicist position held that Mathematics is simply Logic supplemented by some extra axioms; in this view Mathematics and Logic are expressed in one and the same language. This position underwent substantial changes in subsequent developments of Logic. In particular the formalist approach of this course will postulate a whole variety of languages and a whole variety of relations between such languages (translations, reference, disquotation, etc.) Logic and Mathematics will be formulated in two different languages: Logic will be formulated in (what we shall call) *Metalanguage* while Mathematics will be formulated in (what we shall call) an *object language*. Mathematical Logic being part of Mathematics will be formulated in the same object language as Mathematics. Hence Logic and Mathematical Logic will be formulated in two different languages and their only possible interaction is through an inherently imperfect (and arguably problematic) translation. A consequence of this is that the theorems of Mathematical Logic will not be seen as (directly) referring to Mathematics at all.

We will make no attempt to carefully compare our approach here with other approaches usually encountered in textbooks. Let us just mention, however, that most textbooks on Mathematics or/and Logic seem to implicitly adopt a realist/Platonist position by accepting, in their ontology, the existence of a "background set theory" in which statements about sets are "facts" that are "true" and "eternal." By contrast, in our nominalist/formalist ontology, all we accept as existing are time, languages, and subjects that can operate with languages in time. Languages will be used to construct theories. Mathematics will be viewed, at every moment in time, as an example of a theory. Theories will be viewed as "physical objects": they will be "finite" sequences of sentences written in a given language on a physical medium (paper, computer) and existing at a definite moment in time. At any moment in time one can add new symbols and new sentences to a theory thus obtaining a new theory called an *extension* of the old theory. The formation of sentences and theories is governed by rules belonging to (what we shall call "general" or "pre-mathematical") Logic. (This is not to be confused with Mathematical Logic which we view as part of Mathematics.) The rules referred to above show what additions of symbols and sentences are acceptable but the choice among the acceptable additions is free. This freedom is what makes the building of theories a creative act. In particular we will view Mathematics as changing in time (because we identify it at each point in time with a new theory, extension of an older theory) and the change at each step is viewed as unpredictable.

Plan of the course. In this course the order of exposition of the various topics will necessarily be quite different from the historical order of their development. We begin with Logic (as a prerequisite for Mathematics; this prerequisite could be called *pre-mathematical Logic*). This will occupy the first 7 chapters. The rest of the course will be devoted to Mathematics. We will introduce Set Theory as one example of a theory within Logic and we define Mathematics as being (at every moment in time) identical to (the corresponding extension of) Set Theory. We then show how the concepts of Algebra can be defined within Set Theory and how the concepts of Geometry and Calculus can be defined within Algebra. We end by very briefly explaining how the concepts of Mathematical Logic can be defined within Set Theory. In each of these fields (Logic, Set Theory, Algebra, Geometry, Calculus, Mathematical Logic) we will merely scratch the surface of the subject. In Algebra we stop after proving some of the basic properties of finite sets, integer numbers, matrices, and polynomials, including the Law of Large Numbers, Euler's formula for planar graphs, the Fundamental Theorem of Arithmetic, the Fundamental Theorem of Symmetric Polynomials, the multiplicativity of determinants, the Hamilton-Cayley theorem, the Existence Theorem for primitive roots, the Fundamental Theorem of Galois Theory, the Fundamental Theorem of Algebra (which will however be proved using calculus) and the Quadratic Reciprocity Law. In Geometry we stop after proving some of the basic properties of plane algebraic curves of degree 1,2,3 (lines, conics, and cubics) including the theorems of Pascal, Pappus, and the Three Cubics Theorem and after we explain the main concepts of differential geometry including the Fundamental Theorem of Riemannian Geometry. In Calculus we stop after proving some of the main theorems on differentiation and integration of real and complex functions, including the Fundamental Theorem of Calculus, and the Cauchy Integral Formula. Many of the steps in these proofs will be relegated to the Exercises which, however, are usually provided with generous hints. In Mathematical Logic we will only explain some of the basic definitions in model theory without discussing any theorems in the subject. Throughout the text we will also include discussion of some of the basic mathematics behind classical physics (Galilean or relativistic) and quantum physics. An Appendix is included which deals with some of the philosophical background of Logic.

The course can be used as a two semester introduction to proofs and rigorous Mathematics. The first semester would roughly cover Logic, sets, and the integers (up until, and including, the Chapter on Arithmetic and with the Chapter called "Examples" possibly skipped). This is what is usually referred to as "discrete mathematics." The second semester could cover specific topics in Algebra, Geometry, and Calculus, which mostly deal with "continuous mathematics" (involving the reals, the imaginaries, topological ideas, etc.) Needless to say, the distinction between "discrete" and "continuous" is rather arbitrary and should be taken with a grain of salt. In particular some chapters of Algebra (which might be perceived as having a "discrete" character) were included in the "continuous" part of the course simply because some of the main examples (such as Lie groups, plane curves, etc.) involve "continuous" mathematics. The discrete and the continuous will be seen to strongly interact with each other. This will become especially clear in the Chapter on Categories which will offer, among other things, a unified picture of Mathematics.

Contents

CHAPTER 1

Languages

In this course we will view Mathematics as part of Logic and we will view Logic as an apparatus aimed at organizing Language. We will start therefore by looking at Language. We admit that there is a whole array of languages and a whole array of relations among them. We explain this by analyzing an example. We will introduce here two languages. The first, which we refer to as *English*, is a drastically simplified version of one of many existing natural languages such as Chinese, Japanese, French, German, etc. The second, which we refer to as *Formal*, is one of many possible formalized/artificial languages; Tarski called it the *Standard* language. Natural and formalized languages share an array of structural properties in spite of the fact that their genesis, mode of operation, and function are different; it is their common properties that interest us here. We shall then examine the interconnections between English and Formal. The approach we present here was pioneered by Frege, Russell, Whitehead, Tarski, etc. There are significant differences among their theories and our presentation is a simplified (minimalist) version of the core of these approaches; as a result we will gloss over many of the subtle points. The logical analysis of Language predates, of course, these developments; it goes back to Aristotle and was closely intertwined with the development of metaphysics for two millenia. We shall ignore this aspect of the analysis altogether.

Let us start with a discussion of English.

EXAMPLE 1.1. The English language is the collection L_{Eng} of all English words (plus separators such as parentheses, commas, etc.). We treat words as individual symbols (and ignore the fact that they are made out of letters). Sometimes we admit as symbols certain groups of words. One can use words to create strings of words such as

0) *"for all not Hamlet man if"*

The above string is considered "syntactically incorrect." The sentences in the English language are the strings of symbols that are "syntactically correct" (in a sense to be made precise later). Here are some examples of sentences in this language:

1) *"Hamlet is a man"*
2) *"Polonius killed Hamlet"*
3) *"The killer of Polonius is Hamlet"*
4) *"Hamlet killed Polonius and Hamlet is a man"*
5) *"Hamlet is not a man or Polonius is a killer"*
6) *"If Hamlet killed Polonius then Hamlet is a killer"*
7) *"Hamlet did not kill Polonius"*
8) *"A man killed Polonius"*

9) *"If a man killed another man then the first man is a killer"*
10) *"A man is a killer if and only if that man killed another man"*

In order to separate sentences in English from a surrounding text that refers to English we put them between quotation marks (and sometimes we write them in italics). So quotation marks do not belong to English but rather they lie outside it; they belong to the language that "talks about" English which we call *Metalanguage*, as we shall explain. Checking syntax presupposes a partitioning of L_{Eng} into various categories of words; no word should appear in principle in two different categories, but this requirement is often violated in practice (which may lead to different readings of the same text). Here are the categories:

- variables: "*something, someone, a thing, an entity,* x, y, z*,...*"
- constants: "*Hamlet, Polonius, ...*"
- functions: "*the killer of,...*"
- predicates: "*is a man, is a killer, killed, ...*"
- connectives: "*and, or, not, if...then, if and only if*"
- quantifiers: "*for all, there exists*"
- equality: "*is, equals*"
- separators: parentheses "(,)" and comma ","

The above categories are referred to as *logical categories.* (They are quite different from, although related to, the *grammatical categories* of *nouns, verbs*, etc.

Constants and variables stand for singular terms (objects, concrete entities). Constants are names of specific objects (proper nouns or phrases that stand for proper nouns) while variables are names for non-specific (generic) objects. The article "the" generally indicates a constant; the article "a" generally indicates that a quantifier is implicitly assumed.

Predicates say/affirm something about one or several objects; if they say/affirm something about one, two, three objects, etc., they are unary, binary, ternary, etc. (So roughly unary predicates correspond to intransitive verbs; binary predicates correspond to transitive verbs.) "Killed" is a binary predicate; "is a killer" is a unary predicate. Predicates are sometimes called *relations* or *relational symbols.*

Functions have objects as arguments but do not say/affirm anything about them; all they do is refer to (or name, or specify, or point towards) something that could itself be an object. Functions will sometimes be called *functional symbols.* Again they can be unary, binary, ternary, etc., depending on the number of arguments. "The father of" is a unary functional symbol. "The son of ... and ..." is a binary functional symbol (where the two arguments stand for the mother and the father and we assume for simplicity that any two parents have a unique son.)

Connectives connect/combine sentences into longer sentences; they can be unary (if they are added to one sentence changing it into another sentence, binary if they combine two sentences into one longer sentence, ternary, etc.). The connective "*and*" is called *conjunction.* The connective "*or*" is called *alternation.* The connective "*not*" is called *negation.* The connective "*if...then...*" is called *implication* or *conditional.* The connective "*if and only if*" is called *equivalence* or *biconditional* and is sometimes abbreviated as "*iff.*". One can consider yet another connective called *disjunction* which is "*either ... or...*"; the relation between alternation and disjunction is that "*A or B*" is the same as "*either A or B or both*"; and similarly "*either A or B*" is the same as "*A or B but not both.*"

Quantifiers specify quantity and are always followed by variables. The most common ones are the universal quantifier "for all" and the existential quantifier "there exists."

Separators separate various parts of the text from various other parts.

In order to analyze a sentence using the logical categories above one first looks for the connectives and one splits the sentence into simpler sentences; alternatively sentences may start with quantifiers followed by variables followed by simpler sentences. In any case, once one identifies simpler sentences, one proceeds by identifying, in each of them, the constants, variables, functions, and predicates. The above type of analysis (called *logical analysis*) is quite different from the *grammatical analysis* based on the grammatical categories of *nouns, verbs*, etc.

Example 1.2. A concise way of understanding the logical analysis of English sentences as above is to create another language L_{For} (let us call it *Formal*) consisting of the following symbols:

- variables: "$x, y, z, ...$"
- constants: "$H, \pi, ...$"
- functions: "$\downarrow, ...$"
- predicates: "$m, k, \dagger$"
- connectives: "$\wedge, \vee, \neg, \rightarrow, \leftrightarrow$"
- quantifiers: "$\forall, \exists$"
- equality: "$=$"
- separators: parentheses "$(,)$" and comma ","

Furthermore let us introduce a rule (called *symbol translation*, given by a doctionary) that attaches to each symbol in Formal a symbol (or possibly several symbols) in English as follows:

"x, y, z" are translated as *"something"* (or again by "$x, y, z, ...$")
"H, π" are translated as *"Hamlet, Polonius"*
"$\downarrow$" is translated as *"the killer of,..."*
"$m, k, \dagger$" are translated as *"is a man, is a killer, killed"*
"$\wedge, \vee, \neg, \rightarrow, \leftrightarrow$" are translated as *"and, or, not, if...then, if and only if"*
"$\forall, \exists$" are translated as *"for all, there exists"*
"$=$" is translated as *"is" or "is a equal to"*

Consider the following sentences in Formal:

1') "mH"
2') "$\pi \dagger H$"
3') "$\downarrow(\pi) = H$"
4') "$(H \dagger \pi) \wedge mS$"
5') "$(\neg(mH)) \vee k\pi$"
6') "$(H \dagger \pi) \rightarrow (kH)$"
7') "$\neg(H \dagger \pi)$"
8') "$\exists x(x \dagger \pi)$"
9') "$\forall x((mx \wedge (\exists y(my \wedge \neg(x = y) \wedge (x \dagger y))) \rightarrow kx)$"
10') "$\forall x(kx \leftrightarrow (mx \wedge (\exists y(my \wedge \neg(x = y) \wedge (x \dagger y))))$"

Sometimes one writes "$m(H)$" instead of "mH." Also one can write "$\dagger(\pi, H)$" or "$\dagger\pi H$" instead of "$\pi \dagger H$."

One says that the English sentences 1)-10) above are *translations* (sometimes called *interpretations*) of the Formal sentences 1')-10'). One also says that the Formal sentences 1')-10') are *translations* (sometimes called *formalizations*) of the English sentences 1)-10).

REMARK 1.3. (Formalization). The above formalizations are obtained in two steps. The first step in formalization (called *paraphrasing*) consists in replacing 8), 9), 10) by the following sentences, respectively:

8") "*There exists something such that that something is a man and that something killed Polonius*" or simply "*There exists x such that x is a man and x killed Polonius.*"

9") "*For all x if x is man and there exists a y such that y is a man, y is different from x and x killed y then x is a killer*",

10") "*For all x one has that x is a killer if and only if x is a man and there exists y such that y is a man, x is not y, and x killed y*"

One also says that 8), 9), 10) are *analyzed* as 8'), 9'), 10'), respectively; or that 8'), 9'), 10') are *logical forms* of 8), 9), 10), respectively.

For English paraphrasing is achieved by using a list of syntactic schemes (which are given in the grammar of the English language with an eye on formalization and which we are not going to record here. One such scheme is, for instance, passing from "All A are B" to "For all x if x is A then x is B". Another scheme is passing from "Some A are B" to "There exists x such that x is A and x is B." Another scheme is passing from "No A is B" to "It is not the case that that there exists x such that x is A and x is B." (Schemes of this type were first considered by Russell.)

The second step in formalization is a "word for word translation" (i.e. replacement of symbols one by one; certain permutations of words are allowed as in "If P then Q" translated as "$P \to Q$" where "if" and "then" are replaced by $\to$ which is placed between P and Q.). Since every sentence in English may be paraphrased in several ways every sentence in English can have several translations into Formal.

One could have paraphrased 1) by

1") "*Hamlet belongs to the totality of all men.*"

or even

1"') "*Hamlet belongs to the totality of all things that are men.*"

Then one could treat "*belongs to*" as a binary predicate which could be translated in Formal by "$\in$" and one could treat "*the totality of all (things that are) men*" as a constant which could be translated in Formal by "M." Then the translation of 1") or 1"') into Formal would be

1"") $S \in M$

Paraphrasing a unary predicate (such as "*is a man*") by a binary predicate plus a constant (as in "*belongs to the totality of all (things that are) men*") is, however, dangerous. Indeed, if one performs this replacement without special care one is led to contradictions. For instance if one paraphrases the unary predicate "*does not belong to itself*" by a binary predicate plus a constant (as in *"belongs to the totality of things that do not belong to themselves"*) then one is led to Russell's paradox (as we shall see later). Therefore, it is advisable to paraphrase sentences of the form 1) by sentences of the form 1') and not by sentences of the form 1""). Later, in Set Theory, an axiom (the separation axiom) will be introduced that specifies what paraphrases of the the kind we considered here are allowed.

REMARK 1.4. (The verb "to exist"). The word "*exists*" which has the form of a predicate (because it is a verb) is instead considered (most of the time) as part of a quantifier. Sentences like "*philosophers exist*" and "*philosophers are human*" have a totally different logical structure. Indeed "*philosophers exist*" should be paraphrased as "*there exists something such that that something is a philosopher*" or simply "*there exists x such that x is a philosopher*"; on the other hand "*philosophers are human*" should be paraphrased as "*for all x if x is a philosopher then x is a human.*" The fact that "*exist*" should not be viewed as a predicate was recognized already by Kant, in particular in his criticism of the "ontological argument."

REMARK 1.5. (The verb "to be"). The verb *to be* (as in *"is, are,..."*) can be:
i) part of a predicate (as in "*is a man*") in which case we say "*is*" is a *copula*;
ii) part of equality (as in "*is equal, is the same as*");
iii) part of a quantifier (as is "*there is*", an equivalent translation of $\exists$).

REMARK 1.6. (On Aristotle's substances). Consider the sentences "*The sky is blue*" and "*Blue is a color*". In the first sentence "*The sky*" is a constant and is the subject of the predicate"*is blue*". In the second sentence "*Blue*" is a constant and is the subject of the predicate "*is a color*". Recall now that the definition of *substance* in Aristotle involves a syntactic component asking that a substance be never able to be part of a predicate. (This has been upheld by the scholastics and later by Leibniz.) There is an ontological component of their definition of substance. Ignoring the ontological component Aristotle would say that "*The sky*" is a substance (because it cannot be part of a predicate, as one cannot say that a subject "*is sky*") but "*Blue*" is not a substance (because, although it can sometimes be the subject of a predicate as in "*Blue is a color*," it can also be part of a predicate as in "*The sky is blue*"); similarly "*color*" is not a substance.

REMARK 1.7. (Metalanguage). All of our discussion of English and Formal above is itself expressed in yet another language which needs to be distinguished from English itself and which we shall call *Metalanguage*. We will discuss Metalanguage in detail in the next chapter (where some languages will be declared *object languages* and others will be declared *metalanguages*). The very latter sentence is written in Metalanguage; and indeed the whole course is written in Metalanguage.

REMARK 1.8. (Naming) It is useful to give names to sentences. For instance if we want to give the name P to the English sentence "*Hamlet is a man*" we can write the following sentence in Metalanguage:

$$P \text{ equals } \text{“\textit{Hamlet is a man.}”}$$

So neither P nor the word *equals* nor the quotation marks belong to English; and *"Hamlet is a man"* will be viewed in Metalanguage as one single symbol. One can give various different names to the same sentence. In a similar way one can give names to sentences in Formal by writing a sentence in Metalanguage:

$$Q \text{ equals } \text{“}m(s)\text{.”}$$

Alternatively one can write

$$Q = \text{“}m(S)\text{”}$$

where $=$ is equality in metalanguage.

Remark 1.9. (Definitions) One can add new predicates or constants to a language by, at the same time, recording certain sentences, called *definitions*. This will be addressed carefully in the chapter on Theories. As an example for the introduction of a new predicate in English we can add to English the predicate “*is an astrochicken*” using previously available predicates such as “*is a chicken*” and “*is a space ship*” by recording the following sentence:

Definition. Something is an astrochicken if and only if it is a chicken and also a space ship.

Here are alternative ways to give this definition:

Definition. An astrochicken is something which is a chicken and also a space ship.

Definition. Something is called *astrochicken* if and only if it is a chicken and also a space ship.

Similarly, if in Formal we have a binary predicate $\in$ and two constants c and s then one could introduce a new predicate ϵ into Formal and record the definition:

Definition. $\forall x(\epsilon(x) \leftrightarrow ((x \in c) \wedge (x \in s)))$.

The two definitions are related by translating $\in$, c, s, and ϵ as “*belongs to*,” “*chicken*,” “*space ships*,” and “*is an astrochicken*,” respectively. The word *astrochicken* is taken from a lecture by Freeman Dyson.

In a similar way one can introduce new functions or new constants; we will come back to this in the Chapter on Theories.

The above type of Definition is called definition by *intension*. Classically the *intension* of a predicate in a language is its “dictionary” definition (which always relies on other, assumed to be known, predicates). On the other hand the *extension* of a predicate p is, classically, the collection of all “objects in the world” that can be predicated by p, i.e. of which p is “true.” The set $\{x \in A \mid P(x)\}$ attached to a formula $P(x)$ which will be introduced in the Chapter on Sets corresponds to the idea of “extension of P” (although the requirement that $x \in A$ will be a severe restriction to the concept of extension). Finally there is a third way to fix predicates in a language, namely by *ostention*, i.e. by pointing, e.g. by pointing to an astrochicken and saying, “*this is an astrochicken*”. As the very example we gave shows, ostention leads to a very limited array of definitions; in particular it cannot be used to define *astrochickens* simply because there seems to be none that one can point at. One should add a remark about another term, *intention* (with a t instead of an s): this term is essentially synonymous to *reference* which will be discussed later.

In the above discussion we encountered 2 examples of languages that we described in some detail (English and Formal) and one example of language (Metalanguage) that we kept somehow vague. Later we will introduce other languages and make things more precise. We would like to "define" now languages in general; we cannot do it in the sense of the present Remark because definitions in our sense here require a language to begin with. All we can do is describe in English what the definition of a language would look like. So the remark below is NOT a definition in the sense of our Remark here.

REMARK 1.10. (Description in English of the concept of language) A *first order language* (or simply a *language*) is a collection L of symbols (assumed to be "inscribed" in a "physical" medium such as paper or papyrus or computer or brain and assumed to possibly grow in time) with the following properties. The symbols in L are divided into 8 categories called *logical categories*. They are: variables, constants, functions, predicates, connectives, quantifiers, equality, and separators. Some of these may be missing. Finally we assume that the only allowed separators are parentheses $(,)$ and commas; we especially ban quotation marks "..." from the separators allowed in a language (because we want to use them as parts of constants in Metalanguage). As already mentioned we assume that the list of variables and constants may grow in time indefinitely: we can always add new variables and constants. The requirement that a physical medium be involved makes the collection "finite" in space in a naive sense; there is a technical sense of "finite" in Mathematics which will be explained much later.

Examples of (first order) languages (in the sense adopted here) are: natural languages (such as English, Chinese, French, German), formalized languages (such as Formal and its variants), Metalanguage (which uses variants of natural languages), Argot (which uses combinations of natural languages and formalized languages, see below), sign languages, Braille, etc. More general languages (which go beyond the ones we are describing here) are: higher order languages (in which one can "quantify over predicates"), "programing languages" (used to write computer programs), "deeper/internal" languages (as postulated, for instance, in Chomsky's theory), etc.

Given a language L one can consider a collection L^* of strings of symbols (also inscribed in a physical medium hence "finite" in space and also allowed to grow in time). There is an "obvious" way (which will be explained later) to define a *syntactically correct* string in L^*; such a syntactically correct string will be called a *sentence*. The collection of sentences in L^* (at every moment in time) is denoted by L^s. (We sometimes say "sentence in L" instead of sentence in L^*.) As in the examples above we can give names $P, ...$ to the sentences in L; these names $P, ...$ do NOT belong to the original language. A translation of a language L into another language L' is a rule that attaches to every symbol in L a symbol (or several symbols) in L'; we assume constants are attached to constants, variables to variables, etc. Due to syntactical correctness the above type of replacement (supplemented by some allowed permutations of words) attaches to sentences P in L sentences P' in L'; such a process is called *word for word translation* (or *symbol for symbol translation*) and it is too rigid to be useful. More general concepts of translation need to be considered; they are given by syntactic schemes that depend of the pair of languages L, L'; for instance if L is a natural language and L' is Formal then one first replaces sentences A in L by sentences B in L which are *paraphrases* of A and then one translates B symbol for symbol into sentences B'

into Formal; B' is declared to be a translation of A (and A to be a translation of B'). Paraphrasing is performed following syntactic rules given in the grammar of L. Translations and reference are required to satisfy the following condition. If L and L' are equipped with reference (which is always the case but sometimes ignored) and if P is a sentence in L whose translation in L' is P' then we impose the condition that P and P' have *the same reference*. Two sentences in a language will be said to have the same *meaning* if they have the same translations in every other available language. In other words the meaning of a sentence is the collection of all its available translations in all the given languages. We view meaning as having various degrees of clarity: the more translations available the more definite the meaning. We impose the condition that if a sentence is a paraphrase of another sentence then the two sentences have the same meaning; in the practice of natural languages this condition is only approximately satisfied. Adopting this syntactic concept of meaning circumvents the subtle "problem of meaning" in the philosophy of language; for the purposes of our course this approach will be sufficient.

Remark 1.11. (Syntax/semantics/reference/inference/truth in languages)

Syntax deals with rules of formation of "correct" sentences. We will examine these rules in detail in a subsequent chapter.

Semantics deals with meaning which was commented upon earlier.

Reference (or universe of discourse) is "what sentences are about." For a "realist" words in English may refer to the physical or imaginary worlds (including symbols in languages which are also viewed as physical entities); e.g., the English word "*Shakespeare*" refers to the "*physical man Shakespeare*"; the words "*the word Shakespeare*" refers to the "*physical word Shakespeare*" as written on a piece of paper, say; and the word "*Hamlet*" refers to something in the imaginary world. For an internalist (like, say, Kant but also Russell) the "physical man Shakespeare" should be replaced by the "phenomenal" (as opposed to the "noumenal" Shakespeare) or by the "idea" of Shakespeare (which can be considered as a physical sign in an "internal language"). Metalanguage, on the other hand, refers to other languages such as English or Formal; so the universe of discourse of Metalanguage consists of other languages; such a reference will be called *linguistic reference*. Reference to things other than languages will be called *non-linguistic reference*. Sentences in Formal can be attached a reference once they are translated into English, say; then they have the same reference as their translations.

Inference is a process by which we accept indicative sentences based on other indicative sentences that are already accepted; see the comments below on indicative sentences. There is a whole array of processes that may qualify as inference from belief to mechanical proof.

We could also ask if the sentences 1),...,10), 1'),...,10') are "true" or "false." We will not define *truth/falsehood* for sentences in any of our languages. Indeed a discussion of truth would complicate our analysis beyond what we are ready to undertake; on the other hand dropping the concepts of truth and falsehood will not affect, as we shall see, our ability to develop Mathematics. To see the difficulties in handling the concept of truth let us consider the sentence "The cat is on the mat" from the viewpoint of the most commonly held theory of truth, the correspondence theory. According to the correspondence theory of truth, truth is the correspondence between sentence and fact. (This position essentially goes back to Aristotle.) In our case we have the sentence "The cat is on the mat" which we call

S. But do we have a fact F which could be in a relation of "correspondence" to S? It seems that F cannot be expressed independently of S so the second term, F, of the correspondence reduces to the first, S. This seems to make the correspondence theory of truth circular. (We could try to express F by taking a picture P of the cat lying on the mat but this would then raise the question as to how to describe the situation in the picture P other than by stating S. We could also try to express F by stating a sentence S' in a different language such as French, say; but this would reduce F to S' and we have the same problem as before.) There are many other theories of truth which can (and have been) advanced: the coherence theory of truth, Tarski's semantic theory of truth, deflationist theories, etc. We will ignore this issue in what follows.

REMARK 1.12. (Correspondences between languages) Translations are an example of correspondence between languages. Other examples of correspondences between languages are *linguistic reference* (a text referring to another text) and *disquotation* (dropping quotation marks).

REMARK 1.13. (Fixed number of constants) English and Formal are examples of languages. Incidentally in these languages the list of constants ends (there is a "fixed number" of constants). But it is important to not impose that restriction for languages. If instead of English we consider a variant of English in which we have words involving arbitrary many letters (e.g., words like "man," "superman," "supersuperman," etc.) then we have an example of a language with "any number of constants." There is an easy trick allowing one to reduce the case of an arbitrary number of symbols to the case of a fixed number of symbols; one needs to slightly alter the syntax by introducing one more logical category, an *operator* denoted, say, by $'$; then one can form constants $c', c'', c''', ...$ starting from a constant c; one can form variables $x', x'', x''', ...$ from a variable x; and one can do the same with functions, predicates, etc.; we will not pursue this in what follows.

REMARK 1.14. (Alternative translations) We already gave an example of translation of Formal into English. The translation given there for connectives, quantifiers, and equality is called the *standard translation.* But there are alternative translations as follows.

Alternative translations of $\rightarrow$ into English are "*if ... then ...*," "*implies*," or "*by...it follows that*," or "*since...we get*," etc.

An alternative translations of $\leftrightarrow$ into English are "*is equivalent to*," "*if, and only if.*"

Alternative translations of $\forall$ into English are "*for all*," or "*for every*," or "*for each.*" Sometimes one translates $\forall$ as "for any," although the latter may lead to confusion. E.g., the sentence "if any of the elements of A is even then not all elements of A are odd" needs to be paraphrased as "if there exists x in A such that x is even then it is not the case that for all y in A, y is odd"; so "any" (usually followed by "of" and not preceded by "for") sometimes paraphrases "there exists.". In view of this ambiguity one should avoid using "for any" as a paraphrase of $\forall$. On the other hand one often uses sentences of the form "Let x be any A that has property P. Then x has property Q." This is always paraphrased as "For every x if x is A and x has property P then x has property Q." So "any" (without "for") can also paraphrase "for every."

Alternative translations of $\exists$ into English are "*for some*" or "*there is an/a*." And, as we saw before, there are cases when "*any*" (usually followed by "of" and not preceded by "for") is properly used in the translation of $\exists$, as in the sentence "if any of the elements of A is even then not all elements of A are odd" considered above.

REMARK 1.15. (Other connectives and quantifiers). English has other connectives (such as *"before," "after," "but," "in spite of the fact that,"* etc.). Some of these we will ignore; others will be viewed as interchangeable with others; e.g., *"but"* will be viewed as interchangeable with *"and"* (although the resulting meaning is definitely altered). Also English has other quantifiers (such as *"there exists a unique," "many," "most," "quite a few," "half," "for at least three,"* etc.). In Formal "there exists a unique" is formalized as $\exists!$ and is often used in mathematical texts. For instance the sentence

"*There is a unique man who killed Polonius*"

can be paraphrased as

"*There is a unique x such that x is a man and x killed Polonius*"

and can be formalized as

"$\exists! x(m(x) \wedge (x \dagger C))$."

The quantifier "there exists a unique" can be "eliminated" i.e., one can paraphrase sentences containing it by sentences that only contain "there exists" and "for all." For instance the above sentence can be paraphrased as

"*There exists x such that x is a man and x killed Polonius and for all y if y is a man and y killed Polonius then $y = x$.*"

and formalized as

"$(\exists x(m(x) \wedge (x \dagger C))) \wedge (\forall y((m(y) \wedge (y \dagger C)) \rightarrow (y = x)))$"

REMARK 1.16. (Texts) Let us consider the following types of objects:
1) symbols (e.g., $x, y, B, C, \dagger, k, ..., \in, \downarrow, ..., \wedge, \vee, \neg, \rightarrow, \leftrightarrow, \forall, \exists, =, (,)$);
2) collections of symbols (e.g., the collection of symbols above, denoted by L);
$2'$) strings of symbols (e.g., $\exists x \forall y(x \in y)$);
3) collections of strings of symbols (e.g., L^*, L^s encountered above or theories T to be encountered later);
$3'$) strings of strings of symbols (such as the proofs to be encountered later).

In the above, collections are unordered while strings are ordered. The above types of objects $(1, 2, 2', 3, 3')$ will be referred to as *texts*. Texts should be thought of as concrete (physical) objects, like symbols written on a piece of paper or papyrus, words that can be uttered, images in a book or in our minds, etc. We assume we know what we mean by saying that a symbol belongs to (or is in) a given collection/string of symbols; or that a string of symbols belongs to (or is in) a given collection/string of strings of symbols. We will not need to iterate these concepts. We will also assume we know what we mean by performing some simple operations on such objects like: concatenation of strings, deleting symbols from strings, substituting symbols in strings with other symbols, "pairing" strings with other strings, etc. These will be encountered and explained later. Texts will be crucial in introducing our concepts of Logic. Note that it might look like we are already assuming some kind of Logic when we are dealing with texts; so our introduction to Logic might seem circular. **But actually the "Logic" of texts that we are assuming is much more elementary than the Logic we want to introduce later; so what we are doing is not circular.**

REMARK 1.17. (Indicative/declarative/imperative/interrogative sentences) All sentences considered so far were indicative (they make a statement that, in natural languages, can be qualified as true or false). Natural languages have other types of sentences: declarative (making a pronouncement like: "We define this to be that"), imperative (giving a command like: "Lift this weight!") and interrogative (asking a question such as: "Is the electron in this portion of space-time?"). "Indicative" and "declarative" are sometimes viewed as synonyms but it is useful to distinguish between the two. Indeed, indicative and declarative sentences may have the same form but different ways of being interpreted/translated. E.g. "I pronounce you husband and wife" can be interpreted as declarative (if the sentence is meant to institute a new state of affairs, in this case the state of affairs in which this unmarried couple becomes a married couple) but it can also be interpreted as indicative (if the sentence describes what I am doing at this moment.) Another, simpler example: "I want to go" can be declarative (expressing an intension, in which case it does not make sense to ask if the sentence is true or not) or indicative (when one understands this sentence in a way that allows it to be false as in the case when I do not actually want to go in spite of what I am saying). In principle, from now on, we will only consider indicative sentences in our languages. An exception to this is the language called *Argot*; see below.

REMARK 1.18. (Argot) For a language L (such as Formal) we may introduce a new language called *argotic* L (or simply *Argot*), denoted sometimes by L_{Argot}. Most mathematics books, for instance, are written in such a language. The language L_{Argot} has as symbols all the symbols of English together with all the symbols of a language L, to which one adds other categories of symbols such as:

• *imperative phrases expressing commands*: *"consider," "assume," "let...be," "let us...,"* etc.)

• *declarative phrases expressing intension*: *"we want to show," "we need to show," "we seek a contradiction," etc.*

Examples of sentences in Argot are

1) "*Since* $(s \in w) \rightarrow (\rho(s))$ *it follows that* $\rho(t)$"
2) "*Let* c *be such that* $\rho(c)$."
3) "*We want to show that* $\rho(c)$."
4) "*For all* c *we have* $\rho(c)$."

We will not insist on explaining the syntax of Argot which is rather different from that of both English and Formal. Suffices to note that the symbols in Formal do not appear between quotation marks inside sentences of Argot; loosely speaking the sentences in Argot often appear as obtained from sentences in Metalanguage via disquotation. Also Argot sometimes uses the constants of Formal as its own variables: the translation into Argot of a Formal sentence "*For all* x, $\rho(x)$" (with

x a variable in Formal) could be "*For all c we have $\rho(c)$*" (with c a constant in Formal).

For the next exercises one needs to enrich Formal by new symbols as needed. The translations involved will not be word for word.

EXERCISE 1.19. Find formalizations of the following English sentences:
1) *"I saw a man."*
2) *"There is no hope for those who enter this realm."*
3) *"There is nobody there."*
4) *"There were exactly two people in that garden."*
5) *"The cat is on the mat".*
6) *"A cat is on the mat".*
7) *"If two lines have two points in common then the lines coincide.*
8) *"For every line and every point not belonging to the line there is a unique line passing through the point and not meeting the first line.*

Sentence 2 above is, of course, a paraphrase of a line in Dante. Sentences 7 and 8 are axioms of Euclidean geometry.

EXERCISE 1.20. Find formalizations of the following English sentences:
1) "*The movement of celestial bodies is not produced by angels pushing the bodies in the direction of the movement but by angels pushing the bodies in a direction perpendicular to the movement.*"
2) "*I think therefore I am.*"
3) "*Since existence is a quality and since a being cannot be perfect if it lacks one quality it follows that a perfect being must exist.*"
4) "*Since some things move and everything that moves is moved by a mover and an infinite regress of movers is impossible it follows that there is an unmoved mover.*"

Hints: The word "*but*" should be paraphrased as "*and*"; "*therefore*" should be paraphrased as "*implies*" and hence as "*if...then*"; "*since...it follows*" should be paraphrased, again, as "*implies.*"

The sentence 1 above paraphrases a statement in one of Feynman's lectures on gravity. The sentence 2 is, of course Descartes' "cogito ergo sum." The sentence 3 is a version of the "ontological argument" (considered by Anselm, Descartes, Leibniz, Gödel; cf. Aquinas and Kant for criticism). See 7.3 for more on this. The sentence 4 is a version of the "cosmological argument" (Aquinas).

CHAPTER 2

Metalanguage

In the previous chapter we briefly referred to *linguistic reference* as being a correspondence between two languages in which the first language $\widehat{L}$ "talks about" a second language L as a language (i.e., it "talks about" the syntax, semantics, etc. of L). We also say that $\widehat{L}$ refers to (or has as universe of discourse) the language L. Once we have fixed L and $\widehat{L}$ we shall call L the *object language* and $\widehat{L}$ the *metalanguage*. (The term *metalanguage* was used by Tarski in his theory of truth; but our metalanguage differs from his in certain respects, cf. Remark 2.4 below. Also this kind of correspondence between $\widehat{L}$ and L is reminiscent of Russell's theory of types of which, however, we will say nothing here.)

Metalanguages and object languages are similar structures (they are both languages!) but we shall keep them separate and we shall hold them to different standards, as we shall see below. Sentences in metalanguage are called *metasentences*. If we treat English and Formal as object languages then all our discussion of English and Formal was written in a metalanguage (which is called *Metalanguage*) and hence consists of metasentences. Let's have a closer look at this concept. First some examples.

Example 2.1. Assume we have fixed an object language L such as English or Formal (or several object languages $L, L', ...$). In what follows we introduce a metalanguage $\widehat{L}$. Here are some examples of metasentences in $\widehat{L}$. First some examples of metasentences of the type we already encountered (where the object language L is either English or Formal):

1) x is a variable in the sentence "$\forall x(x \in a)$."
2) P equals *"Hamlet is a man."*

Later we will encounter other examples of metasentences such as:

3) $P(b)$ is obtained from $P(x)$ by replacing x with b.
4) Under the translation of L into L' the translation of P is P'.
5) By ... the sentence $P \vee \neg P$ is a tautology.
6) c is a constant
7) The string of sentences $P, Q, R, ..., U$ is a proof of V.
8) V is a theorem.
9) If $P, ..., T$ is a proof then T is a theorem.
10) If x is a variable then x is a term.

The metasentences 1, 3, 6 are explanations of syntax in L (see later); 2 is a definition (referred to as a notation or naming); 4 is an explanation of semantics (see later); 5 is part of a metaproof; and 7, 8 are claims about inference (see later). 9 and 10 are axioms in metalanguage (metaaxioms).

Here are the symbols in $\widehat{L}$.

First we postulate that **there are no variables and no quantifiers in metalanguage.** (The reason for this will be explained below.) Next we have:

• constants: *"Hamlet," "Hamlet is a man,"* "$\wedge$," "$=$," L, L^*, L^s, ..., $P, Q, R, ...$,
• functions: the variables in, the translation of, the proof of, $\wedge$, $\vee$, $\neg$, $\rightarrow$, $\leftrightarrow$, $\exists x$, $\forall x$,...
• predicates: is translated as, occurs in, is obtained from...by replacing...with..., is a tautology, is a proof,..., follows from, by ... it follows that..., by ... one gets that,...
• connectives: and, or, not, if...then, if and only if, because,...
• equality: is, equals,...
• separators: parentheses, comma, period.

Remark 2.2. Note that names of sentences in the object language become constants in metalanguage. The texts of the object language, appearing between quotation marks, also become constants in metalanguage. The connectives of the object language become functions in metalanguage. The symbols "$\wedge, ...$" used as constants, the symbols $\wedge, ...$ used as functions, and the symbols *and,...* viewed as connectives should be viewed as different symbols (normally one should use different notation for them).

Remark 2.3. The above metalanguage can be viewed as a *MetaEnglish* because it is based on English. One can construct a *MetaFormal* metalanguage by replacing the English words with symbols including:
• connectives: & (for and), $\Rightarrow$ (for if...then), $\Leftrightarrow$ (for if and only if)
• equality: $\equiv$ (for is, equals)

We will not proceed this way, i.e., we will always use MetaEnglish as our metalanguage.

Remark 2.4. What Tarski called *metalanguage* is close to what we call *metalanguage* but not quite the same. The difference is that Tarski allows metalanguage to contain the symbols of original object language written *without* quotation marks. So for him (but not for us), if the language is Formal, then the following is a metasentence:

$$\text{“}\forall x \exists y s(x, y)\text{” is true if and only if } \forall x \exists y s(x, y)$$

Allowing the above to be a metasentence helped Tarski introduce his conditions that *truth in a language* should satisfy (the Tarski T scheme); we will not discuss this here but see the Chapter on Models.

Remark 2.5. (Syntax/semantics/reference/inference/truth in metalanguage versus object language)

The syntax of object languages will be regulated by metalanguage. On the other hand metalanguage has a syntax of its own which is simpler (e.g. there are no variables and quantifiers) and we keep less precise than that of object languages so that we avoid the necessity of introducing a metametalanguage which regulates it; that would prompt introducing a metametametalanguage that regulates the metametalanguage, etc. The hope is that metalanguage, kept sufficiently loose from a syntactic viewpoint, can sufficiently well explain its own syntax without leading to contradictions. The very text you are reading now is, in effect, metalanguage explaining its own syntactic problems. The syntactically correct texts in metalanguage are referred to as metasentences. Definitions in metalanguage are called *metadefinitions*. It is crucial to distinguish between words in sentences and words in metasentences which are outside the quotation marks; even if they look the same they should be regarded as different words.

In terms of semantics sentences in object languages are assumed to have a meaning derived from (or rather identified with) translations into other languages but we will generally ignore this meaning. On the other hand, metasentences have a metameaning derived from their own translations into other languages; we shall assume we understand their metameaning and we shall *not* ignore it.

Metasentences have a reference (which could be called *metareference*): they always refer to sentences in the object language. On the other hand we will generally ignore the reference of sentences in object language.

The "double standard" approach towards object languages and metalanguage is necessary if one wants to avoid the introduction of metametalanguages, etc.; it is reasonable because the metameaning and metareference of metasentences in metalanguage is much simpler than the meaning and reference of sentences in object language. Referring to written words as mere words is a much more straightforward business than referring to what the words refer to.

Metasentences, as well as sentences in object language, are assumed to have no truth value (it does not make sense to say they are true or false).

For instance the metasentences

a. The word "*elephants*" occurs in the sentence "*elephants are blue*."
b. The word "*crocodiles*" occurs in the sentence "*elephants are gray*."

can be translated into the "metalanguage of letter searches" (describing how to search a word in a sentence, say). Both metasentences have a metameaning. Intuitively we are tempted to say that (a) is true and (b) is false. As already mentioned we do not want to introduce the concepts of *true* and *false* in this course. Instead we infer sentences in object language, respectively, metasentences; inference of sentences in object languages will be called *proof* or *deduction* and will be "mechanical" (will not involve semantics); inference of metasentences in metalanguage will be called *metaproof* or *showing* or *checking* and will involve a certain degree of semantics (for instance of the words "if" and "then" but not "for all" and "there exists").

For example we agree that (a) above can be metaproved; also the negation of (b) can be metaproved (checked by showing). Metaproof is usually based on a translation into a "computer language": for instance to metaprove (a) take the words of the sentence "*elephants are blue*" one by one starting from the right (say) and ask if the word is "*elephants*"; the third time you ask the answer is yes, which ends the metaproof of (a). A similar discussion applies to some other types of metasentences; e.g., to the metasentences 1-7 in Example 2.1. The metaproof of 5 in Example 2.1 involves, for instance, "showing tables" whose correctness can be checked by inspection by a machine. (This will be explained later.) The situation with the metasentence 8 in Example 2.1 is quite different: there is no method (program) that can decide if there is a metaproof for 8; neither is there a method that can find a metaproof for 8 in case there is one. But if one already has a proof of T in 8 then checking that the alleged proof is a proof can be done mechanically and this provides a metaproof for 8. Most metaproofs consist in checking that a definition applies to a given metasentence. The rules governing the latter would be spelled out in metametalanguage; we will not do this here.

The different standards adopted for object language and metalanguage are based on the following "balance" principle best expressed in a table as follows:

	object language	metalanguage
quantification	present	absent
syntactic structure	strong	weak
semantic structure	weak	strong
ability to infer	strong	weak
ability to refer	weak	strong
truth	undefined	undefined

REMARK 2.6. Since P, Q are constants in metalanguage and $\wedge, \vee, ..., \exists x$ are functions in metalanguage one can form syntactically correct strings

$$P \wedge Q, ..., \exists x P$$

in metalanguage, etc. If

P equals "$p...$"
Q equals "$q...$"

where $p, ..., q, ...$ are symbols in the object language then we introduce the rule ("metaaxiom")

$P \wedge Q$ equals "$(p...) \wedge (q...)$"

The above should be viewed as one of the rules allowed in metaproofs. Similar obvious rules can be given for $\vee, \exists$, etc. Note that the parentheses are many times necessary; indeed without them the string $P \vee Q \wedge R$ would be ambiguous. We will drop parentheses, however, each time there is no ambiguity. For instance we will never write $(P \vee Q) \wedge R$ as $P \vee Q \wedge R$. Note that according to these conventions $(P \vee Q) \vee R$ and $P \vee (Q \vee R)$ are still considered distinct.

REMARK 2.7. Assume we are given a metadefinition:

P equals "$p....$"

Then we say P is a *name* for the sentence "$p....$" We impose the following rule for this type of metadefinition: if two sentences have the same name they are identical (as strings of symbols in the object language; identity means exactly the same symbols in the same order and it is a physical concept). Note on the other hand that the same sentence in the object language can have different names.

In the same spirit if

$P(x)$ equals "$p...x...$"

is a metadefinition in metalanguage then we will add to the object language a new predicate (still denoted by P) by adding, to the definitions of the object language, the following definition:

$\forall x(P(x) \leftrightarrow (p...x...))$.

So the symbol P appears once as a constant in metalanguage and as a predicate in the object language. (We could have used two different letters instead of just P but it is more suggestive to let P play two roles.) This creates what one can call a *correspondence* between part of the metalanguage and part of the language. This correspondence, which we refer to as *linguistic reference*, is not like a translation between languages because constants in metalanguage do not correspond to constants in the object language but to sentences (or to new predicates) in the object language. In some sense this linguistic reference is a "vertical" correspondence between languages of "different scope" whereas translation is a "horizontal" correspondence between languages of "equal scope." The words "vertical" and "horizontal" should be taken as metaphors rather than precise terms.

REMARK 2.8. (Disquotation) There is a "vertical" correspondence (called *disquotation* or *deleting quotation marks*) that attaches to certain metasentences in metalanguage a sentence in the object language (or Argot). Consider for instance the metasentence in MetaEnglish

1) From "*Hamlet is a man*" and "*all men are mortal*" it follows that "*Hamlet is mortal.*"

Its disquotation is the following sentence in English:

2) *Since Hamlet is a man and all men are mortal it follows that Hamlet is mortal.*

Note that 1 refers to some sentences in English whereas 2 refers to something (somebody) called Hamlet. So the references of 1 and 2 are different; and so are their meaning (if we choose to care about the meaning of 2 which we usually don't).

If P equals "*Hamlet is a man*," Q equals "*all men are mortal*," and R equals "*Hamlet is mortal*" then 2 above is also viewed as the disquotation of:

1') From P and Q it follows that R.

Disquotation is not always performable: if one tries to apply disquotation to the metasentence

1) x is a free variable in "*for all x, x is an elephant*"

one obtains

2) *x is a free variable in for all x, x is an elephant*

which is not syntactically correct.

Disquotation is a concept involved in some of the classical theories of truth, e.g., in Tarski's example:

"Snow is white" is true if and only if snow is white.

Since we are not concerned with truth in this course we will not discuss this connection further. For more on this see the Chapter on Models.

We will often apply disquotation without any warning if there is no danger of confusion.

EXERCISE 2.9. Consider the following sentences in English and explain how they are obtained from metasentences in Metalanguage.

1) To be or not to be, that is the question.

2) You say yes, I say no, you say stop, but I say go, go, go.

3) The sentence you are reading is false.

1 is, of course, from Shakespeare. 2 is from the Beatles. 3 is a form of *liar's paradox*.

Hint: 1 and 2 are obtained by disquotation (plus paraphrasing). 3 is obtained by giving a name to 3 and referring to 3 by name. There are many ways to do this.

REMARK 2.10. (Indicative/declarative/imperative/interrogative) All metasentences considered so far were indicative (they state their content). There are other types of metasentences: declarative ("We define this to be..."), imperative (giving a command like: "Prove this theorem!," "Replace x by b in P," "Search for x in P," etc.) and interrogative (asking a question such as: "What is the value of the function f at 5?," "Does x occur in P?," etc.). The syntax of metasentences discussed above only applies to indicative metasentences. We will only use imperative/interrogative metasentences in the exercises or some metaproofs; these other types of metasentences require additional syntactic rules which are clear and we will not make explicit here.

From now on we will make the following convention. In any discussion about languages we will assume we have fixed an object language and a metalanguage. The object language will simply be referred to as the "language." So the word "object" will systematically be dropped.

CHAPTER 3

Syntax

We already superficially mentioned syntax. In this chapter we discuss, in some detail, the syntax of Formal (or similar languages). The syntax of English (or other natural languages), Metalanguage, and Argot are similar but more complicated and will not be explicitly addressed here. All the explanations below are, of course, written in Metalanguage.

As we saw a language is a collection L of symbols. Given L we considered the collection L^* of all strings of symbols in L. In this chapter we explain the definition of sentences (which will be certain special strings in L^*). Being a sentence will involve, in particular, a certain concept of "syntactic correctness." The kind of syntactic correctness discussed below makes L a *first order language*. There are other types of languages whose syntax is different (e.g., second order languages, in which one is allowed to say, for instance, "for every predicate, etc...."; or languages whose syntax is based on grammatical categories rather than logical categories; or computer languages, not discussed in this course at all). First order languages are the most natural (and are entirely sufficient) for developing Mathematics.

In what follows we let L be a collection of symbols consisting of variables $x, y, ...$, constants, functions, predicates, connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$, $\leftrightarrow$ (where $\neg$ is unary and the rest are binary), quantifiers $\forall, \exists$, equality $=$, and, as separators, parentheses $(,)$, and commas. (For simplicity we considered 5 "standard" connectives, 2 "standard" quantifiers, and a "standard" symbol for equality; this is because most examples will be like that. However any number of connectives and quantifiers, and any symbols for them would do. In particular some of these categories of symbols may be missing.) According to our conventions recall that we will also fix a metalanguage $\widehat{L}$ in which we can "talk about" L.

EXAMPLE 3.1. If L has constants $a, b, ...$, a functional symbol f, and a binary predicate $\Box$ then here are examples of strings in L^*:

1) $)(x\forall\Box\exists aby(\rightarrow$
2) $f(f(a))$
3) $\exists y((a\Box x) \rightarrow (a\Box y))$
4) $\forall x(\exists y((a\Box x) \rightarrow (a\Box y)))$

In what follows we will define terms, formulas, and sentences; 1 above will be neither a term, nor a formula, nor a sentence; 2 will be a term; 3 will be a formula but not a sentence; 4 will be a sentence.

For the rule (metaaxiom) below we introduce a unary predicate "*is a term*" into metalanguage.

METAAXIOM 3.2.
1) If x is a variable then x is a term. If c is a constant then c is a term.

2) If $t, s, ...$ are terms and f is a functional symbol then $f(t, s, ...)$ is a term.

Remark 3.3. Note that we did not say "all variables are terms" because we cannot use quantifiers in metalanguage; instead we have a metaaxiom as above for each individual variable and each individual constant. This observation will apply again and again when discussing metaaxioms.

Remark 3.4. Functions may be unary $f(t)$, binary $f(t, s)$, ternary $f(t, s, u)$, etc. When we write $f(t, s)$ we simply mean a string of 5 symbols; there is no "substitution" involved here. Substitution will play a role later, though; cf. 3.14.

Example 3.5. If $a, b, ...$ are constants, $x, y, ...$ are variables, f is a unary functional symbol, and g is a a binary functional symbol, all of them in L, then

$$f(g(f(b), g(x, g(x, y))))$$

is a term. The text above is a metasentence. A metaproof of this is as follows. Since x, y are terms $g(x, y)$ is a term. Hence $g(x, g(x, y))$ is a term. Since b is a term $f(b)$ is a term. So $g(f(b), g(x, g(x, y)))$ is a term, hence $f(g(f(b), g(x, g(x, y))))$ is a term.

For the next metaaxiom we introduce the predicate *is a formula* into metalanguage.

Metaaxiom 3.6.
1) If t, s are terms then $t = s$ is a formula.
2) If $t, s, ...$ are terms and ρ is a predicate then $\rho(t, s, ...)$ is a formula.
3) If Q, Q' are formulae then $Q \wedge Q'$, $Q \vee Q'$, $\neg Q$, $Q \rightarrow Q'$, $Q \leftrightarrow Q'$ are formulae.

Formulae of the form 1) or 2) above are called *atomic formulae*.

Recall our convention that if we have a different number of symbols (written differently) we make similar metadefinitions for them; in particular some of the symbols may be missing altogether. For instance if equality is missing from L we ignore 1.

We introduce a predicate "is in L^f" into metalanguage and we introduce the metaaxioms: P is a formula if and only if P is in L^f.

Remark 3.7. Predicates can be unary $\rho(t)$, binary $\rho(t, s)$, ternary $\rho(t, s, u)$, etc. Again, $\rho(t, s)$ simply means a string of 5 symbols $\rho, (, t, s,)$ and nothing else. Sometimes one uses another syntax for predicates: instead of $\rho(t, s)$ one writes $t\rho s$ or ρts; instead of $\rho(t, s, u)$ one may write ρtsu, etc. All of this is in the language L. On the other hand if some variables $x, y, ...$ appear in a formula P we sometimes write in metalanguage $P(x, y, ...)$ instead of P. In particular if x appears in P (there may be other variables in P as well) we sometimes write $P(x)$ instead of P. Formulas of the form $P \rightarrow Q$ are referred to as conditional formulas. Formulas of the form $P \leftrightarrow Q$ are referred to as biconditional formulas.

Example 3.8. Assume L contains a constant c, a unary predicate ρ, and a unary functional symbol f. Then the following is a formula:
$(\forall x(f(x) = c)) \rightarrow (\rho(f(x)))$

For what follows we need to add a predicate "x *is free in*". If x is free in P we also say x is a free variable in P. Instead of "x is not free in P" we sometimes say "x is bound in P".

METAAXIOM 3.9.

1) If x appears in the term t or in the term s then x is free in the formula $t = s$. If x does not appear in s and does not appear in t then x is bound in $t = s$.

2) If x appears in the term t or in the term s, etc., and if ρ is a predicate then x is free in the formula $\rho(t, s, ...)$. If x does not appear in the terms $t, s, ...$ and if ρ is a predicate then x is bound in the formula $\rho(t, s, ...)$.

3) If x is free in the formula P or in the formula Q then x is free in $P \wedge Q$, $P \vee Q$, $P \rightarrow Q$, $P \leftrightarrow Q$, $\neg P$. If x is bound in the formula P and in the formula Q then x is bound in $P \wedge Q$, $P \vee Q$, $P \rightarrow Q$, $P \leftrightarrow Q$, $\neg P$.

4) If P is a formula and x is free in P then $\exists x P$ and $\forall x P$ are formulas and x is bound in $\exists x P$ and in $\forall x P$.

5) If P is a formula and x and y are free in P then y is free in $\exists x P$ and in $\forall x P$. If P is a formula and y is bound in P and x is free in P then y is bound in $\exists x P$ and in $\forall x P$.

Formulas of the form (i.e., which are equal to one of) $\forall x P$, $\forall x P(x)$ are referred to as universal formulas. Formulas of the form $\exists x P$, $\exists x P(x)$ are referred to as existential formulas.

EXAMPLE 3.10.

1) x is bound in $\forall y \exists x(\rho(x, y))$. (Metaproof of this: x is bound in $\exists x(\rho(x, y))$ hence bound in $\forall y \exists x(\rho(x, y))$.)

2) x is free in $(\exists x(\beta(x))) \vee \rho(x, a)$. (Metaproof of this: x is free in $\rho(x, a)$ so x is free in $(\exists x(\beta(x))) \vee \rho(x, a)$.)

3) x is bound in $\forall x((\exists x(\beta(x)) \vee \rho(x, a))$.

4) y and z are free in $(\forall x \exists y(\alpha(x, y, z))) \wedge \forall u(\beta(u, y))$.

EXAMPLE 3.11. Let p and q be a ternary and a binary predicate respectively. The following is a formula $(\forall y(\exists x\ q(x, y)))) \rightarrow (\neg p(x, y, z))$. The above text is a metasentence. A metaproof of it is as follows. Since x, y, z are terms $q(x, y)$ and $p(x, y, z)$ are formulas. Hence $\neg p(x, y, z)$ is a formula. Since x is free in $q(x, y)$ we have that $\exists x q(x, y)$ is a formula. Since y is free the latter we get that $\forall y(\exists x\ q(x, y)))$ is a formula. Hence $(\forall y(\exists x\ q(x, y)))) \rightarrow (\neg\ p(x, y, z))$ is a formula.

EXAMPLE 3.12. Let p and q be a ternary and a binary predicate respectively. Consider the string of symbols $(\forall y(\exists x\ q(x, y)))) \rightarrow \vee(\neg\ p(x, y, z))$. One cannot metaprove the metasentence "the above is a formula" the way we proceeded in the previous example (the problem being with the symbol $\vee$). This does not mean that we can metaprove the sentence "the above string is not a formula"; all we have is that we cannot conclude that the above string is a formula.

METADEFINITION 3.13. A string in L^* is called a *sentence* if it is a formula (i.e., is in L^f) and has no free variables. Sentences are sometimes called *closed formulae*. Formulae that are not closed are called *open*.

To avoid the implicit quantifier in the above formulation one actually needs to introduce predicates of the form "is a formula with variables x, y, z" (etc.) and replace "no free variables" with "x is bound, y is bound, and z is bound", etc.

Note that

1) If P is in L^s then P is in L^f;
2) If P is in L^f then P is in L^*;
3) If t is a term then t is in L^* and not in L^f.

METADEFINITION 3.14. If x is a free variable in a formula P one can replace all its free occurrences with a term t to get a formula which can be denoted by $P\frac{t}{x}$. More generally if $x, y, ...$ are variables and $t, s, ...$ are terms, we may replace all free occurrences of these variables by $t, s, ...$ to get a formula $P\frac{ts...}{xy...}$. A more suggestive (but less precise) notation is as follows. We write $P(x)$ instead of P and then we write $P(t)$ instead of $P\frac{t}{x}$. Similarly we write $P(t, s, ...)$ instead of $P\frac{ts...}{xy...}$ We will constantly use this $P(t), P(t, s, ...)$, etc. notation from now on.

Similarly if u is a term containing x and t is another term then one may replace all occurrences of x in u by t to get a term which we may denote by $u\frac{t}{x}$; if we write $u(x)$ instead of u then we can write $u(t)$ instead of $u\frac{t}{x}$. And similarly we may replace two variables x, y in a term u by two terms t, s to get a term $u\frac{ts}{xy}$, etc. We will not make use of this latter type of substitution in what follows.

EXAMPLE 3.15. If P equals "*x is a man*" then x is a free variable in P. If a equals "*Hamlet*" then $P(a)$ equals "*Hamlet is a man.*"

EXAMPLE 3.16. If P equals "*x is a man and for all x, x is mortal*" then x is a free variable in P. If a equals "*Hamlet*" then $P(a)$ equals "*Hamlet is a man and for all x, x is mortal.*"

EXERCISE 3.17. Is x a free variable in the following formulas?
1) "$(\forall y \exists x(x^2 = y^3)) \wedge (x \text{ is a man})$"
2) "$\forall y(x^2 = y^3)$"
Here the upper indexes 2 and 3 are unary functions.

EXERCISE 3.18. Compute $P(t)$ if:
1) $P(x)$ equals "$\exists y(y^2 = x)$" and "t" equals "x^4."
2) $P(x)$ equals "$\exists y(y \text{ poisoned } x)$" and "$t$" equals "*Plato's teacher.*"

CHAPTER 4

Tautologies

We start now the analysis of inference within a given language (which is also referred to as deduction or proof). In order to introduce the general notion of proof we need to first introduce tautologies; in their turn tautologies are introduced via certain arrays of symbols in metalanguage called *tables*.

METADEFINITION 4.1. Let T and F be two constants in metalanguage. We also allow separators in metalanguage that are frames of tables. Using the above plus arbitrary constants P and Q in metalanguage we introduce the following strings of symbols in metalanguage (which are actually arrays rather than strings but which can obviously be rearranged in the form of strings). They are referred to as the truth tables of the 5 standard connectives.

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	Q	$P \rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	Q	$P \leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

P	$\neg P$
T	F
F	T

REMARK 4.2. If in the tables above P is the sentence "p...." and Q is the sentence "q...." we allow ourselves, as usual, to identify the symbols $P, Q, P \wedge Q$, etc. with the corresponding sentences "p...," "q...," "(p...)$\wedge$(q...)," etc. Also: the letters T and F evoke "truth" and "falsehood"; but they should be viewed as devoid of any meaning.

Fix in what follows a language L that has the 5 standard connectives $\wedge$, $\vee$, $\neg$, $\rightarrow$, $\leftrightarrow$ (but does not necessarily have quantifiers or equality).

We introduce a new predicate in metalanguage: "is a string" and "is a Boolean string generated by P, Q, R". (Similar predicates "is a Boolean string generated by P, Q", "is a Boolean string generated by P, etc.) We also introduce a new predicates "the formula P belongs to the string S" and "the string S' is obtained from the string S by adding the formula P".

METAAXIOM 4.3.
1) The string
P,
Q,
R

is a Boolean string generated by P, Q, R.

2) If S is a Boolean string generated by P, Q, R and if V', V'' are formulae belonging to S then the string obtained from S by adding any of the formulae

$$V' \wedge V'', V' \vee V'', \neg V', V' \rightarrow V'', V' \leftrightarrow V''.$$

is a Boolean string generated by P, Q, R.

EXAMPLE 4.4. The following is a Boolean string generated by P, Q, R:
P
Q
R
$\neg R$
$Q \vee \neg R$
$P \rightarrow (Q \vee \neg R)$
$P \wedge R$
$(P \wedge R) \leftrightarrow (P \rightarrow (Q \vee \neg R))$

A metaproof of that is as follows:
P,
Q,
R
is a Boolean string generated by P, Q, R. Then
P,
Q,
R,
$\neg R$
is a Boolean string generated by P, Q, R. Then
P,
Q,
R,
$\neg R$,
$Q \vee \neg R$
is a Boolean string generated by P, Q, R. Etc. Note that this metaproof involves the semantics of "if" and "then".

EXAMPLE 4.5. Consider the string
P
Q
R
$T \rightarrow R$
$\neg R$
$Q \vee \neg R$
$P \rightarrow (Q \vee \neg R)$
$P \wedge R$
$(P \wedge R) \leftrightarrow (P \rightarrow (Q \vee \neg R))$

The metasentence "The above is a Boolean string generated by P, Q, R" cannot be metaproved as in the previous example (the problem being that the presence of $T \rightarrow R$ cannot be justified). This does not metaprove the metasentence "the above is not a Boolean string generated by P, Q, R"; all we have is that we cannot conclude that the above string is a Boolean string generated by P, Q, R.

EXAMPLE 4.6. The following is a Boolean string generated by $P \to (Q \vee \neg R)$ and $P \wedge R$:

$P \to (Q \vee \neg R)$
$P \wedge R$
$(P \wedge R) \leftrightarrow (P \to (Q \vee \neg R))$

REMARK 4.7. The same sentence may appear as the last sentence in two different Boolean strings; cf. the previous examples.

REMARK 4.8. Assume we are given a Boolean string generated by P, Q, R. (When more or less than 3 generators the metadefinition is similar.) One is tempted to give the following metadefinition. The truth table attached to this Boolean string and to the fixed system of generators P, Q, R is a string of symbols (or rather plane configuration of symbols thought of as reduced to a string of symbols) as follows:

P	Q	R	U	...	W
T	T	T	...	...	...
T	T	F	...	...	...
F	T	T	...	...	...
F	T	F	...	...	...
T	F	T	...	...	...
T	F	F	...	...	...
F	F	T	...	...	...
F	F	F	...	...	...

Here note that the 3 columns of the generators consist of all 8 possible combinations of T and F. The dotted columns correspond to the sentences other than the generators and are computed by the following rule. Assume V is not one of the generators P, Q, R and assume that all columns to the left of the column of V were computed; also assume that V is obtained from V' and V'' via some connective $\wedge, \vee, ...$ Then the column of V is obtained from the columns of V' and V'' using the tables of the corresponding connective $\wedge, \vee, ...,$ respectively. The above metadefinition is, however, not correct because it implicitly involves quantifiers. So one has to reject this as a metadefiniton and one proceeds, instead, as follows.

METAAXIOM 4.9. One introduces predicates "is a table," "is a truth table," and "is obtained from ... by adding a column by the rule..." into metalanguage and one introduces the metaaxiom "if $\mathcal{T}$ is a truth table and $\mathcal{T}'$ is table obtained from $\mathcal{T}$ by a adding a column by the rule above then $\mathcal{T}'$ is a truth table." We leave the (messy but obvious) details to the reader.

REMARK 4.10. Note that there is a "mechanical procedure" to decide (and check) if a given sentence is a tautology or not.

EXAMPLE 4.11. Consider the following Boolean string generated by P and Q:

P
Q
$\neg P$
$\neg P \wedge Q$

Its truth table is:

P	Q	$\neg P$	$\neg P \wedge Q$
T	T	F	F
T	F	F	F
F	T	T	T
F	F	T	F

Note that the generators P and Q are morally considered "independent" (in the sense that all 4 possible combinations of T and F are being considered for them); this is in spite of the fact that actually P and Q may be equal, for instance, to $\forall x(p(x))$ and $\exists x(\neg p(x))$, respectively.

We introduce the predicate "the sentence ... is a tautology" into metalanguage and we consider the following:

Metaaxiom 4.12. If S is a Boolean string generated by sentences P, Q, R such that the last sentence in the string is V and the truth table attached to the string and the generators P, Q, R has only Ts in the V column then V is a tautology. Same metaaxiom for more or less than 3 sentences.

Remark 4.13. Note that we did not give above a metadefinition of tautology by asking that there exist a Boolean string ending in S with the S column containing only Ts; such a metadefinition would be not correct because it contains quantifiers. In particular, although there is direct way to metaprove that a given sentence is a tautology there is no direct way to metaprove that a given sentence is not a tautology. This can be done in Lathematical Logic but not in the framework of general Logic developed here.

In all the exercises and examples below, $P, Q, ...$ are names of sentences.

Example 4.14. $P \vee \neg P$ is a tautology. To metaprove this consider the Boolean string generated by P,

P
$\neg P$
$P \vee \neg P$

Its truth table is (check!):

P	$\neg P$	$P \vee \neg P$
T	F	T
F	T	T

This ends our metaproof of the metasentence saying that $P \vee \neg P$ is a tautology.

Remark that if we view the same Boolean string

P
$\neg P$
$P \vee \neg P$

as a Boolean string generated by P and $\neg P$ the corresponding truth table is

P	$\neg P$	$P \vee \neg P$
T	T	T
T	F	T
F	T	T
F	F	F

and the last column in the latter table does not consist of Ts only. This does not change the fact that $P \vee \neg P$ is a tautology. Morally, in this latter computation we had to treat P and $\neg P$ as "independent"; this is not a mistake but rather a failed attempt to metaprove that $P \vee \neg P$ is a tautology.

EXAMPLE 4.15. Let S be the sentence $(P \wedge (P \to Q)) \to Q$. This sentence is a tautology which is called *modus ponens*. To metaprove this consider the following Boolean string generated by P, Q:

P
Q
$P \to Q$
$P \wedge (P \to Q)$
$(P \wedge (P \to Q)) \to Q$

Its truth table is:

P	Q	$P \to Q$	$P \wedge (P \to Q)$	S
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

EXERCISE 4.16. Explain how the table above was computed.

EXERCISE 4.17. Give a metaproof of the fact that each of the sentences below is a tautology:
1) $(P \to Q) \leftrightarrow (\neg P \vee Q)$.
2) $(P \leftrightarrow Q) \leftrightarrow ((P \to Q) \wedge (Q \to P))$.

EXERCISE 4.18. Give a metaproof of the fact that each of the sentences below is a tautology:
1) $(P \wedge Q) \to P$.
2) $P \to (P \vee Q)$.
3) $((P \wedge Q) \wedge R) \leftrightarrow (P \wedge (Q \wedge R))$.
4) $(P \wedge Q) \leftrightarrow (Q \wedge P)$.
5) $(P \wedge (Q \vee R)) \leftrightarrow ((P \wedge Q) \vee (P \wedge R))$.
6) $(P \vee (Q \wedge R)) \leftrightarrow ((P \vee Q) \wedge (P \vee R))$.

METADEFINITION 4.19.
1) $Q \to P$ is called the *converse* of $P \to Q$.
2) $\neg Q \to \neg P$ is called the *contrapositive* of $P \to Q$.

EXERCISE 4.20. Give a metaproof of the fact that each of the sentences below is a tautology:
1) $((P \vee Q) \wedge (\neg P)) \to Q$ (modus ponens, variant).
2) $(P \to Q) \leftrightarrow (\neg Q \to \neg P)$ (contrapositive argument).
3) $(\neg(P \wedge Q)) \leftrightarrow (\neg P \vee \neg Q)$ (de Morgan law).
4) $(\neg(P \vee Q)) \leftrightarrow (\neg P \wedge \neg Q)$ (de Morgan law).
5) $((P \to R) \wedge (Q \to R)) \to ((P \vee Q) \to R)$ (case by case argument).
6) $(\neg(P \to Q)) \leftrightarrow (P \wedge \neg Q)$ (negation of an implication).
7) $(\neg(P \leftrightarrow Q)) \leftrightarrow ((P \wedge \neg Q) \vee (Q \wedge \neg P))$ (negation of an equivalence).

REMARK 4.21. 2) in Exercise 4.20 says that the contrapositive of an implication is equivalent to the original implication.

METADEFINITION 4.22. A sentence P is a contradiction if and only if $\neg P$ is a tautology.

CHAPTER 5

Theories

As usual this chapter is written in metalanguage. We introduce here theories. Theories are texts usually written in Argot. Recall the language Argot L_{Argot} is obtained from the symbols of English L_{Eng}, plus the symbols of a language L (such as Formal), plus command symbols (such as "let," "consider," etc.), plus phrases showing intension (such as "we want to show," "we need to show," "we seek a contradiction," etc.)

First we need to clarify the syntax of Argot. Rather than developing a detailed explanation the way we did for languages such as Formal (see the chapter in Syntax) we just proceed by example.

Example 5.1. Here are some sentences in Argot:

1) We want to prove U.
2) Assume P.
3) Since P and Q it follows that R.
4) Since $s = t$ we get $P(s) = P(t)$.
5) So R.
6) We know P.
7) There are 2 cases.
8) The first case is A.
9) The second case is B.
10) We seek a contradiction.
11) Let c be such that $P(c)$.
12) Assume there exists x is such that $P(x)$.
13) By P there exists c such that $Q(c)$.
14) Let c be arbitrary.
15) For all c we have $P(c)$.
16) This proves P.
etc.

Here c is a constant in L; t, s are terms in L; P, Q, R, U, A, B etc. are sentences in L; etc. Note that the constants in Formal may be used as variables in Argot. Various variants of the above sentences will be also called *sentences in Argot* (cf. the examples that follow).

In the metaaxioms below we will use various new predicates in metalanguage such as "is a theory," "is an extension of", "is labeled as axiom," "is labeled as theorem," etc.

Metaaxiom 5.2. (Theories). A sequence T of sentences labeled as axioms or definitions in L is a theory. If T is a theory and if L' is obtained from L by adding symbols and T' is obtained from T by adding new sentences (labeled as axioms, definitions, or theorems) subject to certain rules to be discussed below then T' is a theory and we say T' is an extension of T.

Here are some clarifications of the above metaaxiom.

METAAXIOM 5.3. (On axioms). Axioms are of two kinds: general axioms (present in every theory) and specific axioms (specific to the theory we are dealing with); they are given either as a "finite" list or by a rule to form them which may create an indefinitely growing list the members of which can be added at any point in the theory. The general axioms usually include the following axioms; for terms t, t', t'', functions f, and predicates p we can add at any point in the threory the following:

Axiom. $t = t$
Axiom. $(t = t') \to (t' = t)$
Axiom. $((t = t') \wedge (t' = t'')) \to (t = t'')$
Axiom. $(t = t') \to (f(t) = f(t'))$
Axiom. $(t = t') \to (p(t) \leftrightarrow p(t'))$

METAAXIOM 5.4. (On definitions). New symbols to L can be added only if, at the same time, Definitions of these symbols are being added; these additions are governed by the following rules:

Rule C. (On constants). If "*There exists a unique x such that* $P(x)$" is a Theorem or an Axiom then one can give the following definition of a new constant c:

Definition. $\forall x((x = c) \leftrightarrow P(x))$.

Here the use of the word "can" in Metalanguage needs to be paraphrased as follows. When we say in Metalanguage that "If... then one can give/add the following definition..." this needs to be paraphrased as: "If ... and if one adds the following definition ... to the theory then one obtains an extension of the theory." Similar paraphrasing will apply to our discussion below when we use the words "one is allowed," etc.

The above Definition is translated as "*c is the unique x such that* $P(x)$.

Note that if $P(x)$ equals "$x = t$" where t is a term then the above Definition can be seen (using the rules of proof to be explained below) to be equivalent to:

Definition. $c = t$.

One sometimes writes

$$c := t \quad \text{or} \quad t =: c$$

to indicate that the above is a definition of c. In natural languages the new constants are sometimes words and in the Definitions they are written in italics.

Rule P. (On predicates). If $P(x)$ is a formula with one free variable one can introduce a new predicate p and the following:

Definition. $\forall x(p(x) \leftrightarrow P(x))$.

Similarly for binary, ternary predicates, etc.

Rule F. (On functions). If $F(x, y)$ is a binary predicate and the following is a Theorem: *"For all x there exists a unique y such that $F(x, y)$"* then one can add a functional symbol f and the following:

Definition. $\forall x \forall y((F(x) = y) \leftrightarrow F(x, y)))$

Similarly for functions of several variables.

REMARK 5.5. The fact that constants and functions can only be introduced if an existence and uniqueness theorem or axiom is available is consistent with Russell's theory of descriptions according to which the use of the definite article "*the*" requires existence and uniqueness. His example is: "*The current king of France is bald*" needs to be paraphrased as "*There exists x such that x is the current king of France and for every y and every z if y is the current king of France and z is the current king of France then $y = z$ and for all x if x is the current king of France then x is bold.*" In our setting the existence and uniqueness of the current king of France needs to be an axiom or a theorem, after which we may define a new constant K as being the unique current king of France, after which we can state (and prove) a theorem saying that K is bald.

METAAXIOM 5.6. (On theorems). A theorem can be added to T only if a proof for it is also being supplied right after its statement. The end of a proof is marked by the $\square$ sign or by QED. There are several types of proof: direct proofs, proofs by contradiction, case by case proofs, and combinations of these. All these types are reducible to direct proofs. A sequence of sentences is a direct proof if it constructed by iterating the following rules. (For clarification some extra sentences in Argot may be added such as: "*We know...*," "*We want to prove...*," "*Assume the hypothesis...*," "*We have that...*," "*We conclude that...*," etc.) In the rules below we introduce the predicate "is an accepted sentence" in metalanguage. Rules 1-2 below do not involve quantifiers. Rules 3-6 below involve quantifiers and introduce in metalanguage the predicate "*is (called) a (local) existential/universal witness.*" We assume the theorem to be proved has the form $H \to C$. The sentence H is called the *hypothesis* and C is called the *conclusion*.

Rule T1. The proof starts with "Assume H. We want to prove C." At any point in the proof H, as well as all previously proved theorems, all previous definitions, all previous axioms, and all tautologies are declared to be accepted sentences. The proof ends when C is declared to be an accepted sentence.

Rule T2. If at one point in a proof P and Q are accepted sentences and if $(P \wedge Q) \to R$ is an accepted sentence then at any point later in the proof one is allowed to add the sentence "Since P and Q it follows that R" to the proof and then one declares R to be an accepted sentence. (If P and Q are equal one just adds "Since P it follows that R.") Schematically we have

Proof. .... We have that P .... We have that Q..... Since P and Q it follows that R .... $\square$

Rule T3. If at one point in a proof "$\forall x P(x)$" is an accepted sentence then at any point later in the proof one is allowed to add the sentence "Since $\forall x P(x)$ it follows that $P(c)$" to the proof where c is any constant (that has or has not been used before in the proof) and one declares $P(c)$ to be an accepted sentence. Note that $P(c)$ is obtained from $\forall x P(x)$ by deleting $\forall x$ and replacing x by c. If c is NEW (has not been used before) c must be ADDED to the language before the theorem and CAN NEVER BE USED AGAIN after the proof; in this case c is called a (local) *universal witness*. (Here "local" means it can only be used in one proof.) Schematically we have

Proof. .... So $\forall x P(x)$ .... Since $\forall x P(x)$ it follows that $P(c)$ .... $\square$

Rule T4. If at one point in a proof "$\exists x P(x)$" is an accepted sentence then at any point later in the proof one is allowed to add the sentence "Let c be such that $P(c)$" to the proof where c is a NEW constant and one declares $P(c)$ to be an accepted sentence; c must be ADDED to the language before the proof and CAN NEVER BE USED AGAIN after the proof. Note that $P(c)$ is obtained from $\exists x P(x)$ by deleting $\exists x$ and replacing x by c. Such a constant c is called a (local) *existential witness*. (Here "local" means it can only be used in one proof.) Schematically we have

Proof. .... So $\exists x P(x)$ .... Let c be such that $P(c)$ .... $\square$

Rule T5. If at one point in a proof $P(c)$ is an accepted sentence then at any point later in the proof one is allowed to add the sentence "Since $P(c)$ it follows that $\exists x P(x)$" to the proof, where c can be any constant new or not, and one declares "$\exists x P(x)$" to be an accepted sentence. Schematically we have

Proof. .... So $P(c)$ .... Since $P(c)$ we have that $\exists x P(x)$ .... $\square$

Rule T6. If at one point in the proof (including in the beginning) we say "we want to prove $\forall x P(x)$" then any time after that we may say "Let c be arbitrary; we want to (it is enough to) prove that $P(c)$" where c is a NEW constant that needs to be ADDED to the language before the proof and CAN NEVER BE USED AGAIN after the proof; such a constant is called a (local) *universal witness*. Then one provides a proof of $P(c)$ and we can say "so $\forall x P(x)$". Schematically we have

Proof. .... We want to prove $\forall x P(x)$....... Let c be arbitrary; we want to prove that $P(c)$ .... So $P(c)$. Since $P(c)$ for arbitrary c it follows that $\forall x P(x)$...... $\square$

Rule T7. At any point in the proof we can introduce new symbols (constants, predicates, or functions) following the corresponding rules. This is done by writing "Let ... be such that ...." The new symbols introduced inside the proof must not be used again outside the proof.

We do not provide rules for proofs that are not direct proofs because all other types of proofs can be reduced to direct proofs in a way that will be explained below.

REMARK 5.7. The less important Theorems are usually referred to as *Propositions*. The Propositions whose only role is to help prove other Propositions or Theorems are usually referred to as *Lemmas*. The decision as to which Theorems can be called *Propositions* or *Lemmas* is a matter of choice (of style) and is not codified in any way.

REMARK 5.8. When adding definitions of new constants, functions, or predicates one should ask, of course, that these definitions be introduced in a sequence and at each step in the sequence the symbol that is being introduced has not appeared before (i.e. it is indeed "new"); this guarantees the *predicativity* of the definitions (by which we mean here their non-circularity), at least from a syntactical viewpoint. This device does not get rid of the semantic impredicativity (which was one of the major themes in the controversies around the foundation of Mathematics at the beginning of the 20th century; cf. Russell, Poincaré, etc.) However, since we chose to completely ignore the meaning (semantics) of object languages, semantic impredicativity will not be an issue for us. To be sure, later in Set Theory, semantic impredicativity is everywhere implicit and might be viewed as implicitly threatening the whole edifice of Mathematics.

Remark 5.9. One can ask the following questions:

1) Is there a "mechanical procedure" to check if a theory (in particular its proofs) are syntactically correct?

2) Is there a "mechanical procedure" to decide if a sentence in a given theory possesses a proof?

3) Is there a "mechanical procedure" to find a proof for a theorem in a given theory provided that we know it possesses a proof?

None of these questions is correct in Metalanguage because quantifiers are being implicitly used. But one can "informally" ask these questions in English (viewed as an object language) and give "informal" answers in English. The informal answers are then YES for 1 for all theories, YES for 2 and 3 for some (trivial) theories and NO for 2 and 3 for ("most") other theories. The NO answer for 2 and 3 is what makes Mathematics a "creative activity." The YES for 1 is justified by examining the rules of proof. The NO for 2 and 3 are justified by creating a mirror of these questions inside Mathematics (in what is called *Mathematical Logic*) and proving theorems in Mathematics that "correspond" to the NO answers mentioned above. Whether or not this "correspondence" is acceptable is a philosophical issue that we will not address here. It is interesting to compare the NO answer to question 2 above with the existence of a "mechanical procedure" to decide if a given sentence is a tautology or not (Cf. Remark 4.10); the key difference between deciding provability versus deciding tautology comes from the use of quantifiers in the first context versus the absence of quantifiers in the second context. This reflects the corresponding difference between object languages and Metalanguage and it provides a sharp dividing line between the two.

Example 5.10. We consider in what follows a simple example of theory called *Group Theory.* The language L of group theory has a constant e, variables $x, y, ...$, and a binary functional symbol $\star$. We introduce the following:

Definition 1. For all z, z is a neutral element if and only if

$$\forall x((z \star x = x) \wedge (x \star z = x)).$$

(This defines "*is a neutral element*" as a new unary predicate.)

Definition 2. For all x and all y, y is an inverse of x if and only if $x{\star}y = y{\star}x = e$.

(This defines "*is an inverse of*" as a new binary predicate.)

The specific axioms of the theory are:

Axiom 1. For all x, y, z we have $x \star (y \star z) = (x \star y) \star z$.
Axiom 2. e is a neutral element.
Axiom 3. For all x there is y such that y is an inverse of x.

Theorem 1. For all z if z is a neutral element then $z = e$.

Proof. Let f be arbitrary. We want to show that if f is a neutral element then $f = e$. Since e is a neutral element it follows that $\forall x(e \star x = x)$. By the latter

$e \star f = f$. Since f is a neutral element we get $\forall x(x \star f = x)$. So $e \star f = e$. Hence we get $e = e \star f$ and so $e = f$. □

Theorem 2. For all x, y, z if y and z are inverses of x then $y = z$.

Proof. Let a, b, c be arbitrary and assume that b and c are inverses of a. We want to show that $b = c$. By Axioms 1 and 2 and Definition 2 we have

$$b \star e = b \star (a \star c) = (b \star a) \star c = e \star c = c.$$

□

In view of Axiom 3 and Theorem 2 we can introduce a functional symbol denoted by attaching the term t^{-1} to any term t via the following

Definition 3. $\forall x \forall y((y = x^{-1}) \leftrightarrow (x \star y = y \star x = e))$

One can continue the theory above by introducing more definitions and axioms and by proving more theorems; and then, again, introducing further definitions, further axioms, and proving further theorems, etc. In this way group theory can grow. We are not going to pursue this here. By the way, group theory, in the sense above, will not be very interesting. On the other hand one can create a mirror of group theory inside Set Theory. The *models* (in a technical sense to be clarified later, see the Chapter on Models) of that mirror will be called *groups*. Inside Set Theory examples of groups will be the set of non-zero real numbers with $\star$ the usual multiplication and $e = 1$; or the set of all invertible real matrices with $\star$ the usual multiplication of matrices and e the identity matrix. All of these objects will be introduced later.

EXERCISE 5.11. Explain what rules were used in our proofs of Theorems 1 and 2 above.

CHAPTER 6

Proofs

In what follows we discuss more systematically various examples of types of proofs. In the examples that follow we assume we are dealing with a theory (which is not necessarily the one considered before) in which we have axioms $A, B, \ldots$. For simplicity we begin with examples of proofs where only rules 1-2 above is being used (i.e., the rules 3-6 regarding witnesses are not being used).

EXAMPLE 6.1. (Direct proof). A direct proof could look as follows.

Theorem. $H \to C$.

Proof. Assume H. We want to prove C. Since H and A it follows that X. Since X and S it follows that Z. Since Z it follows that C. $\square$

The above counts as a proof if

$$S, \quad H \wedge A \to X, \quad X \wedge S \to Z, \quad Z \to C$$

are accepted sentences.

EXAMPLE 6.2. (Proof by contradiction). Another method of proving sentences such as $H \to C$ is by contradiction and may look as follows.

Theorem 1. $H \to C$.

Proof. Assume both H and $\neg C$ and seek a contradiction. Since $\neg C$ and A it follows that Y. On the other hand since H and S it follows that $\neg Y$. So we get that $Y \wedge \neg Y$ which is a contradiction. We conclude that $H \to C$. $\square$

The above counts as a proof if

$$S, \quad (\neg C) \wedge A \to Y, \quad H \wedge B \to \neg Y$$

are accepted sentences.

The above proof can be thought of as an abbreviation of the following two Theorems whose proofs are direct proofs:

Theorem 2. $(H \wedge \neg C) \to (Y \wedge \neg Y)$

Proof. Assume H and $\neg C$. Since $\neg C$ and A it follows that Y. On the other hand since H and S it follows that $\neg Y$. So we get that $Y \wedge \neg Y$. $\square$

Theorem 3. $H \to C$

Proof. Assume H. Since $(H \wedge \neg C) \to (Y \wedge \neg Y)$ and H is follows that $H \to C$. $\square$

A direct proof of Theorem 1 can be given by using Theorems 2 and 3 because

$$((H \wedge \neg C) \to (Y \wedge \neg Y)) \to (H \to C)$$

is a tautology and hence it is an accepted sentence.

EXAMPLE 6.3. Direct proofs and proofs by contradiction can be given to sentences which are not necessarily of the form $H \to C$. Here is an example of a proof by contradiction.

Theorem. C.

Proof. Assume $\neg C$ and seek a contradiction. Since $\neg C$ and A it follows that X. Since A and X we get Y. On the other hand since B and $\neg C$ we get $\neg Y$. So $Y \wedge \neg Y$, which is a contradiction. This ends the proof. $\square$

The above counts as a proof if

$$(\neg C) \wedge A \to X, \quad A \wedge X \to Y, \quad B \wedge \neg C \to \neg Y$$

are accepted sentences.

EXAMPLE 6.4. (Case by case proof) Say we want to prove a theorem of the form:

Theorem. $(H' \vee H'') \to C$.

Proof. There are two cases: Case 1 is H'; Case 2 is H''. Assume first that H'. From the axiom A and H' we get P. So C. Now assume that H''. Since B and H'' it follows that X. Since X we get C. So in either case we get C. This ends the proof. $\square$

The above counts as a proof if

$$A \wedge H' \to P, \quad P \to C, \quad H'' \wedge B \to X, \quad X \to C$$

are accepted sentences.

The above "case by case" strategy applies more generally to theorems of the form

Theorem. $H \to C$

Proof. There are two cases:
Case 1: W holds.
Case 2: $\neg W$ holds.
Assume first we are in Case 1. Then by A and W we get P. So C.
Now assume we are in Case 2. Since B and $\neg W$ it follows that X. Since X we get C.
So in either case we get C. This ends the proof. $\square$

The above counts as a proof if W is any sentence and

$$A \wedge W \to P, \quad P \to C, \quad (\neg W) \wedge B \to X, \quad X \to C$$

are accepted sentences.

Note that, in the latter proof, finding a sentence W and dividing the proof in two cases according as W or $\neg W$ holds is usually a creative act: one needs to guess what W will work.

Finally note that the case by case proof above should be viewed, again, as an abbreviation of a direct proof in the same way in which proofs by contradiction were reduced to direct proofs. Indeed what one can do is first give a direct proof of:

Theorem. $(H \wedge W) \to C$

Then give a direct proof of:

Theorem. $(H \wedge \neg W) \to C$

Then finally prove:

Theorem. $H \to C$

using the fact that

$$(((H \wedge W) \to C) \wedge (H \wedge \neg W) \to C) \to (H \to C)$$

is a tautology. We leave the details to the reader. So after all there is only one basic type of proof, the direct proof.

EXAMPLE 6.5. Here is an example that combines proof by contradiction with "case by case" proof. Say we want to prove:

Theorem. $H \to C$.

Proof. Assume H and $\neg C$ and seek a contradiction. There are two cases:
Case 1. W holds.
Case 2. $\neg W$ holds.
In case 1, by ... we get ... hence a contradiction. In case 2, by ... we get ... hence a contradiction. This ends the proof. □

EXAMPLE 6.6. Sometimes a theorem U has the statement:

Theorem. The following conditions are equivalent:
1) P;
2) Q;
3) R.

What is being meant is that U is

$$(P \leftrightarrow Q) \wedge (P \leftrightarrow R) \wedge (Q \leftrightarrow R)$$

One proceeds "in a circle" by proving first $P \to Q$ then $Q \to R$ then $R \to P$.

EXAMPLE 6.7. In order to prove a theorem of the form $P \wedge Q$ one first proves P and proves Q.

EXAMPLE 6.8. In order to prove a theorem of the form $P \vee Q$ one may proceed by contradiction as follows. Assume $\neg P$ and $\neg Q$ and one seeks a contradiction.

EXAMPLE 6.9. In order to prove a theorem U of the form $P \leftrightarrow Q$ one first proves $P \to Q$ and then one proves $Q \to P$.

Next we consider examples of proofs where, in addition to rules 1-2, the rules 3-6 (governing witnesses) are being used as well.

EXAMPLE 6.10. Here is an example of proof by contradiction that involves witnesses. Here $P(x)$ is any formula with exactly one free variable x.

Theorem. $(\forall x(\neg P(x))) \to (\neg(\exists x P(x)))$.

Proof. Assume both $\forall x(\neg P(x))$ and $\neg\neg(\exists x P(x))$ and seek a contradiction. Since $\neg\neg(\exists x P(x))$ it follows that $\exists x P(x)$. Let e be such that $P(e)$. Now since $\forall x(\neg P(x))$ we get in particular $\neg P(e)$, a contradiction. This ends the proof. □

The constant e introduced in the proof is a new constant, is being added to the language before the proof and CAN NEVER BE USED AGAIN after the proof; it is an existential witness.

EXAMPLE 6.11. A proof can start as a direct proof and involve later an embedded argument by contradiction. Here is an example (which also involves witnesses).

Theorem. $(\neg(\exists x P(x))) \rightarrow (\forall x(\neg P(x)))$.

Proof. Assume $\neg(\exists x P(x))$. We want to show that $\forall x(\neg P(x))$. Let a be arbitrary; we want to show that $\neg P(a)$. Assume $\neg\neg P(a)$ and seek a contradiction. Since $\neg\neg P(a)$ it follows that $P(a)$. Since $P(a)$ it follows that $\exists x P(x)$. But we assumed $\neg(\exists x P(x))$. This is a contradiction which ends the proof. $\square$

The constant a in the above proof is a universal witness; it is being added to the language before the proof and CAN NEVER BE USED AGAIN after the proof.

EXERCISE 6.12. Prove the following:

Theorem. $(\neg(\forall x P(x))) \leftrightarrow (\exists x(\neg P(x)))$

EXAMPLE 6.13. In the same way one can prove

Theorem. $(\neg(\forall x \exists y P(x,y))) \leftrightarrow (\exists x \forall y(\neg P(x,y)))$.

Proof. We first prove the implication $\rightarrow$.

Assume $\neg(\forall x \exists y P(x,y))$ and $\neg(\forall x \exists y(\neg P(x,y)))$ and seek a contradiction. By Theorem in Example 6.12 we get $\exists x(\neg(\exists y(\neg P(x,y))))$. Let c be such that we have $\neg(\exists y(\neg((P(c,y)))$. By the Theorem in Example 6.11 we get that $\forall y(\neg\neg P(c,y))$. So $\exists x \forall y P(x,y)$, a contradiction.

We now prove the implication $\leftarrow$.

Assume $\exists x \forall y(\neg P(x,y))$ and $\forall x \exists y P(x,y)$ and seek a contradiction. Let d be such that $\forall y(\neg P(d,y))$. Since $\forall x \exists y P(x,y)$ it follows that $\exists y P(d,y)$. Since $\forall y(\neg P(d,y))$ it follows from the Theorem in Example 6.11 that $\neg(\exists y P(d,y))$, a contradiction.

EXAMPLE 6.14. In the same way one can prove theorems of the form

Theorem. $(\neg(\forall x \forall y \exists z \forall t P(x,y,z,t))) \leftrightarrow (\exists x \exists y \forall z \exists(\neg P(x,y,z,t)))$.

Etc. In other words to negate a sentence that starts with quantifiers one "flips" the quantifiers $\forall$ and $\exists$ and one negates the predicate involved.

EXAMPLE 6.15. If the theorem that is to be proved has the form

Theorem. $\exists x P(x)$

then one proceeds as follows. One starts the proof with the sentence "It is sufficient to prove that $P(c)$" where c is a constant that HAS BEEN USED BEFORE the statement of the theorem; such a constant is called a (non-local) *existential witness* and CAN BE USED after the end of the proof. Then one provides a proof for $P(c)$. Schematically we have

Proof. It is sufficient to prove that $P(c)$. By ... it follows that... Hence .... So we conclude that $P(c)$. $\square$

This is justified by noting that one can first prove the

Theorem. $P(c)$

and then one proves the Theorem $\exists x P(x)$ by using Rule 5.

METADEFINITION 6.16. Let S be a sentence. A negation of S is a sentence R such that $\neg S \leftrightarrow R$ is a theorem. We say that a negation R of S is in reduced form if the connective $\rightarrow$ does not appear in R and the connective $\neg$ only appears in R in front of atomic formulae contained in R.

EXAMPLE 6.17. Let P, Q be atomic sentences. Let S be the sentence

$$P \wedge Q$$

Let S', S'' be the sentences

$$S' = \text{“}\neg(P \wedge Q)\text{”}$$
$$S'' = \text{“}\neg P \vee \neg Q\text{”}$$

Then S' is a negation of S but is it not in reduced form because $\neg$ appears in front of a non-atomic formula. On the other hand S'' is a negation of S in reduced form. It is a negation because

$$\neg(P \wedge Q) \leftrightarrow (\neg P \vee \neg Q)$$

is a tautology (hence a theorem); it is in reduced form because $\neg$ appears in it only in front of atomic formulae and $\rightarrow$ does not appear in S''.

EXAMPLE 6.18. We claim that we have the following

Theorem. $(\neg(\exists x(P(x) \wedge Q(x))) \leftrightarrow (\forall x(\neg P(x) \vee \neg Q(x)))$

So the negation in reduced form of $\exists x(P(x) \wedge Q(x))$ is $\forall x(\neg P(x) \vee \neg Q(x))$. Note that $\neg(\exists x(P(x) \wedge Q(x))$ is also a negation of $\exists x(P(x) \wedge Q(x)$ but it is not in reduced form because $\neg$ in it does not appear in front of an atomic formula. Here is the proof of the Theorem.

Proof. We check the implication $\rightarrow$; the converse is similar.

Assume $(\neg(\exists x(P(x) \wedge Q(x)))$; we want to prove that $\forall x(\neg P(x) \vee \neg Q(x))$. To prove that $\forall x(\neg P(x) \vee \neg Q(x))$ we consider an arbitrary b and we need to show that $\neg P(b) \vee \neg Q(b)$. Hence we need to show that $\neg(P(b) \wedge Q(b))$. Assume $P(b) \wedge Q(b)$ and seek a contradiction. Since $P(b) \wedge Q(b)$ it follows that $\exists x(P(x) \wedge Q(x))$, a contradiction. $\square$

EXERCISE 6.19. Find negations in reduced form of the following sentences. (Provide the answers but do not provide proofs justifying your answers.) In this exercise P, Q are atomic formulae involving two variables.

1) $\forall x \forall y(P(x, y) \wedge Q(x, y))$.
2) $\forall x \forall y(P(x, y) \vee Q(x, y))$.
3) $\exists x \forall y(P(x, y) \rightarrow (\exists z Q(z, y)))$.
4) $(\exists x \forall y P(x, y)) \rightarrow (\forall y \exists x Q(x, y))$.

EXERCISE 6.20. Formalize the following sentences. Find the negations in reduced form of those formalizations. Translate into English those negations.

1) If Plato is a bird then Plato eats nuts.
2) John is not Plato and Paul is not Aristotle.

3) For every line and every point that does not belong to the line there exists no line passing through the point and parallel to the first line.

4) For every line and every point that does not belong to the line there exists at least two distinct lines passing through the point and parallel to the first line.

5) For every ϵ there exists δ such that for all x if $|x-a| < \delta$ then $|f(x)-f(a)| < \epsilon$. (Here f and $|\ |$ are unary functions and $<$ is a binary predicate. The sentence above plays a role in calculus.)

We next introduce the predicate "is inconsistent" into metalanguage.

Metaaxiom 6.21. If T is a theory and A is a sentence in T and $A \wedge \neg A$ is in T then the theory T is inconsistent. If T is an extension of T' then T' is inconsistent.

Remark 6.22. One is tempted to make a metadefinition along the following lines: A theory is inconsistent if there exists an extension of that theory that contains a sentence of the form $A \wedge \neg A$. A theory is consistent if it is not inconsistent. A theory is complete if there is an extension of the theory such that for any sentence A in L either A belongs to the extension or $\neg A$ belongs to the extension. A theory is incomplete if it is not complete. However the above metadefinition is not correct in our setting because it contains quantifiers. The concepts of Logic suggested here can be imitated by concepts in Set Theory (i.e., in Mathematics; cf. our last Chapter) and then theorems about set theoretic completeness and consistency can be proved in Set Theory (Gödel's theorems, for instance, which will not be touched in this course); however these latter theorems are not metatheorems in Logic (i.e., about sentences) but rather theorems in Set Theory (i.e., about nothing).

We end by discussing fallacies. A fallacy is a logical mistake. Here are some typical fallacies:

Example 6.23. *Confusing an implication with its converse.* Say we want to prove that $H \to C$. A typical mistaken proof would be: Assume C; then by ... we get that ... hence H. The error consists of having proved $Q \to P$ rather than $P \to Q$.

Example 6.24. *Proving a universal sentence by example.* Say we want to prove $\forall x P(x)$. A typical mistaken proof would be: By ... there exists c such that ... hence ... hence $P(c)$. The error consists in having proved $\exists x P(x)$ rather than $\forall x P(x)$.

Example 6.25. *Defining a constant twice.* Say we want to prove $\neg(\exists x P(x))$ by contradiction. A mistaken proof would be: Assume there exists c such that $P(c)$. Since we know that $\exists x Q(x)$ let c be (or define c) such that $Q(c)$. By $P(c)$ and $Q(c)$ we get ... hence ..., a contradiction. The error consists in defining c twice in two unrelated ways: first c plays the role of an existential witness for P; then c plays the role of an existential witness for Q. But these existential witnesses are not the same.

Exercise 6.26. Give examples of wrong proofs of each of the above types. If you can't solve this now, wait until we get to discuss the integers.

Remark 6.27. Later, when we discuss induction we will discuss another typical fallacy; cf. Example 14.8.

CHAPTER 7

Examples

In this Chapter we analyze some toy examples of theories that are unrelated to Mathematics. In the next Chapter we will present the main example of theory in this course which is Set Theory (identified with Mathematics itself). The present chapter will not be used in the sequel and may be skipped: it is an "intermezzo" between our treatment of Logic and that of Mathematics.

EXAMPLE 7.1. We begin with "Pascal's wager." The structure of Pascal's wager argument is as follows. If God exists and I believe it exists then I will be saved. If God exists and I do not believe it exists then I will not be saved. If God does not exist but I believe it exists I will not be saved. Finally if God does not exist and I do not believe it exists then I will not be saved. Pascal's conclusion is that if he believes that God exists then there is a one chance in two that he be saved whereas if he does not believe that God exists then there is a zero chance that he be saved. So he should believe that God exists. The next example is a variation of Pascal's wager showing that if one requires "sincere" belief rather than just belief based on Logic then Pascal will not be saved. Indeed assume the specific axioms:

Axiom 1. If God exists and a person does not believe sincerely in its existence then that person will not be saved.

Axiom 2. If God does not exist then nobody will be saved.

Axiom 3. If a person believes that God exists and his/her belief is motivated only by Pascal's wager then that person does not believe sincerely.

We want to prove the following

Theorem 1. If Pascal believes that God exists but his belief is motivated by his own wager only then Pascal will not be saved.

All of the above is formulated in the English language L'. We consider a simpler language L and a translation of L into L'.

The new language L contains among its constant p (for Pascal) and contains 4 unary predicates g, w, s, r whose translation in English is as follows:

g is translated as "*is God*"
w is translated as "*believes motivated only by Pascal's wager*"
s is translated as "*believes sincerely*"
r is translated as "*is saved*"

The specific axioms are

Axiom 1. $\forall y((\exists x g(x)) \wedge (\neg s(y)) \to \neg r(y))$.

Axiom 2. $\forall y((\neg(\exists x g(x))) \to (\neg r(y)))$.
Axiom 3. $\forall y(w(y) \to (\neg(s(y))))$.

In this language Theorem 1 is the translation of the following:

Theorem 2. If *w(p)* then $\neg r(p)$.

So to prove Theorem 1 in L' it is enough to prove Theorem 2 in L. We will do this by using a combination of direct proof and case by case proof.

Proof of Theorem 2. Assume $w(p)$. There are two cases: the first case is $\exists x g(x)$; the second case is $\neg(\exists x g(x))$. Assume first that $\exists x g(x)$. Since $w(p)$, by Axiom 3 it follows that $\neg s(p)$. By Axiom 1 $(\exists x g(x)) \wedge (\neg s(p)) \to \neg r(p))$. Hence $\neg r(p)$. Assume now $\neg(\exists x g(x))$. By Axiom 2 we then get again $\neg r(p)$. So in either case we get $\neg r(p)$ which ends the proof. $\square$

Example 7.2. Here is the structure of "Descartes' cogito." We consider a language L containing the constant "I", the unary predicates "think" and "exist" and the binary predicate "doubts." Also consider the axioms (written in Argot):

Axiom 1. For all y I doubt y. (Absolute doubt.)
Axiom 2. For all x and all y if x doubts y then x thinks. (Doubting is a form of thinking.)
Axiom 3. For all x if x thinks then x exists. (The cogito.)

We have the following

Theorem 1. I exist.

Proof. From Axiom 1 we get that I doubt I. From Axiom 2 (making x and y equal to I) we get that if I doubt I then I think. From the latter and Axiom 3 (with x replaced by I) we get that I exist. $\square$

Example 7.3. The next example is the famous "ontological argument" for the existence of God (cf. Anselm, Descartes, Leibnitz, Gödel). The version below is, in some sense, a "baby version" of the argument; Gödel's formalization (which he never published) is considerably subtler. Cf. (Wang 1996). The structure of the classical ontological argument for the existence of God is as follows. Let us assume that qualities (same as properties) are either positive or negative (and none is both). Let us think of *existence* as having 2 kinds: *existence in mind* (which shall be referred to as *belonging to mind*) and *existence in reality* (which shall be referred to as *belonging to reality*). It is not important that we do not know what mind and reality are; we just see them as English words here. The 2 kinds are not necessarily related: belonging to mind does not imply (and is not implied by) belonging to reality. (In particular we do *not* view mind necessarily as part of reality which we should not: unicorns belong to mind but not to reality.) The constants and variables refer to things (myself, my cat, God,...) or qualities (red, omnipresent, deceiving, eternal, murderous, mortal,...); we identify the latter with their extensions which are, again, things (the Red, the Omnipresent, the Deceiving, the Eternal, the Murderous, the Mortal,...) In particular we consider the following constants: *reality, mind, God, the Positive Qualities.* We also consider the binary predicate *belongs to.* We say a thing has a certain quality (e.g. *my cat is eternal*) if that thing belongs to the extension of that quality (e.g. *my cat belongs to the Eternal*). Assume the following axioms:

Axiom 1. There exists a thing belonging to mind that has all the positive qualities and no negative quality.
Axiom 2. "Being real" is a positive quality.
Axiom 3. Two things belonging to mind that have exactly the same qualities are identical. (This is Leibniz's famous principle of *identity of indiscernibles.*)

Axioms 1 and 3 allow defining God as the only thing in mind which has all positive qualities and no negative quality; then one can prove the following:

Theorem 1. Being real is a quality belonging to God.

In other words God is real.

The above sentences are written in the English language L'. Let us formalize the above in a language L and prove a formal version of Theorem 1 in L whose translation is Theorem 1. Assume L contains among its constants the constants r, m, p and a binary predicate β. We consider a translation of L into L' such that

r is translated as "reality" or "being real";
m is translated as "mind";
p is translated as "the positive qualities"
$x\beta y$ is translated as "x belongs to y".

The specific axioms are:

Axiom 1. $\exists x((x\beta m) \wedge (\forall z((z\beta p) \leftrightarrow (z\beta x))))$.
Axiom 2. $r\beta p$.
Axiom 3. $\forall x\forall y(((x\beta m) \wedge (y\beta m)) \rightarrow ((\forall z(z\beta x \leftrightarrow z\beta y)) \rightarrow (x = y)))$.

Note that later, in Set Theory, we will have a predicate $\in$ which, like β, will be translated as "belongs to" (as in an object belongs to the collection of objects that have a certain quality); but the axioms are different.

In view of Axioms 1 and 3 we may introduce a new constant G and the following:

Definition 1. $\forall x((x = G) \leftrightarrow ((x\beta m) \wedge (\forall z((z\beta p) \leftrightarrow (z\beta x)))))$.

We will translate G as "God". The translation of the above in English is: "*Something is God if and only if that something is in my mind and has all the positive qualities but no negative quality.*" We have the following Theorem whose translation in L' is "*Reality is a quality belonging to God*":

Theorem 3. $r\beta G$.

Proof. By Axiom 1 we have $G\beta m$ and

$$\forall z((z\beta p) \leftrightarrow (z\beta G)).$$

Hence we have, in particular,

$$(r\beta p) \leftrightarrow (r\beta G).$$

By Axiom 2, $r\beta p$. Hence $r\beta G$. $\square$

The argument above is, of course, correct. What is questionable is the choice of the axioms and the reference of L. Also recall that, in our course, the question of truth was not addressed; so it does not make sense to ask whether the English sentence "*God has existence in reality*" is true of false. For criticism of the relevance of this argument (or similar ones) see, for instance, (Kant 1991) and (Wang 1996). However, the mere fact that some of the most distinguished logicians of all times (in particular Leibniz and Gödel) took this argument seriously shows that the argument has merit and, in particular, cannot be dismissed on trivial grounds.

EXERCISE 7.4. Explain why Definition 1 in Example 7.3 above is correct.

EXAMPLE 7.5. The next example is again a toy example and comes from physics. In order to present this example we do not need to introduce any physical concepts. But it would help to keep in mind the two slit experiment in quantum mechanics (for which we refer to Feynman's Physics course, say). Now there are two types of physical theories that can be referred to as *phenomenological* and *explanatory*. They are intertwined but very different in nature. Phenomenological theories are simply descriptions of phenomena/effects of (either actual or possible) experiments; examples of such theories are those of Ptolemy, Copernicus, or that of pre-quantum experimental physics of radiation. Explanatory theories are systems postulating transcendent causes that act from behind phenomena; examples of such theories are those of Newton, Einstein, or quantum theory. The theory below is a baby example of the phenomenological (pre-quantum) theory of radiation; our discussion is therefore not a discussion of quantum mechanics but rather it suggests the necessity of introducing quantum mechanics. The language L' and definitions are those of experimental/phenomenological (rather than theoretical/explanatory) physics. We will not make them explicit. Later we will move to a simplified language L and will not care about definitions.

Consider the following specific axioms (which are the translation in English of the phenomenological predictions of classical particle mechanics and classical wave mechanics, respectively):

Axiom 1. If radiation in the 2 slit experiment consists of a beam of particles then the impact pattern on the photographic plate consists of a series of successive flashes and the pattern has 2 local maxima.

Axiom 2. If radiation in the 2 slit experiment is a wave then the impact pattern on the photographic plate is not a series of successive flashes and the pattern has more than 2 local maxima.

We want to prove the following

Theorem 1. If in the 2 slit experiment the impact consists of a series of successive flashes and the impact pattern has more than 2 local maxima then in this experiment radiation is neither a beam of particles nor a wave.

The sentence reflects one of the elementary puzzles that quantum phenomena exhibit: radiation is neither particles nor waves but something else! And that something else requires a new theory which is quantum mechanics. (A common fallacy would be to conclude that radiation is both particles and waves !!!) Rather than analyzing the language L' of physics in which our axioms and sentence are stated (and the semantics that goes with it) let us introduce a simplified language L as follows.

We consider the language L with constants $a, b, ...$, variables $x, y, ...$, and unary predicates p, w, f, m. Then there is a translation of L into L' such that:

p is translated as "is a beam of particles"
w is translated as "is a wave"
f is translated as "produces a series of successive flashes"
m is translated as "produces a pattern with 2 local maxima"

Then we consider the specific axioms

Axiom 1. $\forall x(p(x) \to (f(x) \wedge m(x)))$.
Axiom 2. $\forall x(w(x) \to (\neg f(x)) \wedge \neg m(x)))$.

Here we tacitly assume that the number of maxima cannot be 1. Theorem 1 above is the translation of the following theorem in L:

Theorem 2. $\forall x((f(x) \wedge (\neg m(x))) \to ((\neg p(x)) \wedge (\neg w(x))))$.

So it is enough to prove Theorem 2. The proof below is, as we shall see, a combination of proof by contradiction and case by case.

Proof. We proceed by contradiction. So assume there exists a such that

$$f(a) \wedge (\neg m(a))$$

and

$$\neg(\neg p(a) \wedge (\neg w(a)))$$

and seek a contradiction. Since $\neg(\neg p(a) \wedge (\neg w(a)))$ we get $p(a) \vee w(a)$. There are two cases. The first case is $p(a)$; the second case is $w(a)$. We will get a contradiction in each of these cases separately. Assume first $p(a)$. Then by Axiom 1 we get $f(a) \wedge m(a)$, hence $m(a)$. But we assumed $f(a) \wedge (\neg m(a))$, hence $\neg m(a)$, so we get a contradiction. Assume now $w(a)$. By Axiom 2 we get $(\neg f(a)) \wedge (\neg m(a))$ hence $\neg f(a)$. But we assumed $f(a) \wedge (\neg m(a))$, hence $f(a)$, so we get again a contradiction. $\square$.

Exercise 7.6. Consider the specific Axioms 1 and 2 in Example 7.5 above and also the specific axioms:

Axiom 3. $\exists x(f(x) \wedge (\neg m(x)))$.
Axiom 4. $\forall x(p(x) \vee w(x))$.

Metaprove that the theory with specific Axioms 1, 2, 3, 4 is inconsistent. Axiom 3 is translated as saying that in some experiments one sees a series of successive flashes and, at the same time, one has more than 2 maxima. Axiom 4 is translated as saying that any type of radiation is either particles or waves. The inconsistency of the theory says that classical (particle and wave) mechanics is not consistent with experiment. (So a new mechanics, quantum mechanics, is needed.) Note that none of the above discussion has anything to do with any concrete proposal for a quantum mechanical theory; all that the above suggests is the necessity of such a theory.

Example 7.7. The next example is a logical puzzle from the Mahabharata. King Yudhishthira loses his kingdom to Sakuni at a game of dice; then he stakes himself and he loses himself; then he stakes his wife Draupadi and loses her too. She objects by saying that her husband could not have staked her because he did not own her anymore after losing himself. Here is a possible formalization of her argument.

We use a language with constants $i, d, ...$, variables $x, y, z, ...$, the binary predicate "*owns*," quantifiers, and equality $=$. We define a predicate $\neq$ by $(x \neq y) \leftrightarrow (\neg(x = y))$. Consider the following specific axioms:

Axiom 1. For all x, y, z if x owns y and y owns z then x owns z.
Axiom 2. For all y there exists x such that x owns y.
Axiom 3. For all x, y, z if y owns x and z owns x then $y = z$.

We will prove the following

Theorem. If i does not own himself then i does not own d.

Proof. We proceed by contradiction. So we assume i does not own i and i owns d and seek a contradiction. There are two cases: first case is d owns i; the second case is d does not own i. We prove that in each case we get a contradiction. Assume first that d owns i; since i owns d, by Axiom 1, i owns i, a contradiction. Assume now d does not own i. By Axiom 2 we know that there exists j such that j owns i. Since i does not own i it follows that $j \neq i$. Since j owns i and i owns d, by Axiom 1, j owns d. But i also owns d. By Axiom 3, $i = j$, a contradiction. $\square$

Example 7.8. This example deals with the concepts of free will from an ontological vs ethical/juridical standpoint. Consider a language that contains constants "ontological free will" and "psychological free will" plus binary predicates "has" and "believes in" plus unary predicates "is pre-determined" (by fate, God, physical laws, etc), "is responsible", "is delusional", plus a function "all the actions of". Consider the following axioms.

Axiom 1. For every x, x has ontological free will if and only if it is not the case that all the actions of x are pre-determined.

Axiom 2. For every x, x has psychological free will if and only if x believes in ontological free will.

Axiom 3. For every x, if x has psychological free will then x is responsible.

Axiom 4. If x believes in y and y does not exist then x is delusional.

Then one has the following

Theorem. For all x, if ontological free will does not exist and if x has psychological free will then x is both responsible and delusional.

In the above "x exists" should be thought of as "there exists y such that $y = x$".

Modern physics seems to imply that ontological free will does not exist; on the other hand human society cannot function unless most people have psychological free will; so by the theorem above most people are both responsible and delusional. Hence morality is based on a delusion. But of course, human society needs a whole array of delusions in order to function at all.

EXAMPLE 7.9. This example illustrates the logical structure of the Newtonian theory of gravitation that unified Galileo's phenomenological theory of falling bodies (the physics on Earth) with Kepler's phenomenological theory of planetary motion (the physics of "Heaven"); Newton's theory counts as an explanatory theory because its axioms go beyond the "facts" of experiment. The language L in which we are going to work has variables $x, y, ...$, constants S, E, M (translated into English as "Sun, Earth, Moon"), a constant R (translated as "the radius of the Earth"), constants $1, \pi, r$ (where r is translated as a particular rock), predicates p, c, n (translated into English as "is a planet, is a cannonball, is a number"), a binary predicate $\circ$ (whose syntax is $x \circ y$ and whose translation in English is "x revolves around the fixed body y"), a binary predicate f (where $f(x, y)$ is translated as "x falls freely under the influence of y"), a binary functional symbol d ("distance between the centers of"), a unary functional symbol a ("acceleration"), a unary functional symbol T (where $T(x, y)$ is translated as "period of revolution of x around y"), binary functions $:, \times$ ("division, multiplication"), and all the standard connectives, quantifiers, and parentheses. Note that we have no predicates for mass and force; this is remarkable because it shows that the Newtonian revolution has a purely geometric content. Now we introduce a theory T in L via its special axioms. The special axioms are as follows. First one asks that distances are numbers:

$\forall x \forall y (n(d(x, y)))$

and the same for accelerations, and times of revolution. (Note that we view all physical quantities as measured in centimeters and seconds.) For numbers we ask that multiplication and division of numbers are numbers:

$(n(x) \wedge n(y)) \to (n(x : y) \wedge n(x \times y))$

and that the usual laws relating $:$ and $\times$ hold. Here are two:

$\forall x(x : x = 1)$.
$\forall x \forall y \forall z \forall u((x : y = z : u) \leftrightarrow (x \times u = z \times y))$.

It is an easy exercise to write down all these laws. We sometimes write

$$\frac{x}{y}, 1/x, xy, x^2, x^3, ...$$

in the usual sense. The above is a "baby Mathematics" and this is all Mathematics we need. Next we introduce an axiom whose justification is in Mathematics, indeed in calculus; here we ignore the justification and just take this as an axiom. The axiom gives a formula for the acceleration of a body revolving in a circle around a fixed body. (See the exercise after this example.) Here is the axiom:

Axiom A. $\forall x \forall y \left(\left(x \circ y \right) \to (a(x, y) = \frac{4\pi^2 d(x,y)}{T^2(x,y)} \right) \Big)$.

To this one adds the following "obvious" axioms

Axiom O1. $\forall x(c(x) \to d(x, E) = R)$,
Axiom O2. $M \circ E$,
Axiom R. $c(r)$,
Axiom K1. $\forall x(p(x) \to (x \circ S))$,

saying that the distance between cannonballs and the center of the Earth is the radius of the Earth; that the Moon revolves around the Earth; that the rock r is a cannonball; and that all planets revolve around the Sun. (The latter is Kepler's first law in an approximate form; the full Kepler's first law specifies the shape of orbits as ellipses, etc.) Now we consider the following sentences (NOT AXIOMS!):

G =“$\forall x \forall y((c(x) \wedge c(y)) \rightarrow (a(x,E) = a(y,E))$”,
K3= “$\forall x \forall y \left((p(x) \wedge p(y)) \rightarrow \left(\frac{d^3(x,S)}{T^2(x,S)} = \frac{d^3(y,S)}{T^2(y,S)}\right)\right)$”,
N= “$\forall x \forall y \forall z \left((f(x,z) \wedge f(y,z)) \rightarrow \left(\frac{a(x,z)}{1/d^2(x,z)} = \frac{a(y,z)}{1/d^2(y,z)}\right)\right)$”.

G represents Galileo's great empirical discovery that all cannonballs (by which we mean here terrestrial airborne objects with no self-propulsion) have the same acceleration towards the Earth. $K3$ is Kepler's third law which is his empirical great discovery that the cubes of distances of planets to the Sun are in the same proportion as the squares of their periods of revolution. Kepler's second law about equal areas being swept in equal times is somewhat hidden in axiom A above. N is Newton's law of gravitation saying that the accelerations of any two bodies moving freely towards a fixed body are in the same proportion as the inverses of the squares of the respective distances to the (center of the) fixed body. Newton's great invention is the creation of a binary predicate f (where $f(x,y)$ is translated into English as “x is in free fall with respect to y”) equipped with the following axioms

Axiom F1. $\forall x(c(x) \rightarrow f(x,E))$
Axiom F2. $f(M,E)$
Axiom F3. $\forall x(p(x) \rightarrow f(x,S))$

expressing the idea that cannonballs and the Moon moving relative to the Earth and planets moving relative to the Sun are instances of a more general predicate expressing “free falling.” Finally let us consider the following

Definition. $g = a(r,E)$

and the following sentence:

$X =$ “$g = \frac{4\pi^2 d^3(M,E)}{R^2 T^2(M,E)}$”.

The main results are the following theorems in T:

Theorem 1. $N \rightarrow X$.

Proof. See the exercise after this example.

Theorem 2. $N \rightarrow G$.

Proof. See the exercise after this example.

Theorem 3. $N \rightarrow K3$.

Proof. See the exercise after this example.

So if one accepts Newton's N then Galileo's G and Kepler's $K3$ follow, that is to say that N “unifies” terrestrial physics with the physics of Heaven. The beautiful thing is, however, that N not only unifies known paradigms but “predicts” new “facts,” e.g., X. Indeed one can verify X using experimental (astronomical and terrestrial physics) data: if one enlarges our language to include numerals and numerical computations and if one introduces axioms as below (justified by measurements) then X becomes a theorem. Here are the additional axioms:

Axiom. $g = 981$ (the number of centimeters per second squared representing g).

Axiom. $\pi = \frac{314}{100}$ (approximate value).

Axiom. R =number of centimeters representing the radius of the Earth (measured for the first time by Eratosthenes using shadows at two points on Earth).

Axiom. $d(M, E)$ =number of centimeters representing the distance from Earth to Moon (measured using parallaxes).

Axiom. $T(M, E)$ =number of seconds representing the time of revolution of the Moon (the equivalent of 28 days).

The fact that X is verified with the above data is the miraculous computation done by Newton that convinced him of the validity of his theory; see the exercise after this example.

REMARK 7.10. This part of Newton's early work had a series of defects: it was based on the circular (as opposed to elliptical) orbits, it assumed the center of the Earth (rather than all the mass of the Earth) as responsible for the effect on the cannonballs, it addressed only revolution around a fixed body (which is not realistic in the case of the Moon, since, for instance, the Earth itself is moving), and did not explain the difference between the d^3/T^2 of planets around the Sun and the corresponding quantity for the Moon and cannonballs relative to the Earth. Straightening these and many other problems is part of the reason why Newton postponed publication of his early discoveries. The final theory of Newton involves the introduction of absolute space and time, mass, and forces. The natural way to develop it is within Mathematics, as mathematical physics; this is essentially the way Newton himself presented his theory in published form. However, the above example suggests that the real breakthrough was not mathematical but at the level of (general) Logic.

EXERCISE 7.11.

1) Justify Axiom A in Example 7.9 above using calculus or even Euclidean geometry plus the definition of acceleration in an introductory physics course.

2) Prove Theorems 1,2,3 in Example 7.9.

3) Verify that with the numerical data for $g, \pi, R, d(M, E), T(M, E)$ in Example 7.9 available from astronomy (find the numbers in astronomy books) the sentence X becomes a theorem. This is Newton's fundamental computation that convinced him of the plausibility of his hypothesis that the Moon and the terrestrial objects are subject to a common law.

CHAPTER 8

Mathematics

Historically Mathematics developed as what we would today call a collection of theories written in an Argot language. Nowadays, however, Mathematics is identified (at every moment in time) with a particular (extension of a) theory T_{set} (called *Set Theory*) in a particular language L_{set}, called the *language of Set Theory*, with specific axioms called the ZFC axioms (the *Zermelo-Fraenkel+Choice* axioms). The specific axioms do not form a "finite" list so they cannot be all listed at the beginning of the theory; they have to be added one by one on a need to use basis to the various successive extensions of the theory. There are rules in Metalanguage regulating what type of new axioms (as well as new symbols) one can add. In spite of the above picture, all mathematical discourse in books and articles (including this course) continues to be presented in Argot; this is done for the sake of comprehensibility with the understanding that the discourse can be, at any time, formalized. The advantage of formalization lies, of course, in the precision of formal language (as opposed to the classical mathematical Argot which is often too vague and may tempt one to use intuitions based on a realist interpretation).

Metadefinition 8.1. The original language L_{set} of Set Theory is the language with variables $x, y, z, ...$, no constants, no functional symbol, a binary predicate $\in$, connectives $\vee, \wedge, \neg, \rightarrow, \leftrightarrow$, quantifiers $\forall, \exists$, equality $=$, and separators $(,),,$. As usual we are allowed to add, whenever it is convenient, new constants, usually denoted by

$$a, b, c, ..., A, B, C, ..., \mathcal{A}, \mathcal{B}, \mathcal{C}, ..., \alpha, \beta, \gamma, ...,$$

and new predicates to L_{set} together with definitions for each of these new symbols. In this way extensions of Set Theory are being obtained which we still refer to as Set Theory. The terms of L_{set} will be called *sets*.

In particular we introduce new predicates $\neq$, $\notin$, $\subset$, $\not\subset$ and the following definition for them:

Definition 8.2.

$\forall x \forall y((x \neq y) \leftrightarrow (\neg(x = y)))$.
$\forall x \forall y((x \notin y) \leftrightarrow (\neg(x \in y)))$.
$\forall x \forall y((x \subset y) \leftrightarrow (\forall z((z \in x) \rightarrow (z \in y))))$.
$\forall x \forall y((x \not\subset y) \leftrightarrow (\neg(x \subset y)))$.

Remark 8.3. We recall the fact that L_{set} being an object language it does not make sense to say that a sentence in it (such as, for instance, $a \in b$) is true or false.

Remark 8.4. Later we will introduce the concept of "countable" set and we will show that not all sets are countable. On the other hand in Set Theory there are always only "finitely many" sets (in the sense that one is using finitely many symbols) although their collection may be increased any time, if necessary. Let us say that such a collection of symbols is "metacountable." This seems to be a paradox which is referred to as the "Skolem paradox." Of course this is not going to be a paradox: "metacountable" and "countable" will be two different concepts. The word "metacountable" belongs to the metalanguage and can be translated into English in terms of arranging symbols on a piece of paper; whereas "*b is countable*" is a definition in Set Theory. We define "*b is countable* $\leftrightarrow C(b)$" where $C(x)$ is a certain formula with free variable x in the language of Set Theory that will be made explicit later.

REMARK 8.5. There is a standard translation of the language L_{set} of Set Theory into the English language (or in Argot) as follows:
$a, b, x, \ldots$ are translated as "the set a," "the set b," "the set x,"......
$\in$ is translated as "*...belongs to the set...*" or as "*...is an element of the set...*"
$=$ is translated as "*...equals...*"
$\subset$ is translated as "*...is a subset of...*" or as "*...is contained in...*"
$\forall$ is translated as "*for all sets*"
$\exists$ is translated as "*there exists a set*"
while the connectives are translated in the standard way.

DEFINITION 8.6. x is called a *proper subset* of y if and only if x is a subset of y and x is not equal to y.

REMARK 8.7. Once we have a translation of L_{set} into English we can speak of Argot and translation of L_{set} into Argot; this simplifies comprehension of mathematical texts considerably.

REMARK 8.8. The standard translation of the language of Set Theory into English (in the remark above) is standard only by convention. A perfectly good different translation is, for instance, the one in which
$a, b, \ldots$ are translated as "*crocodile* a," "*crocodile* b,"...
$\in$ is translated as "*... is dreamt by the crocodile ...*"
$=$ is translated as "*... has the same taste as ...*"
$\forall$ is translated as "*for all crocodiles ...*"
$\exists$ is translated as "*there exists a crocodile ...*"
One could read mathematical texts in this translation; admittedly the English text that would result from this translation would be somewhat strange.

REMARK 8.9. Note that Mathematics uses other symbols as well such as

$$\leq, \circ, +, \times, \sum a_n, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{F}_p, \equiv, \lim a_n, \int f(x)dx, \frac{df}{dx}, \ldots$$

These symbols will be all sets (hence terms) and will be introduced through appropriate definitions; they will all be defined through the predicate $\in$. In particular in L_{set}, the symbols $+$ or $\times$ or $\frac{df}{dx}$ are NOT functional symbols; and $\leq$ or $\equiv$ are NOT predicates. More generally we will introduce a predicate "...*is a function*" in L_{set} but note that in a sentence of the form "f *is a function*" in L_{set} the letter f is always a set (term) and not a functional symbol. One should make a clear distinction between the predicate "*is a function*" in L_{set} and the predicate "*is a function*" in Metalanguage. It is somewhat unfortunate that the word "function" traditionally appears in both object language and Metalanguage; one needs to keep in mind (and steer away from) the possibility of confusion.

We next introduce the (specific) axioms of Set Theory.

AXIOM 8.10. (Singleton axiom)

$$\forall x \exists y((x \in y) \wedge (\forall z((z \in y) \rightarrow (z = x))).$$

The translation in Argot is that for any set x there is a set y whose only element is x.

AXIOM 8.11. (Unordered pair axiom)

$$\forall x \forall y \exists u(\forall z((z \in u) \leftrightarrow ((z = x) \vee (z = y)))).$$

In Argot the translation is that for any two sets x, y there is a set that only has them as elements.

Next is $P(x)$ is a formula in the language of sets, having a free variable x, we may introduce at any point in the theory the following:

AXIOM 8.12. (Separation axiom for $P(x)$)

$$\forall y \exists z \forall x((x \in z) \leftrightarrow ((x \in y) \wedge (P(x))).$$

The translation in Argot is that for every set y there is a set z whose elements are all the elements x of y such that $P(x)$.

EXAMPLE 8.13.

1) Taking $P(x)$ to be $x \notin x$ we may introduce at any point in the theory the axiom

$$\forall y \exists z \forall x((x \in z) \leftrightarrow ((x \in y) \wedge (x \notin x))$$

2) Taking $P(x)$ to be $\forall w(w \in x)$ we may introduce at any point in the theory the axiom

$$\forall y \exists z \forall x((x \in z) \leftrightarrow ((x \in y) \wedge (\forall w(w \in x)))$$

REMARK 8.14. So we have a recipe to produce axioms rather than ONE axiom: for each $P(x)$ one has a different axiom. Such a recipe is referred to as an *axiom scheme*. Since there are "infinitely many" ("metacountably" many) P's one allows "infinitely many" ("metacountably" many) axioms. They are introduced one by one as needed.

REMARK 8.15. One could ask if, for any given P, one can replace the separation axiom by the "modified separation axiom"

$$\exists z \forall x((x \in z) \leftrightarrow (P(x)).$$

The translation in Argot of this is that, "Given P, there is a set z whose elements are all the sets x such that $P(x)$." In spite of the fact that the latter seems quite reasonable such an axiom leads, in fact, to a contradiction. Indeed a contradiction is achieved by taking $P(x)$ to be $x \notin x$ as shown later in Exercise 9.39. This contradiction is called the *Russell paradox*. Since the "culprit" seems to be the formula "$x \notin x$" one would think that by introducing a further axiom implying that "there is no x such that $x \in x$" the Russell paradox might be avoided. We will later introduce an axiom (the Foundation Axiom 8.26 below) which indeed implies that "there is no x such that $x \in x$". However it turns out that even under that additional axiom the separation axiom cannot be replaced the "modified separation axiom" without contradiction; see Exercise 9.39. Replacing the separation axiom by the "modified separation axiom" seems to require the drastic move of changing the very syntax of our language. Such a change was indeed attempted by Russell who proposed instead his theory of types in which variables are of various types (type one: $x, y, ...$; type two $X, Y, ...$; type three $\mathcal{X}, \mathcal{Y}, ...$; etc.) with the predicate $\in$ only used as in $x \in X$, $x \in Y$, $y \in X$, $y \in Y$, $X \in \mathcal{X}$, $X \in \mathcal{Y}$, $Y \in \mathcal{X}$, $Y \in \mathcal{Y}$, etc. Russell's theory of "types" led however to difficulties related to the fact that real numbers may have different "types." Russell abandoned his approach and the theory of types was only later revived but we will ignore here this development.

The Separation Axiom has to be supplemented so as to allow "parameters" as follows. If $P(x, w)$ is a formula in the language of sets, having free variables x, w we may introduce at any point in the theory the following:

Axiom 8.16. (Separation Axiom for $P(x, w)$)

$$\forall y \forall w \exists z \forall x((x \in z) \leftrightarrow ((x \in y) \wedge (P(x, w))).$$

The translation in Argot is that for every set y and every set w there is a set z whose elements are all the elements x of y such that $P(x, w)$.

Similar axioms are allowed for formulas P with more than 2 free variables.

Axiom 8.17. (Extensionality axiom)

$$\forall u \forall v((u = v) \leftrightarrow \forall x((x \in u) \leftrightarrow (x \in v))).$$

The translation in Argot is that two sets u and v are equal if and only if they have the same elements.

Axiom 8.18. (Union axiom)

$$\forall w \exists u \forall x((x \in u) \leftrightarrow (\exists t((t \in w) \wedge (x \in t)))).$$

The translation in Argot is that for every set w there exists a set u such that for every x we have that x is an element of u if and only if x is an element of one of the elements of w. We say that u is the union of the sets in w.

Axiom 8.19. (Empty set axiom)

$$\exists x \forall y(y \notin x).$$

We introduce a predicate "is empty" via the following:

Definition 8.20. For all x, x is *empty* if and only if $\forall y(y \notin x)$.

So the translation in Argot of the Empty set axiom is: "*There exists an empty set.*"

Axiom 8.21. (Power set axiom)

$$\forall y \exists z \forall x((x \in z) \leftrightarrow (x \subset y))).$$

The translation in Argot is that for every set y there is a set z such that a set x is an element of z if and only if all elements of x are elements of y.

For simplicity the rest of the axioms will be formulated in Argot only.

Definition 8.22. Two sets are *disjoint* if and only if they have no element in common. The elements of a set are *pairwise disjoint* if and only if any two elements are disjoint.

AXIOM 8.23. (Axiom of choice) For every non-empty set w whose elements are pairwise disjoint sets there is a set that has exactly one element in common with each of the elements of w.

DEFINITION 8.24. A set x is *inductive* if and only if there is an empty set w such that $w \in x$ and for all $y \in x$ there exists $z \in x$ with the property that z is the set whose only element is y.

AXIOM 8.25. (Axiom of infinity) There exists an inductive set.

Intuitively this axiom guarantees the existence of "infinite" sets; note that we have not defined "finite/infinite" sets yet.

The following axiom is being introduced in order to disallow the existence of any x such that $x \in x$ as well as the existence of similar "exotic" situations; these consequences will be explained later.

AXIOM 8.26. (Axiom of foundation) For every non-empty set x there exists $y \in x$ such that x and y are disjoint.

One finally introduces the following axiom scheme. At any point in the theory, if $P(x, y, z)$ is a formula, one is allowed to introduce the following:

AXIOM 8.27. (Axiom of replacement for P) If for every z and every u we have that $P(x, y, z)$ "defines y as a function of $x \in u$" (i.e., for every $x \in u$ there exists a unique y such that $P(x, y, z)$) then for all z there is a set v which is the "image of this map" (i.e., v consists of all y's with the property that there is an $x \in u$ such that $P(x, y, z)$).

Similar axioms may be introduced with x, z replaced by tuples of variables.

EXERCISE 8.28. Write the axioms of choice, infinity, foundation, and replacement in the language of sets.

METADEFINITION 8.29. All of the above axioms form the ZFC system of axioms (Zermelo-Fraenkel+Choice). Set theory T_{set} is the theory in L_{set} with ZFC axioms. Unless otherwise specified all theorems in the rest of the course are understood to be theorems in T_{set}. By abuse of terminology we continue to denote by T_{set} any extension of T_{set}.

REMARK 8.30. Note the important fact that the axioms did not involve constants. In the next chapter we will show how to introduce constants.

From this moment on all proofs in this course will be written in Argot. Also, unless otherwise stated, all proofs required to be given in the exercises must be written in Argot.

CHAPTER 9

Sets

Recall that we introduced Mathematics/Set Theory as being a specific theory T_{set} in the language L_{set} with no constants and with axioms ZFC described in the last chapter. In this Chapter we will discuss some standard ways to introduce constants and functional symbols in this theory. The terms in T_{set} are also referred to as "sets"; so "set" is another word for "term" in T_{set}. Sets will be denoted by $a, b, ..., A, B, ..., \mathcal{A}, \mathcal{B}, ..., \alpha, \beta, \gamma, ...$

In what follows all definitions will be definitions in the language L_{set} of sets. Sometimes definitions are given in Argotic L_{set}.

Note that "there is only one empty set" in the sense that we have:

THEOREM 9.1. *For all x and y if x and y are empty then $x = y$.*

Proof. Let a and b be arbitrary. Assume they are empty. We want to show that $a = b$. By the extensionality axiom we have

$$(a = b) \leftrightarrow \forall x((x \in a) \leftrightarrow (x \in b)).$$

So it is enough to show that

$$\forall x((x \in a) \leftrightarrow (x \in b)).$$

Let d be arbitrary. We want to show that

$$(d \in a) \leftrightarrow (d \in b).$$

Since $\forall y(y \notin a)$ it follows that $d \notin a$. Since $\forall y(y \notin b)$ it follows that $d \notin b$. Since $d \notin a$ and $d \notin b$ it follows

$$(d \notin a) \leftrightarrow (d \notin b).$$

Hence

$$(d \in a) \leftrightarrow (d \in b).$$

$\square$

In view of the Empty set axiom and Theorem 9.1 we can introduce a new constant $\emptyset$ (called the *empty set*) via the

DEFINITION 9.2. $\forall x((x = \emptyset) \leftrightarrow (\forall y(y \notin x)))$.

Next we introduce a unary functional symbol $\{\ \}$ by the

DEFINITION 9.3. $\forall x \forall y(y = \{x\} \leftrightarrow \forall z((z \in y) \to (z = x)))$

REMARK 9.4. The definition is correct because, by extensionality axiom, the y in the singleton axiom is "unique."

EXERCISE 9.5. Make the above Remark precise.

So if a is a set (i.e. a term) then $\{a\}$ is a set (i.e., a term, because it is obtained by applying a function symbol to a term); we can say (and we will usually say, by abuse of terminology) that $\{a\}$ is "the unique" set containing a only among its elements; we will often use this kind of abuse of terminology. In particular $\{\{a\}\}$ denotes the set whose only element is the set $\{a\}$, etc.

Similarly, one introduces the binary functional predicate $\{\ ,\ \}$ via the

Definition 9.6. $\forall x \forall y \forall z(\{x,y\} = z \leftrightarrow (\forall u((u \in z) \leftrightarrow ((u = x) \vee (u = y)))))$

Again the definition is correct because, by extensionality axiom, the u in the unordered pair axiom is "unique."

So if a, b are sets then $\{a, b\}$ is the set that only has a and b as elements.

Remark 9.7. By the extensionality axiom in order to prove $A = B$ for sets A and B one needs to prove $A \subset B$ and $B \supset A$. I.e. one needs to prove:
1) If $x \in A$ then $x \in B$.
2) If $x \in B$ then $x \in A$.

Exercise 9.8. If $a = b$ then $\{a\} = \{a, b\}$.

Proposition 9.9. *If $b \neq c$ then $\{a\} \neq \{b, c\}$.*

Proof. We proceed by contradiction. So assume $A = \{a\}$, $B = \{b, c\}$, and $A = B$ and seek a contradiction. Indeed since $a \in A$ and $A = B$, by the extensionality axiom we get $a \in B$. Hence $a = b$ or $a = c$. Assume $a = b$ and seek a contradiction. (In the same way we get a contradiction by assuming $a = c$.) Since $a = b$ we get $B = \{a, c\}$. Since $c \in B$ and $A = B$, by the extensionality axiom we get $c \in A$. So $c = a$. Since $a = b$ we get $b = c$. But $b \neq c$ so we get a contradiction. $\square$

Exercise 9.10. Prove that:
1) If $\{a\} = \{b\}$ then $a = b$.
2) $\{a, b\} = \{b, a\}$.
3) There is a set b whose only elements are $\{a\}$ and $\{a, \{a\}\}$; so

$$b = \{\{a\}, \{a, \{a\}\}\}.$$

For $P(x)$ a formula in the language of sets with one free variable x we define a unary functional symbol

$$\{x \in \quad \mid P(x)\}$$

attaching to every set (term) t the set (term)

$$\{x \in t \mid P(x)\}$$

via the following

Definition 9.11.

$$\forall u \forall v((v = \{x \in u \mid P(x)\}) \leftrightarrow (\forall x((x \in v) \leftrightarrow ((x \in u) \wedge P(x)))))$$

The correctness of the definition follows from the separation and extensionality axioms.

Exercise 9.12. Explain in detail the correctness of the above definition.

Remark 9.13. So if A is a set (i.e., term) then $\{x \in A \mid P(x)\}$ is a set (i.e., term) and is translated as "the set whose elements are the elements of A satisfying property P." This set could be called the *extension of P in A* and corresponds to what in philosophical terminology is called the *extension of a concept*; the formula P itself corresponds to the *intension* of the concept.

If f is a functional symbol in T_{set} (introduced at some point in the theory) then we sometimes write

$$\{f(x) \in A \mid P(x)\}$$

instead of

$$\{y \in A \mid \exists x((f(x) = y) \wedge P(x))\}.$$

To make our definitions (notation) more reader friendly we will begin to express them in Argot as in the following example.

We introduce binary functional symbol $\cup$ as follows.

DEFINITION 9.14. For sets A and B the set $A \cup B$ (called the *union* of A and B) is the set such that for all c, $c \in A \cup B$ if and only if $c \in A$ or $c \in B$.

The above definition is a formulation in Argot of the following definition for the binary functional symbol $\cup$:

DEFINITION 9.15. $\forall x \forall y \forall z((z = x \cup y) \leftrightarrow (\forall u((u \in z) \leftrightarrow ((u \in x) \vee (u \in y)))))$

Using a similar type of formulation we introduce the binary functional symbols $\cap$ and $\backslash$ via the following:

DEFINITION 9.16. For sets A and B the *difference* between the set A and the set B is the set

$$A \backslash B = \{c \in A \mid c \notin B\}.$$

DEFINITION 9.17. For sets A and B the *intersection* of the sets A and B is the set

$$A \cap B = \{c \in A \mid c \in B\}.$$

EXERCISE 9.18. Explain in detail the correctness of the above definitions.

EXERCISE 9.19. Prove that if a, b, c are sets then there is a set (which will be denoted by $\{a, b, c\}$) whose only elements are a, b, c; in other words prove the following sentence:

$$\forall x \forall x' \forall x'' \exists y((x \in y) \wedge (x' \in y) \wedge (x'' \in y) \wedge (\forall z(z \in y) \rightarrow ((z = x) \vee (z = x') \vee (z = x''))))$$

Hint: Use the singleton axiom, the unordered pair axiom, and the union axiom, applied to the set $\{\{a\}, \{b, c\}\}$.

Similarly one defines sets $\{a, b, c, d\}$, etc.

REMARK 9.20. Inside any proof, we may introduce a new constant

$$\{a, b, c, ...\}$$

"denoting" any set that contains a, b, c (and possibly other elements); this new constant is a witness for the theorem stating that there exists a set containing a, b, c. Strictly speaking such a new constant should not be used after the end of the proof where it is used because the above definition does not pin down *all* the elements of $\{a, b, c, ...\}$. However, by abuse of notation, the latter rule is not usually enforced.

Exercise 9.21.
1) Prove that $\{\emptyset\} \neq \emptyset$.
2) Prove that $\{\{\emptyset\}\} \notin \{\emptyset, \{\emptyset\}\}$.
3) Prove that $\{\{\{\emptyset\}\}\} \notin \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$.

Exercise 9.22. Prove that:
1) $\{a, b, c\} = \{b, c, a\}$.
2) If $c \neq a$ and $c \neq b$ then $\{a, b\} \neq \{a, b, c\}$.

Exercise 9.23. Let $A = \{a, b, c\}$ and $B = \{c, d\}$ with a, b, c, d distinct (i.e., pairwise non-equal). Prove that
1) $A \cup B = \{a, b, c, d\}$,
2) $A \cap B = \{c\}$, $A \backslash B = \{a, b\}$.

Exercise 9.24. Let $A = \{a, b, c, d, e, f\}$, $B = \{d, e, f, g, h\}$ with a, b, c, d, e, f, g, h distinct. Compute
1) $A \cap B$,
2) $A \cup B$,
3) $A \backslash B$,
4) $B \backslash A$,
5) $(A \backslash B) \cup (B \backslash A)$.

Exercise 9.25. Prove the following:
1) $A \cap B \subset A$,
2) $A \subset A \cup B$,
3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
4) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
5) $(A \backslash B) \cap (B \backslash A) = \emptyset$.

Hint for $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$. Let $x \in (A \cup B) \cap (A \cup C)$; we want to show that $x \in A \cup (B \cap C)$. Since $x \in (A \cup B) \cap (A \cup C)$ we have $x \in A \cup B$ and $x \in A \cup C$. We have two cases. The first case is $x \in A$. In this case $x \in A \cup (B \cap C)$ and we are done. The second case is $x \notin A$. In this case, since $x \in A \cup B$ and $x \notin A$, we get $x \in B$; and similarly since $x \in A \cup C$ and $x \notin A$ we get $x \in C$. Since $x \in B$ and $x \in C$ we get $x \in B \cap C$ hence $x \in A \cup (B \cap C)$ and we are done again.

Next we introduce the unary functional symbol $\mathcal{P}$ via the following:

Definition 9.26. For a set A the set $\mathcal{P}(A)$ is the set whose elements are the subsets of A; we call $\mathcal{P}(A)$ the *power set* of A.

Exercise 9.27. Explain in detail the correctness of this definition.

Example 9.28. If $A = \{a, b, c\}$ with a, b, c distinct then

$$\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

with all listed subsets distinct.

Exercise 9.29. Let $A = \{a, b, c, d\}$. Write down the set $\mathcal{P}(A)$.

Exercise 9.30. Let $A = \{a, b\}$. Write down the set $\mathcal{P}(\mathcal{P}(A))$.

Next we introduce the binary functional symbol $(\ , \ )$ via the following:

Definition 9.31. (Ordered pairs) For sets a and b the *ordered pair* (a, b) is the set $\{\{a\}, \{a, b\}\}$.

We sometimes say “pair” instead of “ordered pair.”

Note that $(a, b) \in \mathcal{P}(\mathcal{P}(\{a, b\}))$. Also note that if $a = b$ then $(a, b) = \{\{a\}\}$.

We next introduce the binary functional symbol $\times$ by the following:

Definition 9.32. For sets A and B we define the *product* of A and B as the set $A \times B$ whose elements are exactly the ordered pairs with first element in A and second element in B. In other words,

$$A \times B = \{z \in \mathcal{P}(\mathcal{P}(A \cup B)) \mid \exists x \exists y((x \in A) \wedge (y \in B) \wedge (z = (x, y)))\}.$$

Clearly $\times$ defines a binary functional symbol.

Proposition 9.33. *$(a, b) = (c, d)$ if and only if $a = c$ and $b = d$.*

Proof. We need to prove that

1) If $a = c$ and $b = d$ then $(a, b) = (c, d)$ and

2) If $(a, b) = (c, d)$ then $a = c$ and $b = d$.

Now 1) is obvious. To prove 2) assume $(a, b) = (c, d)$.

Assume first $a \neq b$ and $c \neq d$. Then by the definition of pairs we know that

$$\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}.$$

Since $\{a\} \in \{\{a\}, \{a, b\}\}$ it follows (by the extensionality axiom) that $\{a\} \in \{\{c\}, \{c, d\}\}$. Hence either $\{a\} = \{c\}$ or $\{a\} = \{c, d\}$. But as seen before $\{a\} \neq \{c, d\}$. So $\{a\} = \{c\}$. Since $a \in \{a\}$ it follows that $a \in \{c\}$ hence $a = c$. Similarly since $\{a, b\} \in \{\{a\}, \{a, b\}\}$ we get $\{a, b\} \in \{\{c\}, \{c, d\}\}$. So either $\{a, b\} = \{c\}$ or $\{a, b\} = \{c, d\}$. Again as seen before $\{a, b\} \neq \{c\}$ so $\{a, b\} = \{c, d\}$. So $b \in \{c, d\}$. So $b = c$ or $b = d$. Since $a \neq b$ and $a = c$ we get $b \neq c$. Hence $b = d$ and we are done in case $a \neq b$ and $c \neq d$.

Assume next $a = b$ and $c = d$. Then by the definition of pairs in this case we have $\{\{a\}\} = \{\{c\}\}$ and as before this implies $\{a\} = \{c\}$ hence $a = c$ so we are done in this case as well.

Finally assume $a = b$ and $c \neq d$. (The case $a \neq b$ and $c = d$ is treated similarly.) By the definition of pairs we get

$$\{\{a\}\} = \{\{c\}, \{c, d\}\}.$$

We get $\{c, d\} \in \{\{a\}\}$. Hence $\{c, d\} = \{a\}$ which is impossible, as seen before. This ends the proof. □

Exercise 9.34. Prove that

1) $(A \cap B) \times C = (A \times C) \cap (B \times C)$,

2) $(A \cup B) \times C = (A \times C) \cup (B \times C)$.

Hint for $\subset$ in 1): Let $x \in (A \cap B) \times C$. Then $x = (y, z)$ with $y \in A \cap B$ and $z \in C$. Since $y \in A \cap B$ we have $y \in A$ and $y \in B$. Since $y \in A$ and $z \in C$ we get $(y, z) \in A \times C$. Since $y \in B$ and $z \in C$ we get $(y, z) \in B \times C$. Since $(y, z) \in A \times C$ and $(y, z) \in B \times C$ it follows that $x = (y, z) \in (A \times C) \cap (B \times C)$.

Exercise 9.35. Prove that

$$\forall x(x \notin x).$$

In Argot this says that no set can be an element of itself.

Hint: Let c be a set such that $c \in c$ and seek a contradiction. Let $a = \{c\}$. Then a is non-empty so by the Axiom of foundation c and a are disjoint. But $c \cap a = c \cap \{c\} = \{c\} \neq \emptyset$, a contradiction.

EXERCISE 9.36. Prove that

$$\forall x \forall y((x \notin y) \vee (y \notin x))$$

In Argot this says that for every two sets one of them is not an element of the other.

Hint: Assume the contrary, let a, b be such that $a \in b$ and $b \in a$ and apply the Axiom of foundation to the set $\{a, b\}$.

EXERCISE 9.37. Prove that

$$\neg(\exists y \forall z(z \in y))$$

In Argot this says that there does not exist a set T such that for every set A we have $A \in T$. (Intuitively there is no set such that all sets belong to it.)

Hint: Assume there is such a T and derive a contradiction. Hence $\forall z(z \in T)$. Hence (by dropping $\forall z$ and replacing z by T) we have $T \in T$ which contradicts Exercise 9.35.

EXERCISE 9.38. Prove that even if we remove the Axiom of foundation from ZFC the following is still a Theorem:

$$\neg(\exists y \forall z(z \in y))$$

Hint: Assume there is such a T and derive a contradiction. So $\forall z(z \in T)$. Consider the set

$$S = \{x \in T \mid x \notin x\}.$$

There are two cases. First case is $S \notin S$. From $\forall z(z \in T)$ it follows that $S \in T$. Since $S \in T$ and $S \notin S$, we get that $S \in \{x \in T \mid x \notin x\}$ hence $S \in S$, a contradiction. The second case is $S \in S$. We get that $S \notin \{x \in T \mid x \notin x\}$ hence $S \notin S$, which is again a contradiction.

EXERCISE 9.39. Show that if for each formula $P(x)$ one replaces the Separation Axiom by the "Modified Separation Axiom"

$$\exists z \forall x((x \in z) \leftrightarrow P(x))$$

then one can derive a contradiction even if one removes from ZFC the Axiom of foundation. (This is called the *Russell paradox*. Roughly speaking it arises from allowing sets of the form $\{x \mid P(x)\}$ rather than sets of the form $\{x \in y \mid P(x)\}$.)

Hint: Take $P(x) =$"$x \notin x$" hence by the Modified Separation Axiom we have $\exists z \forall x((x \in z) \leftrightarrow (x \notin x))$. Let S be a witness of this sentence. So (dropping $\exists z$ and replacing z by S) we have

$$\forall x((x \in S) \leftrightarrow (x \notin x)).$$

So (dropping $\forall x$ and replacing x by S) we have

$$(S \in S) \leftrightarrow (S \notin S).$$

This is a contradiction (as one can see by taking A to be $S \in S$ and using a truth table for the sentence $A \leftrightarrow \neg A$.)

REMARK 9.40. Before the advent of ZFC Russell showed that Cantor's Set Theory leads to a contradiction, cf. the "Russell paradox" discussed earlier. Within ZFC Russell's paradox, in its original form, disappears. Whether there are other forms of this paradox, or similar paradoxes, that survive in ZFC it is not known.

Remark 9.41. One can ask if dropping the Axiom of foundation from ZFC leads to an interesting theory. The answer is YES, although we will not investigate this here. Such a theory has applications in computer science. Also it is interesting to follow a suggestion of Badiou according to which if one applies the ideas of Set Theory to matters of political philosophy then the Axiom of foundation appears to be violated. The example Badiou gives is the thing called "*French Revolution*" which as a set contains elements such as "*Robespierre*," "*The Fall of Bastille*," etc. but it also contains as an element the concept of "*French Revolution*" itself; indeed the concept of "*French Revolution*" cannot be considered complete (the Revolution has not taken place) unless it is made to incorporate itself as an element (i.e., unless the actors of the Revolution perceive the events as a Revolution).

We end by recording the following:

Definition 9.42. A *correspondence* between a set A and a set B is a subset $U \subset A \times B$.

Definition 9.43. (Composition). Assume U is a correspondence between A and B and assume V is a correspondence between B and C. Then one defines a correspondence $V \circ U$ (called the *composition* of V with U) between A and C by

$$V \circ U = \{(x, z) \in A \times C \mid \exists y((y \in B) \wedge ((x, y) \in U) \wedge ((y, z) \in V))\}.$$

Definition 9.44. (Transpose). Assume U is a correspondence between A and B. Then one defines the *transpose* U^t of U as the correspodence

$$U^t = \{(y, x) \in B \times A \mid (x, y) \in U\}.$$

Exercise 9.45. Prove that
1) $U^{tt} = U$,
2) $(V \circ U)^t = U^t \circ V^t$,
3) $W \circ (V \circ U) = (W \circ V) \circ U$.

In the next two Chapters we will study two types of correspondences: maps and relations, respectively.

CHAPTER 10

Maps

The concept of map (or function) has a long history. Originally functions were understood to be given by more or less explicit "formulae" (polynomial, rational, algebraic, and later by series). Controversies around what the "most general" functions should be arose, for instance, in connection with solving partial differential equations (by means of trigonometric series); this is somewhat parallel to the controversy around what the "most general" numbers should be that arose in connection with solving algebraic equations (such as $x^2 = 2$, $x^2 = -1$, or higher degree equations with no solutions expressed by radicals, etc.). The notion of "completely arbitrary" function gradually arose through the work of Dirichlet, Riemann, Weierstrass, Cantor, etc. Here is the definition (in Argot):

Definition 10.1. A *map* (or *function*) from a set A to a set B is a subset $F \subset A \times B$ such that for every $x \in A$ there is a unique $y \in B$ with $(x, y) \in F$.

Remark 10.2. The above Definition is written in Argot. One can paraphrase it as follows: for all u, v, w we have that w is a map from u to v if and only if $w \subset u \times v$ and for every x if $x \in u$ then there exists a unique $y \in v$ such that $(x, y) \in w$. Equivalently: for all u, v, w we have that w is a map from u to v if and only if $w \subset u \times v$ and the following hold:

1) For every x if $x \in u$ then there exists $y \in v$ such that $(x, y) \in w$.

2) For all x', x'', y', y'' if $(x', y'), (x'', y'') \in w$ and $x' = x''$ then $y' = y''$.

To formalize Definition 10.1 one introduces a new ternary predicate $fun(w, u, v)$ (which in Argot is translated as "*w is a map (or function) from u to v*") via the following Definition:

$$\forall w \forall u \forall v (fun(w, u, v) \leftrightarrow (((w \subset u \times v) \wedge (\forall x((x \in u) \rightarrow \exists ! y((x, y) \in w)))))).$$

As noted before one needs to make a clear distinction between the predicate "*...is a map (or function)... between ... and ...*" in L_{set} and the predicate "*... is a functional symbol (or function)* in L_{set}" in Metalanguage. For instance

"$\exp, \sin, +, \times$ *are functions*"

is a sentence in L_{set} while

"$\{\ ,\ \}$," "$\mathcal{P}$," "$\{x \in \quad \mid P(x)\}$" *are functions in L_{set}*

is a metasentence in Metalanguage.

Remark 10.3. Note that every map is a correspondence. Note also that what we call a *map* $F \subset A \times B$ corresponds to what in elementary Mathematics is called the *graph of a map*.

Definition 10.4. If A and B are sets we denote by $Fun(A, B) \subset \mathcal{P}(A \times B)$ the set of all maps $F : A \rightarrow B$.

Remark 10.5. Definition 10.4 is, again, given in Argot. To formalize it one introduces a binary functional symbol "Fun" (where one translates $Fun(u,v)$ as "the set of all maps from u to v") via the following Definition:

$$\forall u \forall v \forall z(((Fun(u,v)=z) \leftrightarrow (\forall w((w \in z) \leftrightarrow fun(w,u,v))))).$$

Remark 10.6. We introduce the following rule in Argot. If "$F \in Fun(A,B)$" and "$(a,b) \in F$" are accepted sentences in T_{set} or Argot (with F, A, B, a, b constants in L_{set}) then one can write in Argot: $F(a) = b$ or $a \mapsto F(a)$ or $F : a \mapsto b$. When F, a, b are constants one can even introduce the symbol $F(a)$ in T_{set} as a new constant via the Definition $F(a) := b$.

However, note that if x is a variable and F, A, B are constants with "$F \in Fun(A,B)$" accepted in T_{set} then the symbol $F(x)$ only makes sense in Argot and not in L_{set} because F is not a functional symbol in L_{set}. Furthermore one cannot introduce in L_{set} a new functional symbol (say, still denoted by F) via the Definition

$$\forall x \forall y((F(x) = y) \leftrightarrow ((x,y) \in F))$$

because after replacing x by A and y by $F(A)$ we get

$$(F(A) = F(A)) \to ((A, F(A)) \in F)$$

which implies $A \in A$, a contradiction! The situation is even worse. Indeed, one often quantifies over functions as when one says in Argot that "*for all functions F if there exists x such that $P(F(x))$ then $Q(F)$.*" If one is to formalize the latter one needs to use a variable y for F but one cannot use the symbol $P(y(x))$ in L_{set} because $y(x)$ is syntactically incorrect in L_{set}. One can attempt to introduce a binary functional symbol in L_{set} whose value for x and y is $x(y)$ but it is easy to see that the attempt leads to a contradiction as before. The moral of these remarks is that every time the symbol $F(\ )$ is being used in an Argot sentence one needs to be able, in principle, to paraphrase (and formalize) the sentence in such a way that the symbol $F(\ )$ disappears. We will not give a complete set of rules for doing this (although this can be done). Rather we will give a few examples here and more examples later. As an example if $P(x)$ is a formula with one free variable x and a is a constant then the Argot sentence

"$P(F(a))$"

is paraphrased as

"*For all y if $(a,y) \in F$ then $P(y)$*"

and formalized as

"$\forall y((a,y) \in F) \to P(y))$."

For a more complicated example, if $P(u)$ is a formula with one free variable u and $Q(u,v)$ is a formula with free variables u, v the Argot sentence

"*For all sets A such that $P(A)$ there exists $F \in Fun(A,A)$ such that for all $x \in A$ we have $Q(F(x), F(F(x)))$*"

is paraphrased as

"*For all u if $P(u)$ then there exists w such that $w \in Fun(u,u)$ and such that for all x, y, z if $x \in u$ and $(x,y) \in w$ and $(y,z) \in w$ then $Q(y,z)$*"

and formalized accordingly (left to the reader).

EXAMPLE 10.7. If a, b, c are distinct the set

$$F = \{(a, a), (b, c)\} \subset \{a, b\} \times \{a, b, c\} \tag{10.1}$$

is a map and $F(a) = a$, $F(b) = c$. On the other hand the subset

$$F = \{(a, b), (a, c)\} \subset \{a, b\} \times \{a, b, c\}$$

is not a map.

DEFINITION 10.8. A map $F : A \to B$ is *injective* (or an *injection*, or *one-to-one*) if and only if for all $x', x'' \in A$ we have that $F(x') = F(x'')$ implies $x' = x''$.

DEFINITION 10.9. A map $F : A \to B$ is *surjective* (or a *surjection*, or *onto*) if and only if for every $y \in B$ there exists $x \in A$ such that $F(x) = y$.

REMARK 10.10. As usual, Definitions 10.8 and 10.9 were given in Argot. To formalize them one can first paraphrase them without the use of the symbol $F(\ )$ as follows:

1) For all w, u, v we have that w is an injective map from u to v if and only w is a map from u to v and for all $x'x'', y', y''$ if $(x', y'), (x'', y'') \in w$ and $y' = y''$ then $x' = x''$.

2) For all w, u, v we have that w us a surjective map from u to v if and only if w is a map from u to v and for every $y \in v$ there exists $x \in u$ such that $(x, y) \in w$.

To formalize Definition 10.8 one introduces a new ternary predicate denoted by "*inj(w,u,v)*" (and translated as "*w is an injective map from u to v*") via an obvious Definition (which we leave to the reader). A similar comment holds for Definition 10.9.

REMARK 10.11. Most Definitions and Theorems that follow are given in Argot and need, in principle, be formalized the way we did above. For the sake of simplicity and readability we will generally avoid doing this from now on. It will be sufficient for us to know that this formalization is possible along the lines already followed in the discussion above.

EXAMPLE 10.12. The map (10.1) is injective and not surjective.

EXERCISE 10.13. Give an example of a map which is surjective and not injective.

DEFINITION 10.14. (Identity map). For every A the *identity* map $I : A \to A$ is defined as $I(x) = x$, i.e.,

$$I = I_A = \{(x, x) \in A \times A \mid x \in A\} \subset A \times A.$$

EXERCISE 10.15. (Inclusion map). Let $A \subset B$. Prove that there is an injective map $i : A \to B$ such that $i(x) = x$ for all $x \in A$. We call i the *inclusion* map; we sometimes say $A \subset B$ is the inclusion map.

EXERCISE 10.16. (Composition) Prove that if $F : A \to B$ and $G : B \to C$ are two maps then the composition of the correspondences G and F, defined by

$$G \circ F = \{(x, z) \in A \times C \mid \exists y((y \in B) \wedge ((x, y) \in U) \wedge ((y, z) \in V))\} \subset A \times C,$$

is a map from A to C and satisfies $(G \circ F)(x) = G(F(x))$ for all $x \in A$.

Definition 10.17. (Restriction) If $F : A \to B$ is a map and $A' \subset A$ then the composition of F with the inclusion map $A' \subset A$ is called the *restriction* of F to A' and is denoted by $F_{|A'} : A' \to B$.

Definition 10.18. (Corestriction) If $F : A \to B$ is a map and $B' \subset B$ is such that for all $x \in A$ we have $F(x) \in B'$ then the map $F' : A \to B'$ defined by $F'(x) = F(x)$ is called the *corestriction* of F to B'.

Definition 10.19. (Commutative diagram) By a *commutative diagram* of sets

$$\begin{array}{ccc} A & \stackrel{F}{\to} & B \\ U\downarrow & & \downarrow V \\ C & \stackrel{G}{\to} & D \end{array}$$

we mean a collection of sets and maps as above with the property that $G \circ U = V \circ F$.

Exercise 10.20. Prove that if $F \circ G$ is surjective then F is surjective. Prove that if $F \circ G$ is injective then G is injective.

Exercise 10.21. Prove that the composition of two injective maps is injective and the composition of two surjective maps is surjective.

Exercise 10.22. Let $F : A \to B$ be an injective map.
1) Prove that the corestriction $F' : A \to F(A)$ of F is a bijection.
2) Prove that there exists a map $G : B \to A$ such that $G \circ F = I_A$. (We call G a *left inverse* or a *retraction* of F.)

Hint for 2): Let $(F')^{-1} : F(A) \to A$ be the inverse of F' and fix $a \in A$. Then define $G(b) = (F')^{-1}(b)$ for $b \in F(A)$ and $G(b) = a$ for all $b \in B \setminus F(A)$.

Exercise 10.23. Let $F : A \to B$ be a surjective map. Prove that there exists a map $G : B \to A$ such that $F \circ G = I_B$. (We call G a *right inverse*, or a *section*, of F.)

Hint: For $b \in B$ let $A_b := \{a \in A \mid F(a) = b\}$ and use the Axiom of choice applied to the set $\{A_b \in \mathcal{P}(A) \mid b \in B\}$.

Definition 10.24. A map is *bijective* (or a *bijection*) if and only if it is injective and surjective.

Here is a fundamental theorem in Set Theory; it was conjectured by Cantor and proved a little later by Bernstein.

Theorem 10.25. *(Bernstein's Theorem) If A and B are sets and if there exist injective maps $F : A \to B$ and $G : B \to A$ then there exists a bijective map $U : A \to B$.*

The proof will have to wait until we get to the chapter on Sequences; we will not use this theorem until it is proved.

Exercise 10.26. Prove that if $F : A \to B$ is bijective then there exists a unique bijective map denoted by $F^{-1} : B \to A$ such that $F \circ F^{-1} = I_B$ and $F^{-1} \circ F = I_A$. F^{-1} is called the *inverse* of F.

Hint: Let F^{-1} be the transpose F^t of F (where we view F as a correspondence).

Exercise 10.27. Let a, b, c, d, e be distinct. Let $F : \{a, b, c\} \to \{c, d, e\}$, $F(a) = d$, $F(b) = c$, $F(c) = e$. Prove that F has an inverse and compute F^{-1}.

EXERCISE 10.28. Prove that if A and B are sets then there exist maps $F : A \times B \to A$ and $G : A \times B \to B$ such that $F(a,b) = a$ and $G(a,b) = b$ for all $(a,b) \in A \times B$. (These are called the *first and the second projection.*)

Hint: For G show that $G = \{((a,b),c); c = b\} \subset (A \times B) \times B$ is a map.

EXERCISE 10.29. Prove that $(A\times B)\times C \to A\times(B\times C)$, $((a,b),c) \mapsto (a,(b,c))$ is a bijection.

DEFINITION 10.30. Write $A \times B \times C$ instead of $(A \times B) \times C$ and write (a,b,c) instead of $((a,b),c)$. We call (a,b,c) a *triple.* Write $A^2 = A\times A$ and $A^3 = A\times A\times A$. More generally adopt this notation for arbitrary number of factors. Elements like $(a,b),(a,b,c),(a,b,c,d)$, etc. will be called *tuples.*

THEOREM 10.31. *If A is a set then there is no bijection between A and $\mathcal{P}(A)$*

Proof. Assume there exists a bijection $F : A \to \mathcal{P}(A)$ and seek a contradiction. Consider the set

$$B = \{a \in A \mid a \notin F(a)\} \in \mathcal{P}(A).$$

Since F is surjective there exists $b \in A$ such that $B = F(b)$. There are two cases: either $b \in B$ or $b \notin B$. If $b \in B$ then $b \in F(b)$ so $b \notin B$, a contradiction. If $b \notin B$ then $b \notin F(b)$ so $b \in B$, a contradiction, and we are done. $\square$

REMARK 10.32. Note the similarity between the above argument and the argument showing that there is no set having all sets as elements (the "Russell paradox"). The above theorem is one of the main discoveries of Cantor: it is the basis for his "creating" his whole "hierarchy of infinities."

DEFINITION 10.33. Let S be a set of sets and I a set. A *family* of sets in S indexed by I is a map $I \to S$, $i \mapsto A_i$. We sometimes drop the reference to S. We also write $(A_i)_{i\in I}$ to denote this family. By the union axiom for every such family there is a set (denoted by $\bigcup_{i\in I} A_i$, called their *union*) such that for all x we have that $x \in \bigcup_{i\in I} A_i$ if and only if there exists $i \in I$ such that $x \in A_i$. Also, for $I \neq \emptyset$, a set (denoted by $\bigcap_{i\in I} A_i$, called the *intersection* of the family) exists such that for all x we have that $x \in \bigcap_{i\in I} A_i$ if and only if for all $i \in I$ we have $x \in A_i$. A family of elements in $(A_i)_{i\in I}$ is a map $I \to \bigcup_{i\in I} A_i$, $i \mapsto a_i$, such that for all $i \in I$ we have $a_i \in A_i$. Such a family of elements is denoted by $(a_i)_{i\in I}$. One defines the *product* $\prod_{i\in I} A_i$ as the set of all families of elements $(a_i)_{i\in I}$.

EXERCISE 10.34. Check that for $I = \{i,j\}$ the above definitions of $\cup, \cap, \prod$ yield the usual definition of $A_i \cup A_j$, $A_i \cap A_j$, and $A_i \times A_j$.

DEFINITION 10.35. Let $(A_i)_{i\in I}$ be a family of sets. The *disjoint union* $\coprod_{i\in I} A_i$ of this family is defined by

$$\coprod_{i\in I} A_i = \{(x,i) \in (\bigcup_{i\in I} A_i) \times I \mid x \in A_i\}.$$

There are natural injective maps $\epsilon_j : A_j \to \coprod_{i\in I} A_i$, $\epsilon_j(x) = (x,j)$. For $I = \{1,2\}$ with $1,2$ two distinct sets we write

$$\coprod_{i\in I} A_i = A_1 \coprod A_2$$

EXAMPLE 10.36. For $I = \{1,2\}$ and $A_1 = \{a,b\}$, $A_2 = \{b,c,d\}$ we have

$$A_1 \coprod A_2 = \{(a,1),(b,1),(b,2),(c,2),(d,2)\}.$$

Definition 10.37. Let $F : A \to B$ be a map and $X \subset A$. Define the *image* of X as the set

$$F(X) = \{y \in B \mid \ \exists x \in X,\ y = F(x)\} \subset B.$$

One often writes, by abuse of notation

$$F(X) = \{F(x) \mid \ x \in X\} \subset B.$$

If $Y \subset B$ define the *inverse image* (or *preimage*) of Y as the set

$$F^{-1}(Y) := \{x \in A \mid F(x) \in Y\} \subset A.$$

For $y \in B$ one often writes, by abuse of notation,

$$F^{-1}(y) := F^{-1}(\{y\}) = \{x \in A \mid F(x) = y\}.$$

Call $F^{-1}(y)$ the *fiber* of F at y. (Note that $F^{-1}(Y)$, $F^{-1}(y)$ are defined even if the inverse map F^{-1} does not exist, i.e., even if F is not bijective.)

Exercise 10.38. Let $F : \{a, b, c, d, e, f, g\} \to \{c, d, e, h\}$, $F(a) = d$, $F(b) = c$, $F(c) = e$, $F(d) = c$, $F(e) = d$, $F(f) = c$, $F(g) = c$. Let $X = \{a, b, c\}$, $Y = \{c, h\}$, all letters being distinct. Compute $F(X)$, $F^{-1}(Y), F^{-1}(c), F^{-1}(h)$.

Exercise 10.39. Prove that if $F : A \to B$ is a map and $X \subset X' \subset A$ are subsets then $F(X) \subset F(X')$.

Exercise 10.40. Prove that if $F : A \to B$ is a map and $(X_i)_{i\in I}$ is a family of subsets of A then

$$F(\cup_{i\in I} X_i) = \cup_{i\in I} F(X_i),$$
$$F(\cap_{i\in I} X_i) \subset \cap_{i\in I} F(X_i).$$

If in addition F is injective show that

$$F(\cap_{i\in I} X_i) = \cap_{i\in I} F(X_i).$$

Give an example showing that the latter may fail if F is not injective.

Exercise 10.41. Prove that if $F : A \to B$ is a map and $Y \subset Y' \subset B$ are subsets then $F^{-1}(Y) \subset F^{-1}(Y')$.

Exercise 10.42. Prove that if $F : A \to B$ is a map and $(Y_i)_{i\in I}$ is a family of subsets of B then

$$F^{-1}(\cup_{i\in I} Y_i) = \cup_{i\in I} F^{-1}(Y_i),$$
$$F^{-1}(\cap_{i\in I} Y_i) = \cap_{i\in I} F^{-1}(Y_i).$$

(So here one does not need injectivity like in the case of unions.)

Exercise 10.43. Let $F : A \to B$ be a surjective map and for every $b \in B$ let $A_b = F^{-1}(b)$. Let $(A_b)_{b\in B}$ be the family defined by the map $B \to \mathcal{P}(A)$, $b \mapsto A_b$. Prove that there is a bijection

$$A \to \coprod_{b\in B} A_b.$$

Hint. Consider the map $A \to \coprod_{b\in B} A_b$ sending every $a \in A$ into the pair $(a, F(a))$.

Exercise 10.44. Let $0, 1$ be two sets. Prove that the map

$$Fun(A, \{0, 1\}) \to \mathcal{P}(A)$$

sending $F : A \to \{0, 1\}$ into $F^{-1}(1) \in \mathcal{P}(A)$ is a bijection.

EXERCISE 10.45. Find a bijection

$$Fun(A, Fun(B, C)) \to Fun(A \times B, C).$$

Hint: Send $F \in Fun(A, Fun(B, C))$, $F : A \to Fun(B, C)$, into the set (map)

$$\{((a, b), c) \in (A \times B) \times C \mid (b, c) \in F(a)\}.$$

EXERCISE 10.46. Prove that the map

$$Fun(A, B \times C) \mapsto Fun(A, B) \times Fun(A, C)$$

defined by

$$F \mapsto (F_1, F_2)$$

where $F(x) = (F_1(x), F_2(x))$ for $x \in A$, is a bijection. (F_1 and F_2 are referred to as the *components* of F.) Generalize this to more that two sets in the product.

The following will be useful later when we introduce the integers.

DEFINITION 10.47. A Peano *triple* is a triple $(N, 1, \sigma)$ where N is a set, $1 \in N$, and $\sigma : N \to N$ is a map such that
1) σ is injective;
2) $\sigma(N) = N \backslash \{1\}$;
3) for every subset $S \subset N$ if $1 \in S$ and $\sigma(S) \subset S$ then $S = N$.

EXERCISE 10.48. Prove that there exists a Peano triple.

Hint: By the Axiom of infinity there exists an inductive set; then N can be taken to be the intersection of all inductive sets contained in that set (which turns out to be itself an inductive set) and one can take σ to be given by $\sigma(x) = \{x\}$. Injectivity of σ follows because if $\{x\} = \{y\}$ then $x = y$.

CHAPTER 11

Relations

A basic notion in Set Theory is that of relation; we shall investigate in some detail two special cases: order relations and equivalence relations.

Definition 11.1. If A is a set then a *relation* on A is a subset $R \subset A \times A$. For all $x, y \in A$ we write xRy if and only if $(x, y) \in R$.

So a relation on A is simply a correspondence between A and A.

Remark 11.2. The above Definition is, again, written in Argot. To formalize it one can define a new binary predicate $rel(z, u)$ translated as "*z is a relation on u*" in the usual way. Also one can introduce a 4-ary predicate $related(x, y, z, u)$ translated as "*x is related to y via the relation z on u*" via the Definition

$$\forall x \forall y \forall z \forall u (related(x, y, z, u) \leftrightarrow ((x \in u) \wedge (y \in u) \wedge (rel(z, u)) \wedge ((x, y) \in z))).$$

Rather than using the formalized version of Definition 11.1 we will use, from now on, its formulation in Argot. In particular we note that the expression "xRy" belongs to Argot and not to the language L_{set} of Set Theory: the sign R is not a predicate in L_{set}. Every time the expression "xRy" appears in an Argot sentence one understands that the formalization of the sentence is obtainable by replacing "xRy" with "$(x, y) \in R$."

Definition 11.3. Let $R \subset A \times A$ be a relation on A and write $a \leq b$ instead of aRb. R is called an *order relation* if and only if for all $x, y, z \in A$, we have:
1) $x \leq x$ (reflexivity),
2) $x \leq y$ and $y \leq z$ imply $x \leq z$ (transitivity),
3) $x \leq y$ and $y \leq x$ imply $x = y$ (antisymmetry).

Definition 11.4. One introduces a new relation $<$ by the condition that $x < y$ iff $x \leq y$ and $x \neq y$. The relation $<$ is not an order.

The use of the word "order" here is unrelated to the use of the same word in "first order logic."

Exercise 11.5. Let R be a relation on A (which we view as a correspondence). Show that
i) The reflexivity condition is equivalent to $I \subset R$.
ii) The transitivity condition is equivalent to $R \circ R \subset R$.
iii) The antisymmetry condition is equivalent to $R \cap R^t \subset I$.

Definition 11.6. An order relation is called a *total order* (or *linear order*) if and only if for every $x, y \in A$ either $x \leq y$ or $y \leq x$. Alternatively we say A is *totally ordered* (by $\leq$).

EXAMPLE 11.7. For instance if $A = \{a, b, c, d\}$ with all letters distinct then

$$R = \{(a, a), (b, b), (c, c), (d, d), (a, b), (b, c), (a, c)\}$$

is an order but not a total order.

EXERCISE 11.8. Let $R_0 \subset A \times A$ be a relation and assume R_0 is contained in an order relation $R_1 \subset A \times A$. Let

$$R = \bigcap_{R' \supset R_0} R'$$

be the intersection of all order relations R' containing R_0. Prove that R is an order relation and it is the smallest order relation containing R_0 in the sense that it is contained in any order relation that contains R_0.

EXERCISE 11.9. Let $A = \{a, b, c, d, e\}$ and $R_0 = \{(a, b), (b, c), (c, d), (c, e)\}$ with all letters distinct. Find an order relation containing R_0. Find the smallest order relation R containing R_0. Show that R is not a total order.

EXERCISE 11.10. Let A be a set. For every subsets $X \subset A$ and $Y \subset A$ write $X \leq Y$ if and only if $X \subset Y$. This defines a relation on the set $\mathcal{P}(A)$. Prove that this is an order relation. Give an example showing that this is not in general a total order.

DEFINITION 11.11. An *ordered set* is a pair $(A, \leq)$ where A is a set and $\leq$ is an order relation on A.

DEFINITION 11.12. Let $(A, \leq)$ and $(A', \leq')$ be ordered sets. A map $F : A \to A'$ is called *increasing* if for every $a, b \in A$ with $a \leq b$ we have $F(a) \leq' F(b)$. It is called *strictly increasing* if for every $a, b \in A$ with $a < b$ we have $F(a) <' F(b)$. One defines *decreasing* and *strictly decreasing* in a similar way.

EXERCISE 11.13. Prove that if $(A, \leq)$, $(A', \leq')$, $(A'', \leq'')$ are ordered sets and $G : A \to A'$, $F : A' \to A''$ are increasing then $F \circ G : A \to A''$ is increasing. And similarly with decreasing functions.

DEFINITION 11.14. Let A be a set with an order $\leq$ and let $B \subset A$. (An important special case of this is $B = A$.)

We say $\beta \in B$ is a *minimal* element of B if and only if for all $b \in B$ such that $b \leq \beta$ we must have $b = \beta$. We stress that if minimal elements of B exist then they belong to B and need not be unique.

We say $m \in B$ is a *minimum* element of B if and only if for all $b \in B$ we have $m \leq b$. If a minimum element exists it is unique (check!) and we denote it by $\min B$. We stress that if $\min B$ exists then, by definition, $\min B$ belongs to B.

We say $\omega \in B$ is a *maximal* element of B if and only if for all $b \in B$ such that $\omega \leq b$ we must have $\beta = b$. We stress that if maximal elements of B exist then they belong to B and need not be unique.

We say $M \in B$ is a *maximum* element of B if and only if for all $b \in B$ we have $b \leq M$. If a maximum element exists it is unique and we denote it by $\max B$. We stress that if $\max B$ exists then by definition it belongs to B. Note that minimal, minimum, maximal, and maximum elements of B depend only on B and not on A

An element $u \in A$ is called an *upper bound* for B in A if and only if $b \leq u$ for all $b \in B$. We also say that B is bounded from above by u.

An element $l \in A$ is called a *lower bound* for B in A if and only if $l \leq b$ for all $b \in B$; we also say B is *bounded from below* by l.

We say B is *bounded* (in A) if and only if it has an upper bound and a lower bound in A.

Let $U_A(B)$ be the set of upper bounds of B in A; if $U_A(B)$ has a minimum element we call it the *supremum* of B in A and we denote it by $\sup_A B$.

Let $L_A(B)$ be the set of lower bounds of B in A; if $L_A(B)$ has a maximum element we call it the *infimum* of B in A and we denote it by $\inf_A B$.

Note that if one of $\sup_A B$ and $\inf_A B$ exists that element is by definition in A, it generally depends on A (and not only on B) and does not necessarily belong to B.

We say B is *bounded* in A if and only if it has both an upper bound and a lower bound in A; this concept also depends on A.

When A is understood from context it is usually dropped from notation and no reference is made to A.

EXERCISE 11.15. Consider the set A and the order $\leq$ defined by the relation R in Exercise 38.11. Does A have a maximum element? Does A have a minimum element? Are there maximal elements in A? Are there minimal elements in A? List all these elements in case they exist. Let $B = \{b, c\}$. Is B bounded? Find the set of upper bounds of B. Find the set of lower bounds of B. Does the supremum of B exist? If yes does it belong to B? Does the infimum of B exist? Does it belong to B?

EXERCISE 11.16. Let $A = \{a, b, c, d, e, f\}$ and R the smallest order on A containing the set

$$\{(a, b), (b, c), (b, d), (c, e), (c, f), (d, e), (d, f)\}$$

Does A have a maximum element? Does A have a minimum element? Are there maximal elements in A? Are there minimal elements in A? List all these elements in case they exist. Let $B = \{a, b, c, d\}$. Is B bounded? Find the set of upper bounds of B. Find the set of lower bounds of B. Does the supremum of B exist? If yes does it belong to B? Does the infimum of B exist? Does it belong to B?

DEFINITION 11.17. A *well ordered* set is an ordered set $(A, \leq)$ such that every non-empty subset $B \subset A$ has a minimum element.

EXAMPLE 11.18. Let $A = \{a, b, c, d\}$ and let $\leq$ be the smallest order relation containing

$$(a, b), (b, c), (c, d)$$

Then $(A, \leq)$ is well ordered.

EXERCISE 11.19. Prove that every well ordered set is totally ordered.

REMARK 11.20. Later, when we will have introduced the ordered set of integers and the ordered set of rational numbers we will see that both are totally ordered, the non-negative integers are well ordered but the non-negative rationals are not well ordered.

The following theorems can be proved (but their proof is beyond the scope of this course):

THEOREM 11.21. *(Zorn's lemma) Assume $(A, \leq)$ is a non-empty ordered set. Assume that every totally ordered subset $B \subset A$ has an upper bound in A. Then A has a maximal element.*

THEOREM 11.22. *(Well ordering principle) Let A be a set. Then there exists an order relation $\leq$ on A such that $(A, \leq)$ is well ordered.*

REMARK 11.23. It can be proved that if one removes from the axioms of Set Theory the axiom of choice then the axiom of choice, Zorn's lemma, and the well ordering principle are all equivalent.

EXERCISE 11.24. Let $(A, \leq)$ and $(B, \leq)$ be totally ordered sets. Define a relation $\leq$ on $A \times B$ by

$$((a,b) \leq (a',b')) \leftrightarrow ((a < a') \vee ((a = a') \wedge (b \leq b'))).$$

Prove that $\leq$ is an order on $A \times B$ (it is called the *lexicographic order*) and that $(A \times B, \leq)$ is totally ordered. (Explain how this order is being used to order words in a dictionary.)

DEFINITION 11.25. Let $R \subset A \times A$ be a relation on A and write $a \sim b$ instead of aRb. R is called an *equivalence* relation if and only if for all $a, b, c \in A$, we have:
1) $a \sim a$ (reflexivity),
2) $a \sim b$ and $b \sim c$ imply $a \sim c$ (transitivity),
3) $a \sim b$ implies $b \sim a$ (symmetry);
we also say that $\sim$ is an equivalence relation.

EXERCISE 11.26. Prove that the symmetry condition is equivalent to $R^t = R$.

EXERCISE 11.27. Let $R_0 \subset A \times A$ be a relation and let

$$R = \bigcap_{R' \supset R_0} R'$$

be the intersection of all equivalence relations R' containing R_0. Prove that R is an equivalence relation and it is the smallest equivalence relation containing R_0 in the sense that it is contained in any other equivalence relation that contains R_0.

DEFINITION 11.28. Given an equivalence relation $\sim$ as above for every $a \in A$ we may consider the set

$$\widehat{a} = \{c \in A \mid c \sim a\}$$

called the *equivalence class* of a. Sometimes, instead of $\widehat{a}$, one writes $\overline{a}$ or $[a]$.

EXERCISE 11.29. Prove that $\widehat{a} = \widehat{b}$ if and only if $a \sim b$.

EXERCISE 11.30. Prove that:
1) if $\widehat{a} \cap \widehat{b} \neq \emptyset$ then $\widehat{a} = \widehat{b}$;
2) $A = \bigcup_{a \in A} \widehat{a}$.

DEFINITION 11.31. If A is a set a *partition* of A is a family $(A_i)_{i \in I}$ of subsets $A_i \subset A$ such that:
1) if $i \neq j$ then $A_i \cap A_j = \emptyset$
2) $A = \bigcup_{i \in I} A_i$.

EXERCISE 11.32. Let A be a set and $\sim$ an equivalence relation on it. Prove that:

1) There exists a subset $B \subset A$ which contains exactly one element of each equivalence class (such a set is called a *system of representatives*. Hint: Use the axiom of choice).

2) The family $(\widehat{b})_{b\in B}$ is a partition of A.

EXERCISE 11.33. Let A be a set and $(A_i)_{i\in I}$ a partition of A. Define a relation R on A as follows:

$$R = \{(a, b) \in A \times A \mid \exists i((i \in I) \wedge (a \in A_i) \wedge (b \in A_i))\}.$$

Prove that R is an equivalence relation.

EXERCISE 11.34. Let A be a set. Prove that there is a bijection between the set of equivalence relations on A and the set of partitions of A.

Hint: Use the above two exercises.

DEFINITION 11.35. The set of equivalence classes

$$\{\alpha \in \mathcal{P}(A) \mid \exists a((a \in A) \wedge (\alpha = \widehat{a})\}$$

is denoted by $A/\sim$ and is called the *quotient* of A by the relation $\sim$.

EXAMPLE 11.36. For instance if $A = \{a, b, c\}$ and

$$R = \{(a, a), (b, b), (c, c), (a, b), (b, a)\}$$

then R is an equivalence relation, $\widehat{a} = \widehat{b} = \{a, b\}$, $\widehat{c} = \{c\}$, and $A/\sim = \{\{a, b\}, \{c\}\}$.

EXERCISE 11.37. Let $A = \{a, b, c, d, e, f\}$ and $R_0 = \{(a, b), (b, c), (d, e)\}$. Find the smallest equivalence relation R containing R_0. Call it $\sim$. Write down the equivalence classes $\widehat{a}, \widehat{b}, \widehat{c}, \widehat{d}, \widehat{e}, \widehat{f}$. Write down the set $A/\sim$.

EXERCISE 11.38. Let S be a set. For every sets $X, Y \in S$ write $X \sim Y$ if and only if there exists a bijection $F : X \to Y$. This defines a relation on S. Prove that this is an equivalence relation.

EXERCISE 11.39. Let $S = \{A, B, C, D\}$, $A = \{a, b\}$, $B = \{b, c\}$, $C = \{x, y\}$, $D = \emptyset$. Let $\sim$ be the equivalence relation on S defined in the previous exercise. Write down the equivalence classes $\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}$ and write down the set $S/\sim$.

DEFINITION 11.40. An *affine plane* is a pair $(A, \mathcal{L})$ where A is a set and $\mathcal{L} \subset \mathcal{P}(A)$ is a set of subsets of A satisfying a series of properties (which we call, by abuse, *axioms*) which we now explain. It is convenient to introduce some terminology as follows. A is called the *affine plane*. The elements of A are called *points*. The elements L of $\mathcal{L}$ are called *lines*; so each such L is a subset of A. We say a point P lies on a line L if $P \in L$; we also say that L passes through P. We say that two lines intersect if they have a point in common; we say that two lines are *parallel* if they either coincide or they do not intersect. We say that 3 points are *collinear* if they lie on the same line. Here are the axioms that we impose:

1) There exist 3 points which are not collinear and every line has at least 2 points.

2) Every 2 distinct points lie on exactly one line.

3) If L is a line and P is a point not lying on L there exists exactly one line through P which is parallel to L.

Remark 11.41. Note that we have not defined 2 or 3 yet; this will be done later when we introduce integers. The meaning of these axioms is, however, clearly expressible in terms that were already defined. For instance axiom 2 says that for every points P and Q with $P \neq Q$ there exists a line through P and Q; we do not need to define the symbol 2 to express this. The same holds for the use of the symbol 3.

Exercise 11.42. Prove that every two distinct non-parallel lines intersect in exactly one point.

Exercise 11.43. Let $A = \{a, b\} \times \{a, b\}$ with $a \neq b$ and let $\mathcal{L} \subset \mathcal{P}(A)$ consist of all subsets of 2 elements; there are 6 of them. Prove that $(A, \mathcal{L})$ is an affine plane. (Again one can reformulate everything without reference to the symbols 2 or 6; one simply uses 2 or 6 letters and writes that they are pairwise unequal.)

Exercise 11.44. Let $A = \{a, b, c\} \times \{a, b, c\}$ with a, b, c distinct . Find all subsets $\mathcal{L} \subset \mathcal{P}(A)$ such that $(A, \mathcal{L})$ is an affine plane. (This is tedious ! Rather than giving all details describe how the solution would be found.)

Definition 11.45. A *projective plane* is a pair $(\overline{A}, \overline{\mathcal{L}})$ where $\overline{A}$ is a set and $\overline{\mathcal{L}} \subset \mathcal{P}(\overline{A})$ is a set of subsets of $\overline{A}$ satisfying a series of axioms which we now explain. Again it is convenient to introduce some terminology as follows. $\overline{A}$ is called the *projective plane*. The elements of $\overline{A}$ are called *points*, P. The elements $\overline{L}$ of $\overline{\mathcal{L}}$ are called *lines*; so each such $\overline{L} \subset \overline{A}$. We say a point P lies on a line $\overline{L}$ if $P \in \overline{L}$; we also say that $\overline{L}$ *passes through* P. We say that two lines *intersect* if they have a point in common; we say that two lines are *parallel* if they either coincide or they do not intersect. We say that 3 points are *collinear* if they lie on the same line. Here are the axioms that we impose:

1) There exist 3 points which are not collinear and every line has at least 3 points.

2) Every 2 distinct points lie on exactly one line.

3) Every 2 distinct lines meet in exactly one point.

Example 11.46. One can attach to every affine plane $(A, \mathcal{L})$ a projective plane $(\overline{A}, \overline{\mathcal{L}})$ as follows. We introduce the relation $\parallel$ on $\mathcal{L}$ by letting $L \parallel L'$ if and only if L and L' are parallel. This is an equivalence relation (see Exercise 11.47). Denote by $\widehat{L}$ the equivalence class of L. Then we consider the set of equivalence classes, $\overline{L}_\infty = \mathcal{L}/\parallel$; call this set the *line at infinity*. Define $\overline{A} = A \coprod \overline{L}_\infty$ and let $\epsilon_1 : A \to \overline{A}$ and $\epsilon_2 : \overline{L}_\infty \to \overline{A}$. Define a line in $\overline{A}$ to be either $\epsilon_2(\overline{L}_\infty)$ or a set of the form $\overline{L} = \epsilon_1(L) \cup \{\epsilon_2(\widehat{L})\}$. Finally define $\overline{\mathcal{L}}$ to be the set of all lines in $\overline{A}$.

Exercise 11.47. With the notation in Example 11.46 prove that the relation $\parallel$ is an equivalence relation.

Exercise 11.48. With the notation in Example 11.46 check that $(\overline{A}, \overline{\mathcal{L}})$ is a projective plane.

Exercise 11.49. Describe the projective plane attached to the affine plane in Exercise 11.43; how many points does it have? How many lines?

CHAPTER 12

Operations

The concept of operation on a set is an abstraction of “familiar” operations such as addition and multiplication of numbers, composition of functions, etc. Sets with operations on them will be referred to as algebraic structures. The study of algebraic structures is referred to as (modern) algebra and took the shape known today through work (in number theory and algebraic geometry) done by Kronecker, Dedekind, Hilbert, Emmy Noether, etc. Here we introduce operations in general, and some algebraic structures such as rings, fields, and Boolean algebras. We prefer to postpone the introduction of other algebraic structures such as groups, vector spaces, etc., until more theory is being developed.

Definition 12.1. A *binary operation* $\star$ on a set A is a map $\star : A \times A \to A$, $(a, b) \mapsto \star(a, b)$. We usually write $a \star b$ instead of $\star(a, b)$. For instance, we write $(a \star b) \star c$ instead of $\star(\star(a, b), c)$. Instead of $\star$ we sometimes use notation like $+, \times, \circ, \ldots.$

Remark 12.2. The above defines a new (binary) predicate *binop* which in Argot reads “*... is a binary operation on ...*” and we may introduce a corresponding new functional symbol (still denoted by $\star$). So *binop* is introduced by the definition:

$$\forall x \forall y (binop(x, y) \leftrightarrow (fun(x, y \times y, y)))$$

Definition 12.3. A *unary* operation $'$ on a set A is a map $' : A \to A$, $a \mapsto {}'(a)$. We usually write a' or ${}'a$ instead of ${}'(a)$. Instead of $'$ we sometimes use notation like $-, i, \ldots.$

Example 12.4. Let $S = \{0, 1\}$ where $0, 1$ are two distinct sets. Then there are 3 interesting binary operations on S denoted by $\wedge, \vee, +$ (and called *supremum, infimum*, and *addition*) defined as follows:

$$0 \wedge 0 = 0, \quad 0 \wedge 1 = 0, \quad 1 \wedge 0 = 0, \quad 1 \wedge 1 = 1;$$

$$0 \vee 0 = 0, \quad 0 \vee 1 = 1, \quad 1 \vee 0 = 1, \quad 1 \vee 1 = 1;$$

$$0 + 0 = 0, \quad 0 + 1 = 1, \quad 1 + 0 = 1, \quad 1 + 1 = 0.$$

The symbol $\wedge$ is also denoted by $\times$ or $\cdot$; it is referred to as multiplication. The symbol $+$ is also denoted by Δ. Also there is a unary operation $\neg$ on S defined by

$$\neg 1 = 0, \quad \neg 1 = 0.$$

Note that if we denote 0 and 1 by F and T then the operations $\wedge, \vee, \neg$ on $\{0, 1\}$ correspond exactly to the “logical operations” on F and T defined in the chapter on tautologies. This is not a coincidence!

Exercise 12.5. Compute $((0 \wedge 1) \vee 1) + (1 \wedge (0 \vee (1 + 1)))$.

In what follows we define some of the basic concepts of algebra: groups, Boolean algebras, rings, ordered rings, integral domains, fields. (More precisely we define, in Set Theory, new predicates: *is a groups*, *is a Boolean algebra*, *is a ring*, etc.)

DEFINITION 12.6. A *group* is a tuple $(G, \star, e)$ where G is a set, $\star$ is a binary operation on G, and $e \in G$ such that the following properties hold:
1) For all $x, y, z \in G$, $x \star (y \star z) = (x \star y) \star z$.
2) For all $x \in G$, $x \star e = e \star x = x$.
3) For all $x \in G$ there exists $y \in G$ such that $x \star y = y \star x = e$.

REMARK 12.7. Properties 1, 2, 3 in Definition 12.6 above are sometimes abusively (or mistakenly) referred to as the "axioms of groups." Indeed, properties 1, 2, 3 in Definition 12.6 above are NOT axioms in T_{set}! Rather, they are part of a Definition in T_{set}. Their form does resemble/mirror the Axioms 1,2,3 of Group Theory in Example 5.10 but their logical status is entirely different. (A similar remark can be made about the definition of Boolean algebras, rings, ordered rings, etc. that will be given below.) This type of "mirroring correspondence" between axioms in one theory (such as Group Theory) and definitions in another theory (such as Set Theory) can be made precise in Metalanguage but is not part of Mathematics (i.e., of Set Theory). Nevertheless this "mirroring correspondence" can be itself "mirrored" inside Mathematics as we will see in the Chapter on Models.

As a consequence, if an exercise in Mathematics (i.e., in Set Theory) says: "Let G be a group; prove that P", then one assumes properties 1, 2, 3 in Definition 12.6 (not as axioms of Set Theory but as hypothesis of the exercise) and one attempts to prove P.

On the other hand if an exercise in Mathematics (i.e., in Set Theory) says: "Let $(G, \star, e)$ be the following triple [...here a definition of triple is given...]; prove that $(G, \star, e)$ is a group," then one needs to prove that properties 1, 2, 3 in Definition 12.6 hold for the given triple (so one cannot assume properties 1, 2, 3 in Definition 12.6 are true, since they are not axioms of Set Theory).

DEFINITION 12.8. A *Boolean algebra* is a tuple

$$(A, \vee, \wedge, \neg, 0, 1)$$

where $\wedge, \vee$ are binary operations, $\neg$ is a unary operation, and $0, 1 \in A$ such that for all $a, b, c \in A$ the following "axioms" are satisfied:
1) $a \wedge (b \wedge c) = (a \wedge b) \wedge c$, $a \vee (b \vee c) = (a \vee b) \vee c$,
2) $a \wedge b = b \wedge a$, $a \vee b = b \vee a$,
3) $a \wedge 1 = a$, $a \vee 0 = a$,
4) $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$, $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$
5) $a \wedge (\neg a) = 0$, $a \vee (\neg a) = 1$.

DEFINITION 12.9. A *commutative unital ring* (or simply a *ring*) is a tuple

$$(R, +, \times, -, 0, 1)$$

(sometimes referred to simply as R) where R is a set, $0, 1 \in R$, $+, \times$ are two binary operations (write $a \times b = ab$), and $-$ is a unary operation on R such that for every $a, b, c \in R$ the following hold:
1) $a + (b + c) = (a + b) + c$, $a + 0 = a$, $a + (-a) = 0$, $a + b = b + a$;
2) $a(bc) = (ab)c$, $1a = a$, $ab = ba$,
3) $a(b + c) = ab + ac$.

The element 1 is referred to as the *identity*; 0 is referred to as the *zero element*. The conditions $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$ are referred to as *associativity*. The conditions $a + b = b + a$ and $ab = ba$ are referred to as *commutativity*.

The above defines a new (6-ary) predicate *ring* which reads "... is a ring with respect to the addition ... the multiplication ... subtraction ... zero element ... and unit element ...". This predicate is introduced by the definition

$$\forall x \forall y \forall z \forall u \forall v \forall w (ring(x, y, z, u, v, w) \leftrightarrow (binop(y, x) \wedge ....))$$

DEFINITION 12.10. We write $a + b + c$ instead of $(a + b) + c$ and abc for $(ab)c$. We write $a - b$ instead of $a + (-b)$.

EXERCISE 12.11. Let R be a ring. Prove that:
1) $x \cdot 0 = 0$ for all $x \in R$.
2) $x \cdot (-1) = -x$ for all $x \in R$.
3) $(-1) \cdot (-1) = 1$

Hint: For 1) start with $0 + 0 = 0$. For 2) and 3) start with $1 + (-1) = 0$

REMARK 12.12. By 1) above if $1 = 0$ then $R = \{0\}$ (in which case R is called a *zero ring*).

DEFINITION 12.13. A ring R is called an *integral domain* if and only if for all $x, y \in R$ if $xy = 0$ then $x = 0$ or $y = 0$.

DEFINITION 12.14. An element a of a ring R is *invertible* if and only if there exists $a' \in R$ such that $aa' = 1$; this a' is then easily proved to be unique. It is called the *inverse* of a, and is denoted by a^{-1}. We denote by $R^\times$ the set of invertible elements of R; clearly if $a, b \in R^\times$ then $1, ab, a^{-1} \in R^\times$. A ring R is called a *field* if $0 \neq 1$ and every non-zero element is invertible, i.e., if $R^\times = R \setminus \{0\}$. For a, b in a field with $b \neq 0$ we write

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

We usually denote fields by different letters (such as K) instead of R.

EXERCISE 12.15. Prove that every field is an integral domain.

DEFINITION 12.16. Let $n \in \mathbb{N}$. By an nth *root of unity* in a ring R we understand an element $\alpha \in R$ such that $\alpha^n = 1$. By a *root of unity* we understand an element which is an nth root of unity for some $n \in \mathbb{N}$.

Clearly the roots of unity in R are invertible.

DEFINITION 12.17. A *Boolean ring* is a commutative unital ring A such that $1 \neq 0$ and for all $a \in A$ we have $a^2 = a$.

EXERCISE 12.18. Prove that in a Boolean ring A we have $a + a = 0$ for all $a \in A$.

EXERCISE 12.19. Prove that
1) $(\{0, 1\}, \vee, \wedge, \neg, 0, 1)$ is a Boolean algebra.
2) $(\{0, 1\}, +, \times, I, 0, 1)$ is a Boolean ring and a field (I is the identity map). This field is denoted by $\mathbb{F}_2$.

EXERCISE 12.20. Prove that if a Boolean ring A is a field then $A = \{0, 1\}$.

DEFINITION 12.21. Let A be a set and let $\mathcal{P}(A)$ be the power set of A. Define the following operations on $\mathcal{P}(A)$:

$$\begin{array}{rcl} X \wedge Y & = & X \cap Y \\ X \vee Y & = & X \cup Y \\ X \Delta Y & = & (X \cup Y) \backslash (X \cap Y) \\ \neg X & = & \mathcal{C}X = A \backslash X. \end{array}$$

EXERCISE 12.22. Prove that
1) $(\mathcal{P}(A), \vee, \wedge, \neg, \emptyset, A)$ is a Boolean algebra;
2) $(\mathcal{P}(A), \Delta, \wedge, I, \emptyset, A)$ is a Boolean ring (I is the identity map).

DEFINITION 12.23. Given a subset $X \subset A$ one can define the *characteristic function* $1_X : A \to \{0,1\}$ by letting $1_X(a) = 1$ if and only if $a \in X$.

EXERCISE 12.24. Prove that
1) $1_{X \vee Y}(a) = 1_X(a) \vee 1_Y(a)$,
2) $1_{X \wedge Y}(a) = 1_X(a) \wedge 1_Y(a)$,
3) $1_{X \Delta Y}(a) = 1_X(a) + 1_Y(a)$,
4) $1_{\neg X}(a) = \neg 1_X(a)$.

DEFINITION 12.25. An *algebraic structure* is a tuple $(A, \star, \bullet, ..., \neg, -, ..., 0, 1, ...)$ where A is a set, $\star, \bullet, ...$ are binary operations, $\neg, -, ...$ are unary operations, and $1, 0, ...$ are given elements of A. (Some of these may be missing; for instance we could have only one binary operation, one given element, and no unary operations.) Assume we are given two algebraic structures

$$(A, \star, \bullet, ..., \neg, -, ..., 0, 1, ...) \quad \text{and} \quad (A', \star', \bullet', ..., \neg', -', ..., 0', 1', ...)$$

(with the same number of corresponding operations). A map $F : A \to A'$ is called a *homomorphism* if for all $a, b \in A$ we have:
1) $F(a \star b) = F(a) \star' F(b)$, $F(a \bullet b) = F(a) \bullet' F(b)$,...
2) $F(\neg a) = \neg' F(a)$, $F(-a) = -' F(a)$,...
3) $F(0) = 0'$, $F(1) = 1'$,....

EXAMPLE 12.26. A map $F : A \to A'$ between two commutative unital rings is called a *homomorphism* (of commutative unital rings) if for all $a, b \in A$ we have:
1) $F(a + b) = F(a) + F(b)$ and $F(ab) = F(a)F(b)$,
2) $F(-a) = -F(a)$ (prove that this is automatic !),
3) $F(0) = 0$ (prove that this is automatic !) and $F(1) = 1$.

EXERCISE 12.27. Prove that if $F : A \to A'$ is a homomorphism of algebraic structures and F is bijective then its inverse $F^{-1} : A' \to A$ is a homomorphism. Such an F will be called an *isomorphism*.

DEFINITION 12.28. Let Ω be a set. A subset $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called a *Boolean algebra* of sets if the following hold:
1) $\emptyset \in \mathcal{A}$, $\Omega \in \mathcal{A}$;
2) If $B, C \in \mathcal{A}$ then $B \cap C \in \mathcal{A}$, $B \cup C \in \mathcal{A}$, $\mathcal{C}B \in \mathcal{A}$.
(Hence $(\mathcal{A}, \vee, \wedge, \mathcal{C}, \emptyset, \Omega)$ is a Boolean algebra.)

EXERCISE 12.29. Prove that if $\mathcal{A}$ is a Boolean algebra of sets then for every $B, C \in \mathcal{A}$ we have $B \Delta C \in \mathcal{A}$. Prove that $(\mathcal{A}, \Delta, \cap, I, \emptyset, \Omega)$ is a Boolean ring.

Definition 12.30. A subset $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called a *Boolean ring* of sets if the following properties hold:

1) $\emptyset \in \mathcal{A}$, $\Omega \in \mathcal{A}$;

2) If $B, C \in \mathcal{A}$ then $B \cap C \in \mathcal{A}$, $B \Delta C \in \mathcal{A}$.

(Hence $(\mathcal{A}, \Delta, \vee, I, \emptyset, \Omega)$ is a Boolean ring.)

Exercise 12.31. Prove that every Boolean ring of sets is a Boolean algebra of sets.

Exercise 12.32. Let $(A, \vee, \wedge, \neg, 0, 1)$ be a Boolean algebra. For every $a, b \in A$ set

$$a + b = (a \vee b) \wedge (\neg(a \wedge b)).$$

Prove that $(A, +, \wedge, I, 0, 1)$ is a Boolean ring (I the identity map).

Exercise 12.33. Let $(A, +, \times, -, 0, 1)$ be a Boolean ring. For every $a, b \in A$ let

$$\begin{aligned} a \vee b &= a + b - ab \\ a \wedge b &= ab \\ \neg a &= 1 - a. \end{aligned}$$

Prove that $(A, \vee, \wedge, \neg, 0, 1)$ is a Boolean algebra.

Exercise 12.34. Let X be a set and $(R, +, \cdot, -, 0, 1)$ a commutative unital ring. Let $Fun(X, R)$ be the set of all functions $X \to R$. For $F, G \in Fun(X, R)$ we define $F + G, F \cdot G, -F, 0, 1 \in Fun(X, R)$ by the formulae

$$(F + G)(x) = F(x) + G(x), \quad (F \cdot G)(x) = F(x) \cdot G(x),$$

$$(-F)(x) = -F(x), \quad 0(x) = 0, \quad 1(x) = x,$$

for all $x \in X$. The operations $F + G$ and $F \cdot G$ are called *pointwise* addition and multiplication of functions. Prove that

$$(Fun(X, R), +, \cdot, -, 0, 1)$$

is a commutative unital ring.

Definition 12.35. A *commutative unital ordered ring* (or simply an *ordered ring*) is a tuple

$$(R, +, \times, -, 0, 1, \leq)$$

where

$$(R, +, \times, -, 0, 1)$$

is a ring, $\leq$ is a total order on R, and for all $a, b, c \in R$ the following axioms are satisfied

1) If $a < b$ then $a + c < b + c$;

2) If $a < b$ and $c > 0$ then $ac < bc$.

We say that $a \in R$ is positive if $a > 0$; and that a is negative if $a < 0$. We say a is non-negative if $a \geq 0$.

Exercise 12.36. Let R be an ordered ring . Prove that for all $x, y \in R$:

1) If $x > 0$ and $y > 0$ then $x + y > 0$ and $xy > 0$

2) If $x < 0$ then $-x > 0$.

2) If $0 \neq 1$ then $0 < 1$.

Hint: For 3) use $(-1) \cdot (-1) = 1$.

Exercise 12.37. Prove that every ordered ring is an integral domain.

EXERCISE 12.38. (Triangle inequality). For $x \in R$ define $|x|$ to be x or $-x$ according as $x \geq 0$ or $x \leq 0$ respectively. Prove that for all $a, b \in R$ we have:

$$|a+b| \leq |a| + |b|, \quad |ab| = |a| \cdot |b|.$$

EXERCISE 12.39. Prove that the ring $(\{0,1\}, +, \times, -, 0, 1)$ has no structure of ordered ring i.e., there is no order $\leq$ on $\{0,1\}$ such that $(\{0,1\}, +, \times, -, 0, 1, \leq)$ is an ordered ring.

REMARK 12.40. We cannot give examples of ordered rings yet. Later we will see that the rings of integers, rationals, and reals have natural structures of ordered rings.

CHAPTER 13

Integers

In this Chapter we introduce the ring $\mathbb{Z}$ of integers and we prove some easy theorems about this concept.

Recall that we defined the concept of operation (Definition 12.1), of ring (Definition 12.9) and of ordered ring (Definition 12.35). For an ordered ring

$$(R, +, \times, -, 0, 1, \leq)$$

we define the sets $R_{\geq 0}$ and $R_{>0}$ of non-negative, respectively positive elements by

$$R_{\geq 0} = \{x \in R \mid x \geq 0\}, \quad R_{>0} = \{x \in R \mid x > 0\}.$$

These sets can be viewed as ordered sets with the order induced by $\leq$.

We have the following Theorem in Set Theory T_{set}:

THEOREM 13.1. *There exists a unique ordered ring* $(R, +, \times, -, 0, 1, \leq)$ *such that the following properties hold:*

1) $R_{\geq 0}$ *is a well ordered set.*

2) $0 = \emptyset$.

3) For all $x \in R_{\geq 0}$ *we have* $x + 1 = \{x\}$.

4) For all $x \in R_{>0}$ *we have* $-x = (0, x)$.

A sketch of proof of this Theorem will be given in Exercise 13.6. In view of Theorem 13.1 we may introduce new constants via the following

DEFINITION 13.2. $(\mathbb{Z}, +, \times, -, 0, 1, \leq)$ is the unique ordered ring with $\mathbb{Z}_{\geq 0}$ well ordered, satisfying the conditions 1, 2, 3 in Theorem 13.1. We call $\mathbb{Z}$ (with these operations, 0, 1, and order) the *ring of integers*. We write $\mathbb{N} = \mathbb{Z}_{>0}$ and we call it the set of *natural numbers*.

Note that $0 \neq 1$ so by Exercise 12.36 we have $0 < 1$, hence $1 = 0 + 1 = \{\emptyset\} \in \mathbb{N}$. Hence the elements

$$\{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, ...$$

belong to $\mathbb{N}$; the first is 1, the next two will later be called 2 and 3, etc. Moreover the elements $-1, -2, -3, ...$ are then:

$$(\emptyset, \{\emptyset\}), (\emptyset, \{\{\emptyset\}\}), (\emptyset, \{\{\{\emptyset\}\}\}), ...$$

REMARK 13.3. The only predicate in the language L_{set} of sets is $\in$ and the terms (in particular the constants) in this language are called *sets*. In particular when we consider the ordered ring of integers $(\mathbb{Z}, +, \times, 0, 1, \leq)$ the symbols $\mathbb{Z}, +, -, \times, 0, 1, \leq, \mathbb{N}$ are all constants (they are sets). In particular $+, \times$ are not originally functions and $\leq$ is not originally a predicate. But, according to our conventions, we may introduce functions (still denoted by $+, -, \times$) and a predicate (still denoted by $\leq$) via appropriate definitions. (This is because "*the set* $+$ *is a binary operation on* $\mathbb{Z}$" is a theorem, etc.)

EXERCISE 13.4. Prove that if $a \in \mathbb{Z}$ then the set $\{x \in \mathbb{Z} \mid a-1 < x < a\}$ is empty.

Hint: It is enough to show that the set $S = \{x \in \mathbb{Z} \mid 0 < x < 1\}$ is empty. Assume S is non-empty and let $m = \min S$. We have $0 < m < 1$. Multiplying the latter by m we get $0 < m^2 < m$, hence $0 < m^2 < 1$, so $m^2 \in S$ and $m^2 < m = \min S$, a contradiction.

EXERCISE 13.5. Prove that if $a \in \mathbb{N}$ then $a = 1$ or $a - 1 \in \mathbb{N}$. Conclude that $\min \mathbb{N} = 1$.

Hint: Use the previous exercise.

EXERCISE 13.6. Prove Theorem 13.1 using the following strategy. Recall from Definition 10.47 the definition of a Peano triple and recall from Exercise 10.48 that there exists a Peano triple. Assume in what follows that $(N, 1, \sigma)$ is a Peano triple. For $y \in N$ let

$$A_y = \{\tau \in Fun(N, N) \mid \tau(1) = \sigma(y), \forall x(\tau(\sigma(x)) = \sigma(\tau(x)))\}.$$

Then one proceeds as follows.

1) One proves that A_y has at most one element. Hint: If $\tau, \eta \in A_y$ and $S = \{x \mid \tau(x) = \eta(x)\}$ then $1 \in S$ and $\sigma(S) \subset S$; so $S = N$.

2) One proves that for every y, $A_y \neq \emptyset$. Hint: If $T = \{y \in N \mid A_y \neq \emptyset\}$ then $1 \in T$ and $\sigma(T) \subset T$; so $T = N$.

3) By 1 and 2 we may write $A_y = \{\tau_y\}$. Then define $+$ on N by $x + y = \tau_y(x)$.

4) One proves that $x + y = y + x$ and $(x + y) + z = x + (y + z)$ on N.

5) One proves that if $x, y \in N$, $x \neq y$, then there exists $z \in N$ such that either $y = x + z$ or $x = y + z$.

6) Define $N^- = \{0\} \times N$, $R = N^- \cup \{0\} \cup N$. One naturally extends $+$ to R.

7) One defines $\times$ on N and then on R in the same style as for $+$.

8) One defines $\leq$ on N and one proves that $(N, \leq)$ is well ordered. One extends this to R.

9) One proves that $(R, +, \times, -, 0, 1, \leq)$ is an ordered ring with $R_{\geq 0}$ well ordered.

So the existence part of Theorem 13.1 follows. The uniqueness part easily follows from the conditions 2, 3, 4 in the Theorem.

DEFINITION 13.7. Define the natural numbers $2, 3, ..., 9$ by

$$\begin{aligned} 2 &= 1 + 1 \\ 3 &= 2 + 1 \\ &.... \\ 9 &= 8 + 1. \end{aligned}$$

Define $10 = 2 \times 5$. Define $10^2 = 10 \times 10$, etc. Define symbols like 423 as being $4 \times 10^2 + 2 \times 10 + 3$, etc. This is called a *decimal representation*.

EXERCISE 13.8. Prove that $12 = 9 + 3$.

Hint: We have:

$$\begin{aligned} 12 &= 10 + 2 \\ &= 2 \times 5 + 2 \\ &= (1 + 1) \times 5 + 2 \\ &= 1 \times 5 + 1 \times 5 + 2 = 5 + 5 + 2 \\ &= 5 + 5 + 1 + 1 = 5 + 6 + 1 = 5 + 7 = 4 + 1 + 7 \\ &= 4 + 8 = 3 + 1 + 8 = 3 + 9 = 9 + 3. \end{aligned}$$

The first proof of this kind for was given by Leibniz (who considered $2 + 2 = 4$).

EXERCISE 13.9. Prove that $18 + 17 = 35$. Prove that $17 \times 3 = 51$.

REMARK 13.10. In Kant's analysis, statements like the ones in the previous exercise were viewed as synthetic; in contemporary Mathematics, hence in the approach we follow, all these statements are, on the contrary, analytic statements. (The definition of analytic/synthetic is taken here in the sense of Leibniz and Kant.)

EXERCISE 13.11. Prove that $7 \leq 20$.

DEFINITION 13.12. For every integers $a, b \in \mathbb{Z}$ the set $\{x \in \mathbb{Z} \mid a \leq x \leq b\}$ will be denoted, for simplicity, by $\{a, ..., b\}$. This set is clearly empty if $a > b$. If other numbers in addition to a, b are specified then the meaning of our notation will be clear from the context; for instance $\{0, 1, ..., n\}$ means $\{0, ..., n\}$ whereas $\{2, 4, 6, ..., 2n\}$ will mean $\{2x \mid 1 \leq x \leq n\}$, etc. A similar convention applies if there are no numbers after the dots.

EXAMPLE 13.13. $\{-2, ..., 11\} = \{-2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$.

Recall that a subset $A \subset \mathbb{N}$ is bounded (equivalently bounded from above) if there exists $b \in \mathbb{N}$ such that $a \leq b$ for all $a \in A$; we say that A is bounded by b from above.

EXERCISE 13.14. Prove that $\mathbb{N}$ is not bounded.

EXERCISE 13.15. Prove that every subset of $\mathbb{Z}$ bounded from above has a maximum.

Hint: If A is bounded from above by b consider the set $\{b - x \mid x \in A\}$.

DEFINITION 13.16. An integer a is *even* if there exists an integer b such that $a = 2b$. An integer is *odd*if it is not even.

EXERCISE 13.17. Prove that if a is odd then $a - 1$ is even.

Hint: Consider the set $\{b \in \mathbb{N} \mid 2b \geq a\}$, and let c be the minimum element of S. Then show that $2(c - 1) < a$. Finally show that this implies $a = 2c - 1$.

EXERCISE 13.18. Prove that if a and b are odd then ab is odd.

Hint: Write $a = 2c + 1$ and $b = 2d + 1$ (cf. the previous exercise) and compute $(2c + 1)(2d + 1)$.

EXERCISE 13.19. Consider the following sentence: There is no bijection between $\mathbb{N}$ and $\mathbb{Z}$. Explain the mistake in the following wrong proof; this is an instance of a fallacy discussed earlier.

"*Proof.*" Assume there is a bijection $f : \mathbb{N} \to \mathbb{Z}$. Define $f(x) = x$. Then f is not surjective so it is not a bijection.

EXERCISE 13.20. Prove that there is a bijection between $\mathbb{N}$ and $\mathbb{Z}$.

CHAPTER 14

Induction

Induction is the single most important method to prove elementary theorems about the integers. (More subtle theorems, such as many of the theorems of "number theory," require more sophisticated methods.) Proofs by induction first appear in the works of Pascal and Jakob Bernoulli in the 17th century.

Let $P(x)$ be a formula in the language L_{set} of sets, with one free variable x. We shall always assume, in what follows, that $P(x)$ is the conjunction of "$x \in \mathbb{N}$" with some other formula; this can be thought of as saying that $P(x)$ is a formula in which x is assumed to be in $\mathbb{N}$. For each such $P(x)$ we have:

THEOREM 14.1. *(Induction Principle for $P(x)$) Assume*
1) $P(1)$.
2) For all $n \neq 1$ if $P(n-1)$ then $P(n)$.
Then for all n we have $P(n)$.

Note that 2) is equivalent to:

2') For all n if $P(n)$ then $P(n+1)$.

There is no advantage in using 2) over 2') or vice versa; we will usually use 2) in what follows.

The above theorem is expressed, as usual, in Argot. The same expressed as a sentence in L_{set} reads:

$$(P(1) \wedge ((\forall x((x \neq 1) \wedge P(x-1)) \rightarrow P(x))) \rightarrow (\forall x P(x)).$$

Proving that "for all n we have $P(n)$" by proving first that $P(1)$ and second that "for all $n \neq 1$ we have $P(n-1) \rightarrow P(n)$" is called a proof by *induction on n*. The proof of $P(n-1) \rightarrow P(n)$ is called the *induction step*. We call $P(n-1)$ the *induction hypothesis*.

Note that we have a theorem for each $P(x)$. Note also that the above Theorem does not say "for all formula P something happens"; that would not be a sentence in the language of sets. It would not be a metasentence in Metalanguage either because it contains quantifiers.

Proof. Let $S = \{n \in \mathbb{N} \mid \neg P(n)\}$. We want to show that $S = \emptyset$. Assume $S \neq \emptyset$ and seek a contradiction. Let m be the minimum of S; in particular $m \in S$. By 1) $m \neq 1$. By Exercise 13.5 $m - 1 \in \mathbb{N}$. By minimality of m, we have $P(m-1)$. By 2) we get $P(m)$ so $m \notin S$, a contradiction. $\square$

Say that an integer a divides an integer b if there exists an integers c such that $b = ac$; write $a|b$ for "a divides b." Define $n^2 = n \times n$, $n^3 = n^2 \times n$, $n^4 = n^3 \times n$, $n^5 = n^4 \times n$ for every integer n.

EXERCISE 14.2. Prove that for every natural n we have $3|n^3-n$.

Hint: Proceed by induction on n as follows. We let $P(n)$ be the sentence: $3|n^3-n$. $P(1)$ is true because $1^3-1=3\times 0$. Assume now that $P(n-1)$ is true (the induction hypothesis) i.e., $(n-1)^3-(n-1)=3q$ for some integer q and let us check that $P(n)$ is true i.e., that $n^3-n=3r$ for some integer r. The equality $(n-1)^3-(n-1)=3q$ reads $n^3-3n^2+3n-1-n+1=3q$. Hence $n^3-n=3(n^2-n+q)$ and we are done by taking $r=n^2-n+q$.

EXERCISE 14.3. Prove that for every natural n we have $5|n^5-n$.

REMARK 14.4. Assume we want to prove a sentence S of the form "for all n and for all m we have $Q(n,m)$" with $Q(x,y)$ having two free variables x,y; so induction does not apply directly. To apply induction we may view this sentence as equal to "for all n we have $P(n)$" where $P(x)=$"$\forall y Q(x,y)$" has now only one free variable x. If we prove "for all n we have $P(n)$" by induction we say that we have proved S by "induction on n." We can of course also view S as equivalent to "for all m we have $R(m)$" with $R(y)=$"$\forall x Q(x,y)$." If we prove "for all m we have $R(m)$" by induction we say that we have proved S by "induction on m." So there are two choices for a proof by induction of a sentence like S and it is sometimes an art to pick the right choice. A similar discussion holds for the case when we have $3,4,\ldots$ variables.

Here is an example of this situation; the statement below "depends on two natural numbers n and m". Recall that we denoted

$$\{1,\ldots,n\}=\{x\in\mathbb{N}\mid 1\le x\le n\}.$$

PROPOSITION 14.5. *For all n and m if there exists a bijection*

$$\{1,\ldots,n\}\to\{1,\ldots,m\}$$

then $n=m$.

Proof. We prove the statement by induction on n. So we let $P(n)$ be the the following sentence: "for all m if there exists a bijection $\{1,\ldots,n\}\to\{1,\ldots,m\}$ then $n=m$." We prove "for all n, $P(n)$" by induction. Clearly $P(1)$ is true; cf. the Exercise below. Assume now $P(n-1)$ is true and let's prove that $P(n)$ is true. So take m arbitrary and consider a bijection $F:\{1,\ldots,n\}\to\{1,\ldots,m\}$; we want to prove that $n=m$. Let $i=F(n)$ and define the map $G:\{1,\ldots,n-1\}\to\{1,\ldots,m\}\backslash\{i\}$ by $G(j)=F(j)$ for all $1\le j\le n-1$. Then clearly G is a bijection. Now consider the map $H:\{1,\ldots,m\}\backslash\{i\}\to\{1,\ldots,m-1\}$ defined by $H(j)=j$ for $1\le j\le i-1$ and $H(j)=j-1$ for $i+1\le j\le m$. (The definition is correct because for every $j\in\{1,\ldots,m\}\backslash\{i\}$ either $j\le i-1$ or $j\ge i+1$; cf. Exercise 13.4.) Clearly H is a bijection. We get a bijection

$$H\circ G:\{1,\ldots,n-1\}\to\{1,\ldots,m-1\}.$$

Since $P(n-1)$ is true we get $n-1=m-1$. (Note that it was crucial that $P(n)$ is a sentence that involves "for all m" because we applied $P(n-1)$ to the case when m was replaced by $m-1$.) Hence $n=m$ and we are done. $\square$

EXERCISE 14.6. Check that $P(1)$ is true in the above Proposition.

REMARK 14.7. Note the general strategy of proofs by inductions. Say $P(n)$ is "about n objects." There are two steps. The first step is the verification of $P(1)$ i.e., one verifies the statement "for one object." For the second step (the induction step) one considers a situation with n objects; one "removes" from that situation "one object" to get a "situation with $n-1$ objects"; one uses the "induction hypothesis" $P(n-1)$ to conclude the claim for the "situation with $n-1$ objects." Then one tries to "go back" and prove that the claim is true for the situation with n objects. So the second step is performed by "removing" one object from an arbitrary situation with n objects and NOT by adding one object to an arbitrary situation with $n-1$ objects. Below is an example of a fallacious reasoning by induction based on "adding" instead of "subtracting" an object.

EXAMPLE 14.8. Here is a wrong argument for the induction step in the proof of Proposition 14.5.

"Proof." Let $G : \{1, ..., n-1\} \to \{1, ..., m-1\}$ *be any bijection and let* $F : \{1, ..., n\} \to \{1, ..., m\}$ *be defined by* $F(i) = G(i)$ *for* $i \leq n-1$ *and* $F(n) = m$. *Clearly* F *is a bijection. Now by the induction hypothesis* $n-1 = m-1$. *Hence* $n = m$. *This ends the proof.*

The mistake is that the above does not end the proof: the above argument only proves that $n = m$ in case our bijection $F : \{1, ..., n\} \to \{1, ..., m\}$ is constructed from a bijection $G : \{1, ..., n-1\} \to \{1, ..., m-1\}$ in the special way described above. But an arbitrary bijection $F : \{1, ..., n\} \to \{1, ..., m\}$ does not always arise the way we defined F in the above "proof." In some sense the mistake we just pointed out is that of defining the same constant twice (cf. Example 13.19): we were supposed to define the symbol F as being an arbitrary bijection but then we redefined F in a special way through an arbitrary G. The point is that if G is arbitrary and F is defined as above in terms of G then F will not be arbitrary (because F will always send n into m).

DEFINITION 14.9. A set A is *finite* if and only if there exists an integer $n \geq 0$ and a bijection $F : \{1, ..., n\} \to A$. (Note that n is then unique by Proposition 14.5.) We write $|A| = n$ and we call this number the *cardinality* of A or the *number of elements* of A. (Note that $|\emptyset| = 0$.) If $F(i) = a_i$ we write $A = \{a_1, ..., a_n\}$. (This notation conflicts with the notation $\{a, ...b\}$ introduced in Definition 13.12. However the meaning of the notation should be clear, in each case, from context.) A set is infinite if it is not finite.

EXERCISE 14.10. Prove that $|\{2, 4, -6, 9, -100\}| = 5$.

EXERCISE 14.11. For every finite sets A and B we have that $A \cup B$ is finite and

$$|A \cup B| + |A \cap B| = |A| + |B|.$$

Hint: First examine the case $A \cap B = \emptyset$. In this case if $F : \{1, ..., a\} \to A$ and $G : \{1, ..., b\} \to B$ are bijections prove that $H : \{1, ..., a+b\} \to A \cup B$ defined by $H(i) = F(i)$ for $1 \leq i \leq a$ and $H(i) = G(i-a)$ for $a+1 \leq i \leq a+b$ is a bijection. The case when $A \cap B \neq \emptyset$ can be reduced to the previous case using the fact that since $A \cup B = A \cup (B \backslash A)$ and $B = (B \backslash A) \cup (A \cap B)$ we have

$$|A \cup B| + |A \cap B| = |A| + |B \backslash A| + |A \cap B| = |A| + |B|.$$

EXERCISE 14.12. For every finite sets A and B prove that $A \times B$ is finite and

$$|A \times B| = |A| \times |B|.$$

Hint: Induction on $|A|$.

EXERCISE 14.13. For every finite sets A and B prove that $Fun(A, B)$ is finite and

$$|Fun(A, B)| = |B|^{|A|}.$$

Hint: Induction on $|A|$.

EXERCISE 14.14. Let $F : A \to B$ be a surjective map between finite sets of cardinalities $|A| = a$ and $|B| = b$. Let $c \in \mathbb{N}$ and assume that for all $y \in B$ we have that

$$|F^{-1}(y)| = c.$$

Prove that $a = bc$.

Hint. Induction on b.

EXERCISE 14.15. Let $S = \{1, ..., n\}$ and $A_1, ..., A_n$ finite sets. Prove the following formula (called the *inclusion-exclusion principle*):

$$| \cup_{i \in S} A_i| = \sum_{I \in \mathcal{P}(S) \setminus \{\emptyset\}} (-1)^{|I|+1} | \cap_{i \in I} A_i|.$$

Hint: Use induction on n and Exercise 14.11.

EXERCISE 14.16. Let $F : \{1, ..., n\} \to R$ be a map, where R is a ring, and write $F(i) = a_i$. We refer to such a map as a (finite) family of numbers. Prove that there exists a unique map $G : \{1, ..., n\} \to R$ such that $G(1) = a_1$ and $G(k) = G(k-1)+a_k$ for $2 \leq k \leq n$.

Hint: Induction on n.

DEFINITION 14.17. In the notation of the above Exercise define the (finite) sum $\sum_{i=1}^{n} a_i$ as the number $G(n)$. We also write $a_1 + ... + a_n$ for this sum. If $a_1 = ... = a_n = a$ the sum $a_1 + ... + a_n$ is written as $a + ... + a$ (n times).

EXERCISE 14.18. Prove that for every $a, b \in \mathbb{N}$ we have

$$a \times b = a + ... + a \ (b \text{ times}) = b + ... + b \ (a \text{ times}).$$

EXERCISE 14.19. Define in a similar way the (finite) product $\prod_{i=1}^{n} a_i$ (which is also denoted by $a_1...a_n = a_1 \times ... \times a_n$). Prove the analogues of associativity and distributivity for sums and products of families of numbers. Define $b^a = b \times ... \times b$ (a times) for $a, b \in \mathbb{N}$ and prove that $b^{a+c} = b^a \times b^c$ and $(b^a)^c = b^{ac}$.

EXERCISE 14.20. Prove that if a is an integer and n is a natural number then

$$a^n - 1 = (a-1)(a^{n-1} + a^{n-2} + ... + a + 1).$$

Hint: Induction on n.

EXERCISE 14.21. Prove that if a is an integer and n is an integer then

$$a^{2n+1} + 1 = (a+1)(a^{2n} - a^{2n-1} + a^{2n-2} - ... - a + 1).$$

Hint: Set $a = -b$.

Exercise 14.22. Prove that a subset $A \subset \mathbb{N}$ is bounded if and only if it is finite.

Hint: We first prove by induction on b that if a set is bounded by b then it is finite. For the induction step assume A is bounded by b. If $b \notin A$ then A is bounded by $b-1$ so it is finite. If $b \in A$ then $A \setminus \{b\}$ is bounded by $b-1$ so $A \setminus \{b\}$ is finite so there is a bijection $F : A \backslash \{b\} \to \{1, ..., m\}$ and one constructs a bijection $G : A \to \{1, ..., m+1\}$ by setting $G(i) = F(i)$ for $i \leq m$ and $G(m+1) = b$. To prove that finite sets are bounded one can use again induction. A different argument would be by contradiction: assume this is false and let n be minimum natural number with the property that there is a finite subset $A \subset \mathbb{N}$ of cardinality n which is not bounded. Let $F : \{1, ..., n\} \to A$ be a bijection, $a_i = F(i)$. Then $\{a_1, ..., a_{n-1}\}$ is bounded from above by some b and conclude that A is bounded from above by either b or a_n.

Exercise 14.23. Prove that every subset of a finite set is finite.

Hint: Use the previous exercise.

Exercise 14.24. Prove that if A is a finite set and $F : A \to A$ is a map then F is injective if and only if it is surjective.

Hint: By Exercise 14.11

$$|A| = |F(A) \cup (A \backslash F(A))| = |F(A)| + |A \backslash F(A)|.$$

So if F is injective then $|A| = |F(A)|$ so $|A \backslash F(A)| = 0$ so F is surjective. Assume now F is surjective. One proves first that there is a function $G : A \to A$ such that $F \circ G = I_A$ (use either the axiom of choice, cf. Exercise 10.23, or another induction); then G is injective, hence surjective, hence bijective, which implies F bijective, hence surjective.

(Note that Cantor's original definition of finite sets is: A is finite if and only if every injective map from A to A is surjective; so what the exercise did is to show that our definition of finite sets implies Cantor's. One can also prove that Cantor's definition of finite sets implies ours; cf. Exercise 19.9.)

Definition 14.25. Let A be a set and $n \in \mathbb{N}$. Define the set A^n to be the set $Fun(\{1, ..., n\}, A)$ of all maps $\{1, ..., n\} \to A$. Call

$$A^\star = \coprod_{n \in \mathbb{N}} A^n$$

the set of *words* with letters in A. We identify A^n with subsets of A^* via the maps $x \mapsto (x, n)$, i.e., we write $(a_1, ..., a_n)$ instead of $((a_1, ..., a_n), n)$.

Definition 14.26. If $f : \{1, ..., n\} \to A$ and $f(i) = a_i$ we write f as a "tuple" $(a_1, ..., a_n)$ and sometimes as a "word" $a_1...a_n$; in other words we add to the definitions of Set Theory the following definitions

$$f = (a_1, ..., a_n) = a_1...a_n.$$

Exercise 14.27.

1) Show that the map $A \to A^1$ sending $a \mapsto f_a$, where $f_a(1) = a$, is a bijection. From now we identify A^1 with A via this bijection.

2) Show that the maps $A^n \times A^m \to A^{n+m}$,

$$((a_1, ..., a_n), (b_1, ..., b_m)) \mapsto (a_1, ..., a_n, b_1, ..., b_m)$$

(called *concatenations*), are bijections. From now on we identify $A^n \times A^m$ and A^{n+m} via these bijections. In particular A^2 is identified with $A \times A$, etc.

3) Consider the induced (non-injective) binary operation $A^\star \times A^\star \to A^\star$, $(u, v) \to uv$. Prove that $u(vw) = (uv)w$.

REMARK 14.28. This Remark involves concepts that we have not defined in this course and therefore may be skipped. Assume, in this Remark only, that we know some space geometry, in particular we know what a convex polyhedron is. (Pyramids, prisms, dodecahedra, etc., are examples of convex polyhedra.) Let Π be a convex polyhedron and denote by $V = V(\Pi)$, $E = E(\Pi)$, $F = F(\Pi)$ the sets of vertices, edges, and vertices, respectively. Then a famous theorem of Euler (which had also been independently discovered by Descartes but kept secret by him) says that

$$|V| - |E| + |F| = 2.$$

(One can check this is true, say for pyramids, prisms, dodecahedra, etc.)

We present in what follows a wrong "proof" of this theorem. The "proof" will by induction on the number of vertices and the mistake will be quite instructive.

Let $P(n)$ be the sentence that for every polyhedron Π with $|V(\Pi)| = n$ the formula above holds. We will attempt to prove by induction that for all n we have $P(n)$. Assume we have a polyhedron Π with $|V(\Pi)| = n - 1$. By the induction hypothesis $|V(\Pi)| - |E(\Pi)| + |F(\Pi)| = 2$. Now pick a face $A_1A_2...A_k$ of Π and add a vertex A_{k+1} to Π close to that face. We obtain a new polyhedron Π^+ with $|V(\Pi^+)| = |V(\Pi)| + 1 = n - 1 + 1 = n$ vertices, $|E(\Pi^+)| = |E(\Pi)| + k$ edges (the new edges are $VA_1, ..., VA_k$) and $|F(\Pi^+)| = |F(\Pi)| + k - 1$ faces (because the face $A_1A_2...A_k$ of Π ceases to be a face in Π^+ but Π^+ has acquired new faces $A_{k+1}A_1A_2, ..., A_{k+1}A_kA_1$). We get that

$$\begin{aligned} |V(\Pi^+)| - |E(\Pi^+)| + |F(\Pi^+)| &= |V(\Pi)| + 1 - |E(\Pi)| - k + |F(\Pi)| + k - 1 \\ &= |V(\Pi)| - |E(\Pi)| + |F(\Pi)| \\ &= 2, \end{aligned}$$

and $P(n)$ follows.

The mistake in the above "proof" is that the above did not prove $P(n)$ for an arbitrary polyhedron with n vertices: it only proved $P(n)$ for polyhedra Π^+ which can be obtained from some polyhedron Π with $n - 1$ vertices by adding a vertex close to some face of Π. But clearly an arbitrary polyhedron with n vertices is not always obtained this way: for instance a cube cannot be obtained in this way from a polyhedron with 7 vertices. The mistake is similar to the one in the wrong proof presented in Example 14.8.

One can ask for a correct proof of Euler's theorem. This will not be done in the present course. But one can ask if a proof of $\forall n P(n)$ above can be achieved by induction assuming some knowledge of geometry. A correct attempt would be to start with an arbitrary polyhedron Σ with $|V(\Sigma)| = n$, remove a vertex (and the edges adjacent to the vertex) and try deform the remaining configuration until one obtains a new polyhedron Σ^- with $n - 1$ vertices, apply the induction hypothesis to Σ^- to get a certain equality of numbers and try use that equality to prove $P(n)$ for Σ. Such an attempt would be correct but runs into serious difficulties some related to the possibility of deforming the configuration (try again the example of the cube). A way to see the difficulty is to take into consideration the fact that Euler's formula actually fails for some non-convex polyhedra so convexity needs to be used at some point in the proof. Bottom line: one needs a better approach and this is done in a chapter of Mathematics called *Topology* (but not in the Topology Chapter of our course).

CHAPTER 15

Fractions

With the integers at our disposal one can use the axioms of Set Theory to construct a whole array of familiar sets of numbers such as the rationals, the reals, the imaginaries, etc. We start here with the rationals. More generally we construct fields of fractions for every integral domain. Recall that an integral domain is a ring such that the product of every two non-zero elements is non-zero.

DEFINITION 15.1. Let R be an integral domain. For every $a, b \in R$ with $b \neq 0$ define the *fraction* $\frac{a}{b}$ (also written as a/b) to be the set of all pairs (c, d) with $c, d \in R$, $d \neq 0$ such that $ad = bc$. Denote by $K := Frac(R)$ the set of all fractions. (For $R = \mathbb{Z}$ we write $\mathbb{Q} = Frac(\mathbb{Z})$ and we call the latter the set of *rational numbers*.) So

$$\frac{a}{b} = a/b = \{(c, d) \in R \times R \mid d \neq 0, ad = bc\},$$

$$Frac(R) = \{\frac{a}{b} \mid a, b \in R, b \neq 0\}.$$

REMARK 15.2. Note that the meaning of the notation $\frac{a}{b}$ used here has nothing to do at this point with the meaning of the same notation introduced in Definition 12.14. However we will soon see that the two meanings for this notation coincide in our case.

EXAMPLE 15.3.

$$\frac{6}{10} = \{(6, 10), (-3, -5), (9, 15), ...\} \in \mathbb{Q}.$$

EXERCISE 15.4. Prove that $\frac{a}{b} = \frac{c}{d}$ if and only if $ad = bc$.

Hint: Assume $ad = bc$ and let us prove that $\frac{a}{b} = \frac{c}{d}$. We need to show that $\frac{a}{b} \subset \frac{c}{d}$ and that $\frac{c}{d} \subset \frac{a}{b}$. Now if $(x, y) \in \frac{a}{b}$ then $xb = ay$; hence $xbd = ayd$. Since $ad = bc$ we get $xbd = bcy$. Hence $b(xd - cy) = 0$. Since $b \neq 0$ we have $xd - cy = 0$ hence $xd = cy$ hence $(x, y) \in \frac{c}{d}$. We proved that $\frac{a}{b} \subset \frac{c}{d}$. The other inclusion is proved similarly. So the equality $\frac{a}{b} = \frac{c}{d}$ is proved. Conversely if one assumes $\frac{a}{b} = \frac{c}{d}$ one needs to prove $ad = bc$; we leave this to the reader.

EXERCISE 15.5. On the set $A = R \times (R \backslash \{0\})$ one can consider the relation: $(a, b) \sim (c, d)$ if and only if $ad = bc$. Prove that $\sim$ is an equivalence relation. Then observe that $\frac{a}{b}$ is the equivalence class

$$\widehat{(a, b)}$$

of (a, b). Also observe that $Frac(R) = A/\sim$ is the quotient of A by the relation $\sim$.

EXERCISE 15.6. Prove that the map $A \to Frac(R)$, $a \mapsto \frac{a}{1}$ is injective.

Definition 15.7. By abuse we identify $a \in R$ with $\frac{a}{1}$ and write $\frac{a}{1} = a$; this identifies R with a subset of $Frac(R)$. Such identifications are very common and will be done later in similar contexts.

Definition 15.8. Define binary operations $+$ and $\times$ on $K := Frac(R)$ by the formulae $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$.

Note that the above definition is correct only if one shows that there exist maps $F : K \times K \to K$ and $G : K \times K \to K$ with the property that for all $\frac{a}{b}, \frac{c}{d} \in K$ we have

$$F\left(\frac{a}{b}, \frac{c}{d}\right) = \frac{ad + bc}{bd}$$

and

$$G\left(\frac{a}{b}, \frac{c}{d}\right) = \frac{ac}{bd}.$$

Exercise 15.9. Show that there exist maps F and G as above, i.e., show that the sets

$$\{\left(\left(\frac{a}{b}, \frac{c}{d}\right), \frac{ad + bc}{bd}\right) \in (K \times K) \times K \mid a, b, c, d \in \mathbb{Z}, bd \neq 0\}$$

and

$$\{\left(\left(\frac{a}{b}, \frac{c}{d}\right), \frac{ac}{bd}\right) \in (K \times K) \times K \mid a, b, c, d \in \mathbb{Z}, bd \neq 0\}$$

are maps $K \times K \to K$.

Hint: To check this it is necessary and sufficient to check that if $\frac{a}{b} = \frac{a'}{b'}$, $\frac{c}{d} = \frac{c'}{d'}$ then $\frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'}$ and $\frac{ac}{bd} = \frac{a'c'}{b'd'}$.

Remark 15.10. To see why we need to check the correctness of Definition 15.8 let us look at the following example. We claim, for instance, there exists no map $F : \mathbb{Q} \to \mathbb{Z}$ such that for all $\frac{a}{b} \in \mathbb{Q}$ we have

$$F\left(\frac{a}{b}\right) = a.$$

Indeed saying that such a map exists is equivalent to the claim that the set

$$F = \{\left(\frac{x}{y}, z\right) \in \mathbb{Q} \times \mathbb{Z} \mid x = z\}$$

is a map $\mathbb{Q} \to \mathbb{Z}$. But note that

$$\left(\frac{3}{5}, 3\right), \left(\frac{6}{10}, 6\right) \in F, \quad \frac{3}{5} = \frac{6}{10}, \quad 3 \neq 6,$$

so F is not a map. Alternatively, if a map F as above existed we would have

$$3 = F\left(\frac{3}{5}\right) = F\left(\frac{6}{10}\right) = 6,$$

a contradiction. This example suggests that the existence of F and G in Exercise 15.9 is far from being automatic.

Exercise 15.11. Let $K = Frac(R)$.

1) Prove that K (with the operations $+$ and $\times$ defined above and with the elements $0, 1$) is a field. It is called the *field of fractions* of R.

2) Check that for $a, b \in R$ we have that $\frac{a}{b}$ in the sense of Definition 12.14 equals $\frac{a}{b}$ in the sense of Definition 15.1.

Definition 15.12. Assume R is an ordered ring (e.g., $\mathbb{Q}$). For $\frac{a}{b}, \frac{c}{d} \in \mathbb{K} := Frac(R)$ with $b, d > 0$ write $\frac{a}{b} \leq \frac{c}{d}$ if $ad - bc \leq 0$. Also write $\frac{a}{b} < \frac{c}{d}$ if $\frac{a}{b} \leq \frac{c}{d}$ and $\frac{a}{b} \neq \frac{c}{d}$.

Exercise 15.13. Prove that K above, equipped with $\leq$ is an ordered ring (ordered field) but $K_{\geq 0}$ is not a well ordered set.

Exercise 15.14. Prove by induction the following equalities in $\mathbb{Q}$:

$$1 + 2 + ... + n = \frac{n(n+1)}{2}.$$

$$1^2 + 2^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

$$1^3 + 2^3 + ... + n^3 = \frac{n^2(n+1)^2}{4}.$$

More generally Bernoulli (17th century) studied sums of the form

$$1^k + 2^k + \ldots + n^k$$

for arbitrary $k \in \mathbb{N}$; he discovered formulae for them (involving the so-called *Bernoulli polynomials*) which turned out to play a role in many mathematical contexts.

Exercise 15.15. Prove that there is no $x \in \mathbb{Q}$ such that $x^2 = 2$.

Hint: Assume there exists $x \in \mathbb{Q}$ such that $x^2 = 2$ and seek a contradiction. Let $a \in \mathbb{N}$ be minimal with the property that $x = \frac{a}{b}$ for some b. Now $\frac{a^2}{b^2} = 2$ hence $2b^2 = a^2$. Hence a^2 is even. Hence a is even (because if a were odd then a^2 would be odd). Hence $a = 2c$ for some integer c. Hence $2b^2 = (2c)^2 = 4c^2$. Hence $b^2 = 2c^2$. Hence b^2 is even. Hence b is even. Hence $b = 2d$ for some integer d. Hence $x = \frac{2c}{2d} = \frac{c}{d}$ and $c < a$. This contradicts the minimality of a which ends the proof.

Remark 15.16. The above proof is probably one of the "first" proofs by contradiction in the history of Mathematics; this proof appears, for instance, in Aristotle, and it is believed to have been discovered by the Pythagoreans. The impossibility of solving $x^2 = 2$ for $x \in \mathbb{Q}$ was translated by the Greeks as evidence that arithmetic is insufficient to control geometry ($\sqrt{2}$ is the length of the diagonal of a square with side 1) and arguably created the first crisis in the history of Mathematics, leading to a separation of algebra and geometry that lasted until Fermat and Descartes.

CHAPTER 16

Combinatorics

Combinatorics is about counting elements in (i.e., finding cardinalities of) finite sets. The origins of combinatorics are in the work of Pascal, Jakob Bernoulli, and Leibniz; these origins are intertwined with the origins of probability theory and the early development of calculus.

DEFINITION 16.1. For $n \in \mathbb{N}$ define the *factorial* of n (read n factorial) by

$$n! = 1 \times 2 \times ... \times n \in \mathbb{N}.$$

Also set $0! = 1$.

DEFINITION 16.2. For $0 \leq k \leq n$ in $\mathbb{Z}$ define the *binomial coefficient*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \in \mathbb{Q}.$$

One also reads this "n choose k."

EXERCISE 16.3. Prove that

$$\binom{n}{k} = \binom{n}{n-k}$$

and

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n.$$

EXERCISE 16.4. Prove that

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

Hint: Direct computation with the definition.

EXERCISE 16.5. Prove that

$$\binom{n}{k} \in \mathbb{Z}.$$

Hint: Proceed by induction on n; use Exercise 16.4.

EXERCISE 16.6. For every a, b in any ring we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Here if c is in a ring R and $m \in \mathbb{N}$ then $mc = c + ... + c$ (m times).

Hint: Induction on n and use Exercise 16.4.

EXERCISE 16.7. (Subsets) Prove that if $|A| = n$ then $|\mathcal{P}(A)| = 2^n$. (A set with n elements has 2^n subsets.)

Hint: Induction on n; if $A = \{a_1, ..., a_{n+1}\}$ use

$$\mathcal{P}(A) = \{B \in \mathcal{P}(A) \mid a_{n+1} \in B\} \cup \{B \in \mathcal{P}(A) \mid a_{n+1} \notin B\}.$$

EXERCISE 16.8. (Combinations) Let A be a set with $|A| = n$, let $0 \leq k \leq n$, and set

$$\mathrm{Comb}(k, A) = \{B \in \mathcal{P}(A) \mid |B| = k\}.$$

Prove that

$$|\mathrm{Comb}(k, A)| = \begin{pmatrix} n \\ k \end{pmatrix}.$$

In other words a set of n elements has exactly $\begin{pmatrix} n \\ k \end{pmatrix}$ subsets with k elements. A subset of A having k elements is called a *combination* of k elements from the set A.

Hint: Proceed by induction on n. If $A = \{a_1, ..., a_{n+1}\}$ use Exercise 16.4 plus the fact that $\mathrm{Comb}(k, A)$ can be written as

$$\{B \in \mathcal{P}(A) \mid |B| = k, a_{n+1} \in B\} \cup \{B \in \mathcal{P}(A) \mid |B| = k, a_{n+1} \notin B\}.$$

EXERCISE 16.9. (Permutations) For a set A let $\mathrm{Perm}(A) \subset Fun(A, A)$ be the set of all bijections $F : A \to A$. A bijection $F : A \to A$ is also called a *permutation*. Prove that if $|A| = n$ then

$$|\mathrm{Perm}(A)| = n!.$$

So the exercise says that a set of n elements has $n!$ permutations.

Hint: Let $|A| = |B| = n$ and let $\mathrm{Bij}(A, B)$ be the set of all bijections $F : A \to B$; it is enough to show that $|\mathrm{Bij}(A, B)| = n!$. Proceed by induction on n; if $A = \{a_1, ..., a_{n+1}\}$, $B = \{b_1, ..., b_{n+1}\}$ then use the fact that

$$\mathrm{Bij}(A, B) = \bigcup_{k=1}^{n+1} \{F \in \mathrm{Bij}(A, B) \mid F(a_1) = b_k\}.$$

For $d \in \mathbb{N}$ and X a set let X^d be the set of all maps $\{1, ..., d\} \to X$. We identify a map $i \mapsto a_i$ with the tuple $(a_1, ..., a_d)$.

EXERCISE 16.10. (Combinations with repetition) Let

$$\mathrm{Combrep}(k, n) = \{(x_1, ..., x_n) \in \mathbb{Z}^n \mid x_i \geq 0,\ x_1 + ... + x_n = k\}.$$

Prove that

$$|\mathrm{Combrep}(k, n)| = \begin{pmatrix} k + n - 1 \\ n - 1 \end{pmatrix}.$$

Hint: Let $A = \{1, ..., k + n - 1\}$. Prove that there is a bijection

$$\mathrm{Comb}(n - 1, A) \to \mathrm{Combrep}(k, n).$$

The bijection is given by attaching to every subset

$$\{i_1, ..., i_{n-1}\} \subset \{1, ..., k + n - 1\},$$

where $i_1 < ... < i_{n-1}$, the tuple $(x_1, ..., x_n)$ where

1) $x_1 = |\{i \in \mathbb{Z} \mid 1 \leq i < i_1\}|$,
2) $x_j = |\{i \in \mathbb{Z} \mid i_j < i < i_{j+1}\}|$, for $2 \leq j \leq n - 1$, and
3) $x_n = |\{i \in \mathbb{Z} \mid i_{n-1} < i \leq k + n - 1\}|$.

Exercise 16.11. (Arrangements). Let A be a set with $|A| = n$, let $0 \le k \le n$, and let $\mathrm{Arr}(k, A)$ be the subset of $A^k = A \times ... \times A$ (k times) consisting of all tuples $(a_1, ..., a_k)$ with $a_1, ..., a_k \in A$ distinct. (An element of $\mathrm{Arr}(k, A)$ is called an *arrangement* of k elements of A.)

1) Prove that there is a bijection

$$\mathrm{Arr}(n, A) \to \mathrm{Perm}(A).$$

2) Prove that

$$|\mathrm{Arr}(k, A)| = \frac{n!}{(n-k)!}.$$

Hint for 1: Fix a bijection $\sigma : A \to \{1, ..., n\}$ and consider the map that sends an arrangement $(a_1, ..., a_n)$ into the bijection $A \to A$ that sends each $x \in A$ into $a_{\sigma(x)}$.

Hint for 2: Consider the map $F : \mathrm{Arr}(k, A) \to \mathrm{Comb}(k, A)$ defined by

$$F(a_1, ..., a_k) = \{a_1, ..., a_k\}$$

and use Exercise 14.14.

Hint for an alternative proof of 2: Proceed by induction on k by considering the map $F : \mathrm{Arr}(k, A) \to \mathrm{Arr}(k-1, A)$ defined by

$$F(a_1, ..., a_k) = (a_1, ..., a_{k-1})$$

and use Exercise 14.14. Note that if we use this alternative proof of 2 plus Exercise 14.14 then one can give an alternative proof for Exercise 16.8.

Exercise 16.12. (Bose-Einstein state space). Let $A = \{1, \dots, n\}$ for some $n \in \mathbb{N}$, let $k \in \mathbb{N}$, and let $\sim$ be the equivalence relation on $A^k = A \times ... \times A$ (k times) such that

$$(a_1, ..., a_k) \sim (b_1, ..., b_k)$$

if and only if there exists a bijection $\sigma : \{1, ..., k\} \to \{1, ..., k\}$ such that for all $i \in \{1, ..., k\}$ we have $b_i = a_{\sigma(i)}$. Find the cardinality of $A^k/\sim$.

(In applications $A^k/\sim$, is taken to be the Bose-Einstein space of "all possible states for a system of k bosons (e.g. of photons) each of which can be in one of n possible states." Bosons are "indistinguishable" and this is why we take the quotient of A^k by $\sim$. By the way the set A^k is referred to as the Maxwell-Boltzman state space of all possible states for k distinguishable particles (such as molecules in a gas) each of which can be in one of n possible states.)

Hint: Show that the set $A^k/\sim$ is in bijection with the set $\mathrm{Combrep}(k, n)$ of all tuples $(x_1, ..., x_n)$ of integers $x_i \ge 0$ such that $x_1 + ... + x_n = k$ as follows: to the equivalence class of $(a_1, ..., a_k) \in A^k$ one attaches the tuple $(x_1, ..., x_n) \in \mathbb{Z}^n$ where $x_i \ge 0$ is the number of times i appears among $a_1, ..., a_k$, i.e.,

$$x_i = |\{j \in \{1, ..., k\} \mid a_j = i\}|.$$

Exercise 16.13. (Fermi-Dirac state space). Let $A = \{1, \dots, n\}$ for some $n \in \mathbb{N}$, let $k \in \mathbb{N}$, $k \le n$, recall that $\mathrm{Arr}(k, A) \subset A^k$ denotes the set of all tuples $(a_1, ..., a_k) \in A^k$ such that $a_1, ..., a_k$ are distinct and let $\sim$ be the equivalence relation on $\mathrm{Arr}(k, A)$ induced by the one in the previous exercise. Find the cardinality of $\mathrm{Arr}(k, A)/\sim$.

(In applications $\mathrm{Arr}(k, A)$ is taken to be the Fermi-Dirac space of "all possible states for a system of k fermions (e.g. electrons) each of which can be in one of n possible states". Fermions are indistinguishable, hence the quotient by $\sim$, but also, in addition, they satisfy the Pauli exclusion principle according to which no two fermions can be in the same state.)

Hint: Show that the set $\mathrm{Arr}(k, A)/\sim$ is in bijection with the set $\mathrm{Comb}(k, A)$ of all subsets of A of cardinality k as follows: to the equivalence class of $(a_1, ..., a_k) \in \mathrm{Arr}(k, A)$ one attaches the set $\{a_1, ..., a_k\}$.

CHAPTER 17

Probability

In this chapter K is an ordered field and $K_{\geq 0} := \{a \in K \mid a \geq 0\}$. Note that there is a unique ring homomorphism $\mathbb{Q} \to K$ which we shall view as an inclusion. The main examples for K will be $\mathbb{Q}$ itself and, later, the field $\mathbb{R}$ of real numbers.

DEFINITION 17.1. Let Ω be a finite set. A *probability measure* on Ω is a map $\mu : \mathcal{P}(\Omega) \to K_{\geq 0}$ satisfying the property that $\mu(\Omega) = 1$ and for every $B, C \in \mathcal{P}(\Omega)$ with $B \cap C = \emptyset$ we have

$$\mu(B \cup C) = \mu(B) + \mu(C).$$

If $\Omega = \{x_1, ..., x_n\}$ and $\mu(\{x_i\}) = p_i \in K_{\geq 0}$ then the family (p_j) is called the *probability distribution* of μ and satisfies $\sum_{j=1}^{n} p_j = 1$. We say that $B, C \in A$ are independent if $\mu(B \cap C) = \mu(B) \cdot \mu(C)$.

Clearly the probability distribution of μ determines μ uniquely.

DEFINITION 17.2. A *random variable* is a function $f : \Omega \to K$. If $\Omega = \{x_1, ..., n\}$ with x_i distinct and $f(\Omega) = \{y_1, ..., y_m\}$ with y_i distinct and if $A_i := f^{-1}(y_i)$ then the *mean value* (or *expectation value*) of f is defined as

$$\mathbb{E}(f) := \mathbb{E}_{\mu}(f) := \sum_{i=1}^{m} y_i \cdot \mu(A_i) = \sum_{j=1}^{n} f(x_j) p_j =: \int f d\mu.$$

EXERCISE 17.3. Prove that for f, A_i, and y_i as above we have

$$f = \sum_{i=1}^{m} y_i \cdot 1_{A_i}.$$

EXERCISE 17.4. Prove that for a random variable f as above if $(C_k)_{k \in \{1,...,N\}}$ is a partition of Ω such that for each k there is a $z_k \in K$ with $f(C_k) = \{z_k\}$ (with z_k not necessarily distinct) then

$$\mathbb{E}(f) = \sum_{k=1}^{N} z_k \mu(C_k).$$

EXERCISE 17.5. Prove that

$$\mathbb{E}(1_B) = \mu(B).$$

EXERCISE 17.6. Prove that

$$|\mathbb{E}(f)| \leq \mathbb{E}(|f|).$$

EXERCISE 17.7. (Cauchy-Schwarz). Prove that

$$(\mathbb{E}(|fg|))^2 \leq \mathbb{E}(f^2) \cdot \mathbb{E}(g^2).$$

Definition 17.8. Two random variables f and g are called *independent* if upon writing $f = \sum y_i \cdot 1_{A_i}$ and $g = \sum_j z_j \cdot 1_{B_j}$ with y_i distinct and z_j distinct, $A_i = f^{-1}(y_i)$, $B_j = g^{-1}(z_j)$, we have that A_i and B_j are independent for all i and j, i.e.,

$$\mu(A_i \cap B_j) = \mu(A_i) \cdot \mu(B_j).$$

Exercise 17.9. Prove that for random variables f and g and $c \in K$ we have

$$\mathbb{E}(f+g) = \mathbb{E}(f) + \mathbb{E}(g), \quad \mathbb{E}(cf) = c \cdot \mathbb{E}(f).$$

If in addition f and g are independent then

$$\mathbb{E}(fg) = \mathbb{E}(f) \cdot \mathbb{E}(g).$$

Hint: Additivity is clear. If in addition f and g are independent then

$$\begin{aligned}
\mathbb{E}(fg) &= \textstyle\sum_{ij} y_i \cdot z_j \cdot \mu(A_i \cap B_j) \\
&= \textstyle\sum_{ij} y_i \cdot z_j \cdot \mu(A_i) \cdot \mu(B_j) \\
&= \mathbb{E}(g) \cdot \mathbb{E}(g).
\end{aligned}$$

Here we used Exercise 17.4 as $y_i z_j$ are not necessarily distinct.

Exercise 17.10. Generalize the definition of independence for more than 2 random variables and prove the corresponding property for the expectation value of a product.

Definition 17.11. The *variance* of a random variable f is defined as

$$\mathbb{V}(f) := \mathbb{V}_\mu(f) := \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2.$$

Exercise 17.12.
1) Check the last equality in the equation above.
2) Prove that if f and g are independent then

$$\mathbb{V}(f+g) = \mathbb{V}(f) + \mathbb{V}(g).$$

Hint for 2): One checks that $f-a$ and $g-b$ are still independent for all $a, b \in K$. We have

$$\begin{aligned}
\mathbb{V}(f+g) &= \mathbb{E}(((f - \mathbb{E}(f)) + (g - \mathbb{E}(g)))^2) \\
&= \mathbb{V}(f) + \mathbb{V}(g) + 2\mathbb{E}((f - \mathbb{E}(f))(g - \mathbb{E}(g))) \\
&= \mathbb{V}(f) + \mathbb{V}(g) + 2\mathbb{E}(f - \mathbb{E}(f))\mathbb{E}(g - \mathbb{E}(g)) \\
&= \mathbb{V}(f) + \mathbb{V}(g) + 2(\mathbb{E}(f) - \mathbb{E}(f))(\mathbb{E}(g) - \mathbb{E}(g)) \\
&= \mathbb{V}(f) + \mathbb{V}(g).
\end{aligned}$$

Exercise 17.13. (Chebyshev's inequalities).
1) Prove that if $f \geq 0$ is a random variable then for all $\epsilon > 0$,

$$\mu(\{x \in \Omega \mid f(x) \geq \epsilon\}) \leq \frac{\mathbb{E}(f)}{\epsilon}.$$

2) Prove that if f is a random variable then for all $\epsilon > 0$,

$$\mu(\{x \in \Omega \mid |f(x) - \mathbb{E}(f)| \geq \epsilon\}) \leq \frac{\mathbb{V}(f)}{\epsilon^2}.$$

Hint: 2) follows from 1) because

$$\{x \in \Omega \mid |f(x) - \mathbb{E}(f)| \geq \epsilon\} = \{x \in \Omega \mid (f(x) - \mathbb{E}(f))^2 \geq \epsilon^2\}.$$

To check 1) note that

$$f = f \cdot 1_{\{x|f(x)\geq\epsilon\}} + f \cdot 1_{\{x|f(x)<\epsilon\}} \geq f \cdot 1_{\{x|f(x)\geq\epsilon\}} \geq \epsilon \cdot 1_{\{x|f(x)\geq\epsilon\}}$$

Taking $\mathbb{E}$ we get

$$\mathbb{E}(f) \geq \epsilon \cdot \mathbb{E}(1_{\{x|f(x)\geq\epsilon\}}) = \epsilon \cdot \mu(\{x|f(x) \geq \epsilon\}).$$

Exercise 17.14. (Bernoulli's Law of Large Numbers). Let $f_1, ..., f_n : \Omega \to \{0, 1\}$ be independent random variables and $p, q \in K_{>0}$, $p + q = 1$, such that for all i we have

$$\mu(f_i^{-1}(1)) = p, \quad \mu(f_i^{-1}(0)) = q.$$

Let $s_n = f_1 + ... + f_n$. Prove that

1) $\mathbb{E}(s_n/n) = p$.

2) $\mathbb{V}(f_1) = ... = \mathbb{V}(f_n) = pq$.

3) $\mathbb{V}(s_n) = npq$.

4) $\mu(\{x \mid |\frac{s_n(x)}{n} - p| \geq \epsilon\}) \leq \frac{1}{4n\epsilon^2}$ for $\epsilon > 0$.

Hint: 1) follows from the additivity of $\mathbb{E}$ plus $\mathbb{E}(f_i) = p$. 3) follows from 2) plus the additivity of $\mathbb{V}$ for independent random variables. 2) follows from the computation

$$\mathbb{V}(f_i) = \mathbb{E}((f_i - \mathbb{E}(f_i))^2) = \mathbb{E}((f_i - p)^2) = (1-p)^2 p + p^2 q = q^2 p + p^2 q = pq(p+q) = pq.$$

4) follows from the Chebyshev inequality:

$$\mu(\{x \mid |\frac{s_n(x)}{n} - p| \geq \epsilon\}) \leq \frac{\mathbb{V}(s_n/n)}{\epsilon^2} = \frac{\mathbb{V}(s_n)}{n^2\epsilon^2} = \frac{npq}{n^2\epsilon^2} = \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

Exercise 17.15. Assume Ω is a non-empty finite set. Prove that the map $\mu : \mathcal{P}(\Omega) \to \mathbb{Q}_{\geq 0} \subset K_{\geq 0}$ defined by

$$\mu(B) = \frac{|B|}{|\Omega|}$$

is a probability measure. (It is called the *uniform probability measure* on Ω and it corresponds to the probability distribution (p_i) with $p_i = 1/|\Omega|$. For all B, $\mu(B)$ is called the *probability* (or measure) of X. If $B = \{x \in \Omega \mid P(x)\}$ where $P(x)$ is a formula with one free variable we say that $\mu(B)$ is the probability that P holds.)

Exercise 17.16. Assume Ω is a non-empty finite set and μ is the uniform probability measure on it.

1) Prove that if $B = C \neq \Omega, \emptyset$ then B and C are not independent.

2) Prove that if $B \cap C = \emptyset$ and $B \neq \emptyset$, $C \neq \emptyset$ then B and C are not independent.

3) Prove that if $\Omega = \Omega_1 \times \Omega_2$, $B = B_1 \times \Omega_2$, $C = \Omega_1 \times C_2$, $B_1 \subset \Omega_1$, $C_2 \subset \Omega_2$, then B and C are independent.

REMARK 17.17. Bernoulli's Law of Large Numbers applies, for instance to the following situation. We let $\Omega = \{0,1\}^n$, we let μ be the uniform probability measure on Ω, and we let $f_i : \Omega \to \mathbb{Q}$ be defined by

$$f_i(x_1, ..., x_n) = x_i.$$

Then f_i are independent (check!) with $\mu(f_i^{-1}(1)) = 1/2 =: p$ and

$$s_n(x_1, ..., x_n) = x_1 + ... + x_n.$$

The Bernoulli's Law gives (check!)

$$\sum_{\{k \ | \ |(k/n - 1/2| \geq \epsilon\}} 2^{-n} \left(\begin{array}{c} n \\ k \end{array} \right) \leq \frac{1}{4n\epsilon^2}.$$

The number $2^{-n} \left(\begin{array}{c} n \\ k \end{array} \right)$ is the probability that when tossing a coin n times one gets k tails. The inequality above says that "when ϵ is fixed and n grows the probability that s_n/n deviates from $1/2$ by more than ϵ approaches 0."

EXAMPLE 17.18. (Dice). Let $A = \{1, ..., 6\}$ and $\Omega = A \times A$. Let $X \subset \Omega$ be defined by $X = \{(1,3), (3,1)\}$ so $|\Omega| = 36$, $|X| = 2$, hence for μ the uniform probability measure on Ω we have

$$\mu(X) = \frac{2}{36} = \frac{1}{18}.$$

(In applications A is called the set of possible outcomes (or states) for one dice (or molecule in a classical gas, say). Ω is called the set of possible outcomes (or states) for a pair of dice (or pair of molecules in a classical gas). Note that $(1,3)$ and $(3,1)$ count as different outcomes: the dice have "individuality" in that they are assumed to be not identical. Assuming all outcomes are equiprobable (total ignorance as to whether the dice are fair) the probability that the outcome is in X is by definition $1/18$.)

EXAMPLE 17.19. (Bosons). Let $A = \{1, ..., 6\}$ and

$$\Omega = (A \times A)/ \sim$$

where $\sim$ is the smallest equivalence relation on $A \times A$ such that for all a and b we have $(a,b) \sim (b,a)$. Let $X \subset \Omega$ be defined by $X = \{\widehat{(1,3)}\}$ and $\Omega = \{\widehat{(a,b)} \mid a, b \in A\}$. Note that $|X| = 1$ and $|\Omega| = 1 + 2 + ... + 5 = \frac{5 \times 6}{2} = 15$. Hence for μ the uniform probability measure on Ω we have

$$\mu(X) = \frac{1}{15}.$$

(In applications to quantum physics A could be the set of possible states for one boson that has a fixed energy; this is not realistic for the number 6 does not really appear in the theory. Ω is taken by definition to be the set of possible states for a pair of bosons: unlike dice which "have individuality" the two bosons are assumed to have "no individuality" so the pairs in $(1,3)$ and $(3,1)$ in $A \times A$ correspond to one and the same state in Ω. Assuming that all states in Ω are equiprobable (total ignorance apart from the "no individuality" assumption under the assumption that the energy is constant) the probability is by definition $1/15$.

EXAMPLE 17.20. (Fermions). Let

$$A = \{1, ..., 6\}, \quad \mathrm{Arr}(2, A) = \{(a, b) \mid a, b \in A, a \neq b\} \subset A \times A, \quad \Omega = \mathrm{Arr}(2, A)/\sim,$$

where $\sim$ is the smallest equivalence relation on $A \times A$ such that for all a and b we have $(a, b) \sim (b, a)$. Let $X \subset \Omega$ be defined by $X = \{\widehat{(1,3)}\}$ and $\Omega = \{\widehat{(a,b)} \mid a, b \in A, a \neq b\}$. Note that $|X| = 1$ and $|A| = 1+2+3+4 = 10$. Hence for μ the uniform probability measure on Ω we have

$$\mu(X) = \frac{1}{10}.$$

(In applications to quantum physics A could be the set of possible states for one fermion with fixed energy. Ω is the set of possible states for a pair of fermions with fixed energy: again, unlike dice which "have individuality" the two fermions have "no individuality" so $(1, 3)$ and $(3, 1)$ correspond again to the same state in Ω; however, unlike bosons, no two fermions can be in the same state (Pauli exclusion principle) so the state $(2, 2)$ is not allowed. Assuming that all states in Ω are equiprobable (fixed energy and total ignorance apart from the "no individuality" assumption and the "Pauli exclusion principle") the probability is by definition $1/10$.

EXAMPLE 17.21. We continue to denote by K an ordered field. In statistical physics one starts with a set Ω (which we still assume, for now, to be finite and which is identified with the state space for a physical system S) and one is given a surjective function $H : \Omega \to \mathcal{E}$ (referred to as *Hamiltonian* or *energy*) where $\mathcal{E} \subset K$ is referred to as the *energy spectrum*. One also fixes a function $e : K \to K_{>0}$ (to be identified later) which, for normalization purposes, is assumed to satisfy $e(0) = 1$. One then defines the *canonical probability measure* μ_{can} on Ω, attached to (Ω, H), by letting

$$\mu_{\mathrm{can}}(\{x\}) = \frac{1}{|H^{-1}(H(x))|} \cdot \frac{e(H(x))}{Z}, \quad \text{where} \quad Z = \sum_{E \in \mathcal{E}} e(E), \quad x \in \Omega.$$

Z is called the *partition function*. Note that

$$\mu_{\mathrm{can}}(H^{-1}(E)) = \frac{e(E)}{Z}.$$

The canonical probability measure and the uniform probability measure on Ω coincide, of course, if the energy function H is a constant function. Finally recall the *expectation value* $\mathbb{E}(H)$ of H,

$$\mathbb{E}(H) := \sum_{E \in \mathcal{E}} E \cdot \frac{e(E)}{Z}.$$

In order to pin down the function e we impose the following "compatibility with sums". To explain this consider two finite sets Ω_1, Ω_2 equipped with "energy functions"

$$H_i : \Omega_i \to \mathcal{E}_i, \quad i = 1, 2.$$

We define the *sum* of (Ω_1, H_1) and (Ω_2, H_2) to be the pair (Ω, H) where $\Omega := \Omega_1 \times \Omega_2$ and H is the function

$$H : \Omega \to \mathcal{E}_1 + \mathcal{E}_2, \quad H(x_1 + x_2) = H_1(x_1) + H_2(x_2)$$

(referred to as the *total energy*). We say that (Ω_1, H_1) and (Ω_2, H_2) are *independent* if the map the map

$$\mathcal{E}_1 \times \mathcal{E}_2 \to K, \quad (E_1, E_2) \mapsto E_1 + E_2$$

is injective. Let us say that the function e is *compatible with sums* if for all independent (Ω_1, H_1) and (Ω_2, H_2) and all $E_1 \in \mathcal{E}_1$, $E_2 \in \mathcal{E}_2$ if (Ω, H) is the sum of (Ω_1, H_1) and (Ω_2, H_2) and $\mu_{\text{can},1}, \mu_{\text{can},2}, \mu_{\text{can}}$ are the canonical probability measures attached to (Ω_1, H_1), (Ω_2, H_2), (Ω, H), respectively, then

$$\mu_{\text{can}}(H^{-1}(E_1 + E_2)) = \mu_{\text{can},1}(H_1^{-1}(E_1)) \cdot \mu_{\text{can},2}(H_2^{-1}(E_2)).$$

This condition puts a strong restriction on e; cf. the Exercise below.

Exercise 17.22.

1) Prove that $e : K \to K_{>0}$ is compatible with sums if and only if for all $y_1, y_2 \in K$ we have

$$e(y_1 + y_2) = e(y_1) \cdot e(y_2).$$

2) Prove that no such function (except $e(y) = 1$) exists for $K = \mathbb{Q}$.

Note: The exponential function provides examples for $K = \mathbb{R}$ (see Remark 17.24 below).

Hint for 1): The "if" part is trivial. For the "only if" part take $|\mathcal{E}_1| = 2$, $|\mathcal{E}_2| = 1$, $0 \in \mathcal{E}_1$.

Remark 17.23. In physical terms the compatibility with sums amounts to the following. We assume the physical system S is obtained by juxtaposition of two physical systems S_1 and S_2 which do not interact; the set of states of S is then the cartesian product of the sets of states of S_1 and S_2 and the energy function of the system S is the sum of the energy functions of the two systems S_1 and S_2. Then our condition asks that the probability that S has energy $E_1 + E_2$ is the product of the probabilities that S_1 has energy E_1 and S_2 has energy E_2.

Remark 17.24. This Remark involves concepts (from calculus) that we have not yet defined and therefore may be skipped and revisited after we discuss calculus. We assume the notation and setting of Example 17.21 and we assume in addition $K = \mathbb{R}$, the field of real numbers (to be defined later) and that e is smooth (cf. a definition to be given later). Then one can prove that e satisfies the condition

$$e(y_1 + y_2) = e(y_1) \cdot e(y_2) \quad \text{for all} \quad y_1, y_2 \in \mathbb{R}$$

if and only if

$$e(y) = \exp(-\beta y)$$

for some $\beta \in \mathbb{R}$, where exp is the exponential function (to be defined later). Assume this is the case. Then we have

$$\mathbb{E}(H) = -Z^{-1}\frac{\partial Z}{\partial \beta}.$$

One also defines the *Boltzman (or micro-canonical) entropy* $S : \mathcal{E} \to \mathbb{R}$ by the formula

$$S(E) = \log(|H^{-1}(E)|), \quad E \in \mathcal{E}$$

where log is the logarithm function (to be defined later).

Planck's introduction of the quantum idea (in 1900) was based on the observation that if one takes

$$\mathcal{E} = \{nh\nu \mid n \in \mathbb{Z},\ 0 \leq n \leq N\},$$

(with $N \in \mathbb{N}$, $\nu \in \mathbb{R}_{>0}$ the frequency, and $h \in \mathbb{R}$ the "Planck constant") then the *expectation value*, $\mathbb{E}(H)$, satisfies

$$\lim_{N\to\infty} \mathbb{E}(H) \to \mathbb{E}_{\text{Planck}}(H) := \frac{1}{\beta} \cdot \frac{\beta h\nu}{\exp(\beta h\nu) - 1}$$

and $\mathbb{E}_{\text{Planck}}(H)$ fits the value $\mathbb{E}_{\text{exper}}(H)$ obtained from the experimental data. Planck's work was motivated by the fact that the value $\mathbb{E}_{\text{exper}}(H)$ did not fit the theoretical value $\mathbb{E}_{\text{class}}(H)$ obtained from the classical 19th century model of statistical physics. This classical model was based on the assumption that $\mathcal{E}$ was the interval $\mathbb{R}_{\geq 0}$ (rather than a discrete subset of $\mathbb{R}_{\geq 0}$) and the corresponding partition function was

$$Z_{\text{class}} = \int_0^\infty \exp(-\beta E)dE := \lim \int_0^n \exp(-\beta E)dE = \frac{1}{\beta}$$

which gives

$$\mathbb{E}_{\text{class}}(H) = -Z_{\text{class}}^{-1}\frac{\partial Z_{\text{class}}}{\partial \beta} = \frac{1}{\beta} \gg \mathbb{E}_{\text{Planck}}(H) \sim \mathbb{E}_{\text{exper}}(H) \quad \text{for} \quad \beta h\nu \gg 1.$$

Here $\gg$ means "much bigger than." (Note that by the above β can be interpreted as the "inverse of the classical temperature.")

Planck's hypothesis on the discrete energy spectrum was an ad hoc fix to the discrepancy between classical statistical mechanics and experiment: there was no full theory behind it. The quantum theory was developed later (in the 1920s) by Heisenberg and Schrödinger.

CHAPTER 18

Graphs

Graph theory starts with some remarkable insights of Euler in the 18th century and became an important branch of "discrete mathematics" in the 20th century.

DEFINITION 18.1. A *graph* is a pair $G = (V, E)$ where V is a finite set (called the set of *vertices*) and $E \subset \mathcal{P}(V)$ is a set (called the set of *edges*) such that for all $e \in E$ we have $|e| = 2$. (In other words $E \subset \text{Comb}(2, V)$.) If $v \in e$ for $v \in V$ and $e \in E$ we say that v and e are *incident*. If $e = \{v, w\}$ we say v and w are *adjacent*. For $v \in V$ we let $\deg(v)$ (called the *degree* of v) be the cardinality of the set of all vertices that are adjacent to v. The cardinalities $|V|$ and $|E|$ are called the *order* and *size* of G, respectively

The above use of the word "order" has nothing to do with the use of the same word in phrases such as "first order logic" or "order relation" or "order of an element in a group" or "order of a group."

DEFINITION 18.2.

1) Two graphs (V, E) and (V', E') are called *equivalent* if and only if there exists a bijection $F : V \to V'$ inducing a bijection between the corresponding sets of edges, i.e. $F(e) \in E'$ and $F^{-1}(e') \in E$ for all $e \in E$, $e' \in E'$.

2) A *subgraph* of a graph (V, E) is a graph (V', E') such that $V' \subset V$ and $E' \subset V'$.

REMARK 18.3. This Remark involves concepts that we have not yet defined and therefore may be skipped. Assume, in this Remark only, some familiarity with plane geometry including the concept of curve in the plane. Then given a graph G one can identify V with a set of points in the plane and one can draw a curve between two points v and w in the plane, representing vertices, if and only if $\{v, w\} \in E$. The resulting plane configuration (of points and curves) can be said to represent the graph. Note that the curves are allowed to intersect (by which we mean "intersect in points that are not vertices"). If a configuration representing a graph G can be chosen such that the curves are pairwise non-intersecting then we say that G is a *planar graph* and the configuration is called *non-intersecting*. This definition assumes we have defined the notions of "plane" and "curve" between two points. Such a definition can be given if one assumes one knows what the field $\mathbb{R}$ of real numbers are and what a continuous map is: the plane is defined as being $\mathbb{R}^2$ and a curve between two points $v, w \in \mathbb{R}^2$ is defined as being a continuous map from the closed interval $[0, 1]$ to $\mathbb{R}$ mapping 0 and 1 into v and w, respectively. Any serious result about planar graphs requires, for its proof, some non-trivial topological arguments which lie beyond this course. Nevertheless, it is useful to use configurations representing graphs. A useful exercise, for instance, is to find plane configurations for various particular graphs such as those in the following example.

Other examples to be considered are obtained from polyhedra as follows. Given a polyhedron in space define a graph by taking the vertices of the graph to be the vertices of the polyhedron and by declaring a set of two vertices of the graph to be an edge of the graph if and only if there exists an edge of the polyhedron between the two vertices. One can prove that the graph attached to a convex polyhedron is a planar graph. It is an instructive exercise to draw plane configurations of the corresponding graphs in case the polyhedron is a tetrahedron, a cube, an octohedron, a dodecahedron, or an icosahedron; these 5 polyhedra are called the Platonic polyhedra.

EXAMPLE 18.4.
1) A graph of the form

$$(\{v_1, ..., v_k\}, \{\{v_1, v_2\}, ..., \{v_{k-1}, v_k\}\})$$

with v_i distinct and $k \geq 2$ will be denoted by $L(v_1...v_k)$ and will be called a *linear* graph.
2) A graph of the form

$$(\{v_1, ..., v_k\}, \{\{v_1, v_2\}, ..., \{v_k, v_1\}\})$$

with v_i distinct and $k \geq 3$ will be denoted by $C(v_1...v_k)$ and will be called a *cyclic* graph.
3) A graph of the form

$$(\{v_1, ..., v_k\}, \{\{v_i, v_j\} \mid 1 \leq i < j \leq k\})$$

with v_i distinct and $k \geq 1$ will be denoted by $K(v_1, ..., v_k)$ and will be called a *complete* graph.

DEFINITION 18.5. A *cycle* in a graph G is a subgraph C of G which is cyclic of order ≥ 3. A *path* in a graph G is a subgraph of G which is linear of order ≥ 2. A graph is called *connected* if and only if for every two distinct vertices v and w there exists a path containing v and w. A graph is called a *tree* if and only if it is connected and has no cycles.

EXERCISE 18.6. Let (V, E) be a graph, $V = \{v_1, ..., v_n\}$, $|E| = s$. Prove that

$$\sum_{i=1}^{n} \deg(v_i) = 2s.$$

Hint. Consider the set $A = \{(v, e) \in V \times E \mid v \in e\}$. Considering the map $A \to E$, $(v, e) \mapsto e$ we get $|A| = 2|E|$. On the other hand writing A as a disjoint union of the sets $A_i := \{(v, e) \in A \mid v = v_i\}$ we get $|A| = \sum_{i=1}^{n} \deg(v_i)$.

EXERCISE 18.7. Assume G is a tree.
1) Prove that G has a vertex of degree 1.
2) Prove that if G has order $|V| = n$ then G has size $|E| = n - 1$.

Hint for 1): Assume no such vertex exist and consider a path of maximum length.

Hint for 2): Induction on n. Remove a vertex of degree 1.

DEFINITION 18.8. By a *face enhanced graph* we will understand a triple (V, E, F) where $G = (V, E)$ is a graph and F is a set of cycles of G (called the set of *faces* of the enhanced graph). An enhanced graph will be called a *combinatorial polyhedron* if and only if every edge in E belongs to exactly two faces in F. Two combinatorial polyhedra are called *equivalent* if and only if there is a bijection between their sets of vertices which induces a bijection between their sets of edges and between their sets of faces.

EXAMPLE 18.9. Let $C = (V, E)$ be a cyclic graph. Then one can define a face enhanced graph (V, E, F) by taking $F = \{C\}$. Call such a face enhanced graph a *combinatorial polygon*. Note a combinatorial polygon is not a combinatorial polyhedron.

EXERCISE 18.10. Let (V, E, F) be a combinatorial polyhedron with $r = |F|$ faces $F = \{C_1, ..., C_r\}$ and s edges $s = |E|$. Then

$$\sum_{i=1}^{r} \deg(C_i) = 2s.$$

Hint: Argument similar to that for Exercise 18.6.

REMARK 18.11. This Remark involves concepts that we have not yet defined and therefore may be skipped. In the terminology of Remark 18.3 if we are given a non-intersecting configuration of a planar graph $G = (V, E)$ then one can define a face enhanced graph (V, E, F) by taking the faces to be the "boundaries" of the resulting " regions" of the plane. If the graph corresponds to a convex polyhedron then the faces of the corresponding face enhanced graph correspond to the faces of the convex polyhedron and our face enhanced graph is in fact a combinatorial polyhedron.

DEFINITION 18.12. Let (V, E, F) be a face enhanced graph, set $G = (V, E)$, and let $C \in F$, $C = C(v_1...v_k)$. Let $m \geq 0$ and consider a set $W = \{w_1, ..., w_m\}$ of cardinality m disjoint from V. (For $m = 0$ we take $W = \emptyset$.) Also let $1 \leq i < j \leq k$ and for $m = 0$ assume that $j \neq i + 1$, $(i, j) \neq (1, k)$, and $\{v_i, v_j\} \notin E$. Let L be the linear graph

$$L = (V_L, E_L) = L(v_i w_1 ..., w_m v_j)$$

(where here and later, for $m = 0$, we omit $w_1, ..., w_m$). Consider the cyclic graphs

$$C_1 = C(v_1 ..., v_i w_1 ..., w_m v_j ... v_k), \quad C_2 = C(v_i ... v_j w_m ... w_1).$$

We construct a new face enhanced graph (V', E', F') as follows. We let

$$V' = V \cup V_L, \quad E' = E \cup E_L,$$

and we let

$$F' = F \cup \{C_1, C_2\} \quad \text{if} \quad G = C$$

and

$$F' = (F \setminus \{C\}) \cup \{C_1, C_2\} \quad \text{if} \quad G \neq C.$$

We say that (V', E', F') is obtained from (V, E, F) by *face bisection*; we also say that (V', E', F') is obtained from (V, E, F) via *bisecting* C by L.

EXERCISE 18.13. Assume the notation in the previous definition. Assume that (V, E, F) is either a combinatorial polygon or a combinatorial polyhedron. Prove that (V', E', F') is a combinatorial polyhedron.

Definition 18.14. A combinatorial polyhedron (V, E, F) is called *combinatorially planar* if and only if there exist

$$(V_1, E_1, F_1), ..., (V_N, E_N, F_N)$$

with $N \geq 2$ such that (V_1, E_1, F_1) is a combinatorial polygon, $(V_N, E_N, F_N) = (V, E, F)$, and for each $k \in \{2, ..., N\}$ we have that (V_k, E_k, F_k) is obtained from $(V_{k-1}, E_{k-1}, F_{k-1})$ by a face bisection.

Exercise 18.15. Assume (V', E', F') is a combinatorial polyhedron obtained either from a combinatorial polygon or from a combinatorial polyhedron (V, E, F) via face bisection of some cycle by a linear graph of order $m + 2 \geq 2$. Prove that

$$|V'| = |V| + m, \quad |E'| = |E| + m + 1, \quad |F'| = |F| + 1.$$

Exercise 18.16. (Combinatorial version of Euler's theorem). Prove that if a combinatorial polyhedron (V, E, F) is combinatorially planar then we have "Euler's formula"

$$|V| - |E| + |F| = 2.$$

Hint: Induction on the number N in Definition 18.14.

Remark 18.17. This Remark uses, again, some previously undefined terms and may be skipped. One can prove that if a combinatorial polyhedron (V, E, F) is combinatorially planar then the graph (V, E) is planar. But it is not clear a priori that the combinatorial polyhedron corresponding to a convex polyhedron is combinatorially planar. (It is a trivial exercise to show that this is the case for the Platonic polyhedra or for other familiar polyhedra such as arbitrary prisms, pyramids, or simple combinations of these. Check the latter statement!) So our purely combinatorial arguments above are not sufficient to prove Euler's theorem about convex polyhedra (cf. Remark 14.28): the latter is a non-trivial topological fact rather than a purely combinatorial fact.

Here is an example showing that the combinatorial polyhedron corresponding to the cube is combinatorially planar. One starts with a combinatorial polygon attached to the cycle $C(1234)$. One bisects it by the linear graph $L(1562)$. Then one bisects the cycle $C(156234)$ by the linear graph $L(683)$. Then one bisects the cycle $C(156834)$ by the linear graph $L(874)$. Finally one bisects the cycle $C(156874)$) by the linear graph $L(75)$. The resulting combinatorial polyhedron corresponds to the cube. Similar constructions can be performed for the other 4 Platonic polyhedra.

Definition 18.18. A combinatorial polyhedron is called *Platonic* if and only if:

1) It is combinatorially planar;
2) All vertices have the same degree;
3) All faces have the same degree.

Exercise 18.19. (Combinatorial version of the classification theorem for Platonic polyhedra). There exist exactly 5 equivalence classes of Platonic combinatorial polyhedra.

Hint: Assume all vertices have degree $d \geq 3$ and all faces have degree g. Let n, s, r be the number of vertices, edges, and faces. By Exercises 18.6 and 18.10 we have

$$2s = dn = gr.$$

By Exercise 18.16 we have

$$0 < 2 = n - s + r = \left(\frac{2}{d} - 1 + \frac{2}{g}\right) s.$$

We get

$$\frac{1}{d} + \frac{1}{g} > \frac{1}{2}.$$

Since the left hand side of the above is $\leq \frac{1}{3} + \frac{1}{g}$ it follows that $g \leq 5$. Similarly $d \leq 5$. So There are only 5 possibilities for the pair (d, g),

$$(3,3), (3,4)(4,3), (3,5), (5,3),$$

hence there are only 5 possibilities for s. One can then list all equivalence classes of graphs that may occur and one finds exactly 5 classes of Platonic combinatorial polyhedra. (By the way they correspond to the 5 Platonic polyhedra.)

CHAPTER 19

Sequences

DEFINITION 19.1. A *sequence* in a set A is a map $F : \mathbb{N} \to A$. If we write $F(n) = a_n$ we also say that $a_1, a_2, ...$ is a sequence in A or that (a_n) is a sequence in A.

THEOREM 19.2. *(Recursion theorem) Let A be a set, $a \in A$ an element, and let $F_1, F_2, ...$ be a sequence of maps $A \to A$. Then there is a unique map $G : \mathbb{N} \to A$ such that $G(1) = a$ and*

$$G(n+1) = F_n(G(n))$$

for all $n \in \mathbb{N}$.

Sketch of proof. First one proves by induction on n that for every n there exists a map

$$G_n : \{1, ..., n\} \to A$$

such that for every $k < n$

$$G_n(k+1) = F_k(G_n(k)).$$

Next, for each natural numbers $n \leq N$ one proves by induction on k that for every k if $k \leq n$ then we have

$$G_n(k) = G_N(k).$$

In other words G_n is the restriction of G_N to $\{1, ..., n\}$. Finally, seeing G_n as a subset of $\{1, ..., n\} \times A$, and hence of $\mathbb{N} \times A$, one defines

$$G := \bigcup_{n \in \mathbb{N}} G_n \subset \mathbb{N} \times A$$

and one proves G is a map with the desired properties. $\square$

EXERCISE 19.3. Fill in the details of the above proof.

One can generalize the above in various ways. Here is an immediate generalization:

EXERCISE 19.4. Let A be a set, let $a_1, a_2, ..., a_m \in A$, and let $F_1, F_2, ...$ be a sequence of maps $A^m \to A$. Prove that there is a unique map $G : \mathbb{N} \to A$ such that $G(1) = a_1, ..., G(m) = a_m$, and

$$G(n+m) = F_n(G(n), G(n+1), ..., G(n+m-1))$$

for all $n \in \mathbb{N}$. (The latter equality is called a *recurrence relation* of degree m.)

Hint: use the same type of argument as the one above.

Here are some applications of recursion. First we are ready to give a proof of:

THEOREM 19.5. *(Bernstein's Theorem) If there exist injections $F : A \to B$ and $G : B \to A$ then there exists a bijection $U : A \to B$*

Proof. Let $C = G(B)$. Then clearly G defines a bijection between B and C and so it is enough to show that there is bijection between A and C. Let $H = G \circ F$. Then $H : A \to C$ is an injection. We are reduced to proving the Lemma below. $\square$

LEMMA 19.6. *Let A be a set and $C \subset A$ a subset. Assume there is an injection $H : A \to C$. Then there is a bijection $U : A \to C$.*

Proof. Define by recursion sequences as follows: $A_1 = A$, $A_{n+1} = H(A_n)$ and $C_1 = C$, $C_{n+1} = H(C_n)$. Define $U : A \to C$ by $U(x) = H(x)$ if $x \in A_n \backslash C_n$ for some $n \in \mathbb{N}$ and $U(x) = x$ otherwise. We claim that $U : A \to C$ is a bijection. Indeed note that we have

$$A_\infty \subset ... \subset C_n \subset A_n \subset ... \subset C_2 \subset A_2 \subset C_1 \subset A_1$$

where

$$A_\infty := \bigcap_n A_n = \bigcap_n C_n.$$

Also note that the sets

$$A_1 \backslash C_1, C_1 \backslash A_2, A_2 \backslash C_2, C_2 \backslash A_3, ..., A_\infty$$

define a partition of A and the sets

$$C_1 \backslash A_2, A_2 \backslash C_2, C_2 \backslash A_3, ..., A_\infty$$

define a partition of C. We are done by noting that U induces bijections

$$A_1 \backslash C_1 \to A_2 \backslash C_2,\ A_2 \backslash C_2 \to A_3 \backslash C_3, ...$$

and is the identity on

$$C_1 \backslash A_2,\ C_2 \backslash A_3, ..., A_\infty$$

hence U is bijection. $\square$

Here is another application of recursion:

PROPOSITION 19.7. *Let $(A, \leq)$ be an ordered set that has no maximal element. Then there is a sequence $F : \mathbb{N} \to A$ such that for all $n \in \mathbb{N}$ we have $F(n) < F(n+1)$.*

Proof. Let $B = \{(a, b) \in A \times A \mid a < b\}$. By hypothesis the first projection $F : B \to A$, $(a, b) \mapsto a$ is surjective. By the axiom of choice there exists $G : A \to B$ such that $F \circ G = I_A$; cf. Exercise 10.23. Then $G(a) > a$ for all a. By the recursion theorem there exists $F : \mathbb{N} \to A$ such that $F(n+1) = G(F(n))$ for all n and we are done. $\square$

EXERCISE 19.8. Prove that if A is an infinite set then there exists an injective map $F : \mathbb{N} \to A$.

Hint. First show there exists a map $H : \mathcal{P}(A) \backslash \{A\} \to A$ with the property that $H(B) \notin B$ for all B; to do this use the axiom of choice applied to the disjoint union

$$\coprod_{B \in \mathcal{P}(A) \backslash \{A\}} (B \backslash A).$$

Cf. Exercise 10.23. Next, let $G : \mathcal{P}(A) \to \mathcal{P}(A)$ be the map defined by $G(A) = A$ and

$$G(B) = B \cup \{H(B)\}, \quad B \in \mathcal{P}(A)\backslash\{A\}.$$

Let $a \in A$. Next, by the recursion theorem there exists a unique map $K : \mathbb{N} \to \mathcal{P}(A)$ such that $K(1) = \{a\}$ and for all $n \in \mathbb{N}$ we have $K(n+1) = G(K(n))$. One proves by induction that for all n we have that $K(n)$ is a finite set (in particular it is not equal to A). Finally define $F : \mathbb{N} \to A$ by the formula $F(n) = G(K(n))$ and one proves that F is injective.

Exercise 19.9. Prove that if A is an infinite set then there exists an injective map $A \to A$ which is not surjective.

Hint. By Exercise 19.8 there exists an injective map $F : \mathbb{N} \to A$. Let $a_n = F(n)$. Then define the map $K : A \to A$ by letting $K(x) = x$ for $x \notin F(\mathbb{N})$ and $K(a_n) = a_{n+1}$ for all $n \in \mathbb{N}$. One checks that K is injective but not surjective.

Definition 19.10. A set A is *countable* if and only if there exists a bijection $F : \mathbb{N} \to A$. A set is *at most countable* if and only if it is either finite or countable. A set is *uncountable* if and only if it is neither finite nor countable.

Example 19.11. The set of all squares $S = \{n^2 \mid n \in \mathbb{N}\}$ is countable; indeed $F : \mathbb{Z} \to S$, $F(n) = n^2$ is a bijection.

Exercise 19.12. Any infinite subset of a countable set is countable.

Hint: It is enough to show that every infinite subset $A \subset \mathbb{N}$ is countable. Let $F \subset \mathbb{N} \times \mathbb{N}$ be the set

$$F = \{(x, y) \in \mathbb{N} \times \mathbb{N} \mid y = \min\{x \in A \mid z > x\})$$

which is of course a map. By the recursion theorem there exists $G : \mathbb{N} \to \mathbb{N}$ such that $G(n+1) = F(G(n))$. One checks that G is injective and its image is A.

Exercise 19.13. Let A and B be infinite sets and $F : A \to B$ a map.
1) Prove that if F is injective and B is countable then A is countable.
2) Prove that if F is surjective and A is countable then B is countable.

Hint. For 1) use the previous exercise. For 2) use the Axiom of choice to get a map $G : B \to A$ such that $F \circ G = I_B$, cf. Exercise 10.23; then G is injective and use 1).

Exercise 19.14.
1) Prove that $\mathbb{N} \times \mathbb{N}$ is countable.
2) Prove that $\mathbb{N}^k$ is countable for all $k \in \mathbb{N}$.

Hint: For 1) one can find an injection $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$; e.g., $(n, m) \mapsto 2^n 3^m$. For 2) use induction on k.

Exercise 19.15. Prove that if A and B are countable then $A \cup B$ and $A \times B$ are countable.

Exercise 19.16.
1) Prove that $\mathbb{Z}$ is countable.
2) Prove that $\mathbb{Q}$ is countable.

Hint: For 1) write $\mathbb{Z} = \mathbb{N} \cup ((-\mathbb{N}) \cup \{0\})$. For 2) consider the surjection $\mathbb{Z} \times \mathbb{N} \to \mathbb{Q}$ given by

$$(a, b) \mapsto \frac{a}{b}.$$

EXERCISE 19.17. Prove that $Fun(\mathbb{N}, \{0,1\})$ is uncountable.

Solution 1. Hint: Since $\mathcal{P}(\mathbb{N})$ is in bijection with $Fun(\mathbb{N}, \{0,1\})$ (cf. Exercise 10.44) this is a consequence of the more general Theorem 10.31 saying that there is no bijection between a set A and its power set $\mathcal{P}(A)$.

Solution 2. Hint: Assume there is a bijection $F : \mathbb{N} \to Fun(\mathbb{N}, \{0,1\})$ and seek a contradiction. Denote $F(n)$ by $F_n : \mathbb{N} \to \{0,1\}$. Construct a map $G : \mathbb{N} \to \{0,1\}$ by the formula

$$G(n) = \neg(F_n(n))$$

where $\neg : \{0,1\} \to \{0,1\}$, $\neg 0 = 1$, $\neg 1 = 0$. (The definition of G does not need the recursion theorem; one can define G as a "graph" directly (check!).) Since F is surjective there exists m such that $G = F_m$. In particular:

$$G(m) = F_m(m) = \neg G(m),$$

a contradiction. This is called *Cantor's diagonal argument*.

EXERCISE 19.18. Prove that if (A_n) is a sequence of sets such that each A_n is at most countable then the union

$$\bigcup_{n \in \mathbb{N}} A_n$$

is at most countable. Deduce that the set of words A^* with letters in a finite non-empty set A is countable.

EXERCISE 19.19. Let us call $\mathcal{F}$ the set of all functions $f : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$ where $\infty \notin \mathbb{N}$. If $f(n) = \infty$ we say that f with input n *does not terminate* (or *runs forever without giving an output*). If $f(n) = m \in \mathbb{N}$ we say that f with input n *terminates and gives output* m.

Let (f_n) be a sequence of elements of $\mathcal{F}$. Its oracle is defined as the function $g \in \mathcal{F}$ satisfying:

1) $g(n) \in \{1, 2\}$ for all n; in particular g with any input n always terminates.
2) $g(n) = 1$ if f_n with input n terminates (i.e. $f_n(n) \in \mathbb{N}$),
3) $g(n) = 2$ if f_n with input n does not terminate (i.e. $f_n(n) = \infty$).

Given (f_n) one can define a function $f \in \mathcal{F}$ (which we refer to as the Turing function attached to (f_n)) as follows. Let g be the oracle of (f_n). Then we let $f(n) = f_n(n) + 1$ if $g(n) = 1$ and we let $f(n) = 1$ if $g_n(n) = 2$.

Prove that if (f_n) is a sequence of functions in $\mathcal{F}$ then the Turing function f attached to (f_n) is not a member of (f_n).

Hint: if f_n with input n does not terminate then f cannot be equal to f_n because f terminates with input n; on the other hand if f_n with input n terminates then again $f \neq f_n$ because they have different outputs.

REMARK 19.20. In this Remark only we assume some intuitive knowledge about "programming." Our discussion here will be informal and imprecise but worth including since it gives a hint as to the applications of Exercise 19.19. Indeed this exercise contains the idea behind Turing's proof that there is no "program" that decides if an arbitrary given "program" with an arbitrary given input terminates. The problem of existence of such a "program" is called the *halting problem* and Turing's result is that the halting problem has a negative answer. To explain this "recall" that Turing defined the notion of "program" as being a special type of function in $\mathcal{F}$. We do not explain his definition here and this will make our discussion not rigorous. One proves there are countably many "programs." This is a consequence of the fact that the set of finite sequences of words with letters in a finite set is countable. Let (f_n) be the sequence of all "programs." (One assumes that given a "program" there is a mechanical procedure to attach to it a label n and hence identifying it with the function f_n.) One proves that if the oracle g of this sequence is a "program" then the Turing function attached to (f_n) is a "program," a contradiction. One concludes that g cannot be a "program" hence the halting problem has a negative answer.

Another way to explain the above is to introduce the following definitions. A subset $S \subset \mathbb{N}$ is called *listable* if there exists n such that $f_n(\mathbb{N}) = S$. One says that a subset $S \subset \mathbb{N}$ is *decidable* if there exists n such that for all $x \in \mathbb{N}$ we have that $f_n(x) = 1$ for $x \in S$ and $f_n(x) = 0$ for $x \notin S$.

One can prove that a set S is decidable if and only if both S and $\mathbb{N} \backslash S$ are listable. For the "if" part let f_p and f_q have images S and $\mathbb{N} \backslash S$, respectively. Consider the sequence $f_p(1), f_q(1), f_p(2), f_q(2), \ldots.$ If x shows up in an even place set $f(x) = 1$; if x shows up in an odd place set $f(x) = 0$. One proves that f is a program and we are done. To check the "only if" part let f_n "decide" S. Consider the sequence $f_n(1), f_n(2), \ldots.$ If $f_n(x) = 1$ print x; if $f_n(x) = 0$ do not print x. The printed list is S which ends the proof of the "only if" part.

Let $\beta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be any natural (i.e., explicitly given by a formula) bijection and let $F_n : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \cup \{\infty\}$ be defined by $F_n(n, m) = f_n(\beta(n, m))$. Then Exercise 19.19 easily implies that the "Turing set"

$$S = \beta(\{(n, m) \in \mathbb{N} \times \mathbb{N} \mid f_n(m) \neq \infty\})$$

is listable but not decidable. This is another way of formulating Turing's result that the halting problem has a negative answer.

The rest of this Remark involves concepts that we have not yet defined and therefore may be skipped and revisited after we discuss polynomials. For what follows we assume we are familiar with polynomials. A set $S \subset \mathbb{N}$ is said to be *Diophantine* if there exists a polynomial $F = F(x, y)$ in $m+1$ variables (where x is an m-tuple of variables and y is one variable) with coefficients in $\mathbb{Z}$ such that $b \in S$ if and only if there exists $a \in \mathbb{Z}^m$ with $F(a, b) = 0$. Every Diophantine set is easily seen to be listable. A celebrated problem of Hilbert asked if every Diophantine set is decidable. Matjasevich proved that every listable set is Diophantine. (This is one of the most remarkable results in the theory of computation.) Since there exist listable sets that are not decidable (e.g., the "Turing set") it follows that there exist Diophantine sets that are not decidable, hence Hilbert's problem has a negative answer.

REMARK 19.21. Consider the following sentence called the *continuum hypothesis*:

For every set A if there exists an injection $A \to \mathcal{P}(\mathbb{N})$ then either there exists an injection $A \to \mathbb{N}$ or there exists a bijection $A \to \mathcal{P}(\mathbb{N})$.

One can ask if the above is a theorem. Answering this question (raised by Cantor) leads to important investigations in Set Theory. The answer (given by two theorems of Gödel and Cohen in the framework of Mathematical Logic rather than

Logic) turned out to be rather surprising; we are not going to discuss these issues in this course.

CHAPTER 20

Arithmetic

Our main aim here is to introduce some of the basic "arithmetic" of $\mathbb{Z}$. In its turn arithmetic can be used to introduce the finite rings $\mathbb{Z}/m\mathbb{Z}$ of residue classes modulo m and, in particular, the finite fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, where p is a prime. In addition one can introduce the ring of p-adic integers, $\mathbb{Z}_p$, which plays a key rile in modern number theory. The arithmetic of $\mathbb{Z}$ to be discussed below already appears in Euclid. Congruences and residue classes were introduced by Gauss at the turn of the 19th century. The p-adic integers were introduced by Hensel at the turn of the 20th century.

Definition 20.1. For integers a and b we say a *divides* b if there exists an integer n such that $b = an$. We write $a|b$. We also say a is a *divisor* of b; or that b is a *multiple* of a. If a does not divide b we write $a \not| b$.

Example 20.2. $4|20$; $-4|20$; $6 \not| 20$.

Exercise 20.3. Prove that
1) if $a|b$ and $b|c$ then $a|c$;
2) if $a|b$ and $a|c$ then $a|b+c$;
3) $a|b$ defines an order relation on $\mathbb{N}$ but not on $\mathbb{Z}$.

Theorem 20.4. *(Euclid division) For every $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ there exist unique $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r < b$.*

The number r above is called the *remainder* when a is divided by b and will be denoted by $r_b(a)$.

Proof. We prove the existence of q, r. The uniqueness is left to the reader. We may assume $a \in \mathbb{N}$. We proceed by contradiction. So assume there exists b and $a \in \mathbb{N}$ such that for all $q, r \in \mathbb{Z}$ with $0 \leq r < b$ we have $a \neq qb + r$. Fix such a b. We may assume a is minimum with the above property. If $a < b$ we can write $a = 0 \times b + a$, a contradiction. If $a = b$ we can write $a = 1 \times a + 0$, a contradiction. If $a > b$ set $a' = a - b$. Since $a' < a$, there exist $q', r \in \mathbb{Z}$ such that $0 \leq r < b$ and $a' = q'b + r$. But then $a = qb + r$, where $q = q' + 1$, a contradiction. □

Definition 20.5. For $a \in \mathbb{Z}$ denote $a\mathbb{Z}$ the set $\{na \mid n \in \mathbb{Z}\}$ of integers divisible by a. For $a, b \in \mathbb{Z}$ denote by $a\mathbb{Z} + b\mathbb{Z}$ the set $\{ma + nb \mid m, n \in \mathbb{Z}\}$ of all numbers expressible as a multiple of a plus a multiple of b.

Proposition 20.6. *For every integers a, b there exists an integer c such that*

$$a\mathbb{Z} + b\mathbb{Z} = c\mathbb{Z}.$$

Proof. If $a = b = 0$ we can take $c = 0$. Assume a, b are not both 0. Then the set $S = (a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{N}$ is non-empty. Let c be the minimum of S. Clearly $c\mathbb{Z} \subset a\mathbb{Z} + b\mathbb{Z}$. Let us prove the opposite inclusion. Let $u = ma + nb$ and let us prove that $u \in c\mathbb{Z}$.

By Euclidean division $u = cq + r$ with $0 \leq r < c$. We want to show $r = 0$. Assume $r \neq 0$ and seek a contradiction. Write $c = m'a + n'b$. Then $r \in \mathbb{N}$ and also

$$r = u - cq = (ma + nb) - (m'a + n'b)q = (m - m'q)a + (n - n'q)b \in a\mathbb{Z} + b\mathbb{Z}.$$

Hence $r \in S$. But $r < c$. So c is not the minimum of S, a contradiction. $\square$

PROPOSITION 20.7. *If a and b are integers and have no common divisor > 1 then there exist integers m and n such that $ma + nb = 1$.*

Proof. By the above Proposition $a\mathbb{Z} + b\mathbb{Z} = c\mathbb{Z}$ for some $c \geq 1$. In particular $c|a$ and $c|b$. The hypothesis implies $c = 1$ hence $1 \in a\mathbb{Z} + b\mathbb{Z} = c\mathbb{Z}$. $\square$

One of the main definitions of number theory is

DEFINITION 20.8. An integer p is *prime* if and only if $p > 1$ and if its only positive divisors are 1 and p.

PROPOSITION 20.9. *If p is a prime and a is an integer such that $p \not| a$ then there exist integers m, n such that $ma + np = 1$.*

Proof. a and p have no common divisor > 1 and we conclude by Proposition 20.7. $\square$

PROPOSITION 20.10. *(Euclid Lemma) If p is a prime and $p|ab$ for integers a and b then either $p|a$ or $p|b$.*

Proof. Assume $p|ab$, $p \not| a$, $p \not| b$, and seek a contradiction. By Proposition 20.9 $ma + np = 1$ for some integers m, n and $m'b + n'p = 1$ for some integers m', n'. We get

$$1 = (ma + np)(m'b + n'p) = mm'ab + p(nm' + n'm + nn').$$

Since $p|ab$ we get $p|1$, a contradiction. $\square$

THEOREM 20.11. *(Fundamental Theorem of Arithmetic) Any integer $n > 1$ can be written uniquely as a product of primes, i.e., there exist primes $p_1, p_2, ..., p_s$, where $s \geq 1$, such that*

$$n = p_1 p_2 ... p_s.$$

Moreover any such representation is unique in the following sense: if

$$p_1 p_2 ... p_s = q_1 q_2 ... q_t$$

with p_i and q_j prime and $p_1 \leq p_2 \leq ...$, $q_1 \leq q_2 \leq ...$ then $s = t$ and $p_1 = q_1$, $p_2 = q_2$, ...

Proof. To prove the existence part let S be the set of all integers > 1 which are not products of primes. We want to show $S = \emptyset$. Assume the contrary and seek a contradiction. Let n be the minimum of S. Then n is not prime. So $n = ab$ with $a, b > 1$ integers. So $a < n$ and $b < n$. So $a \notin S$ and $b \notin S$. So a and b are products of primes. So n is a product of primes, a contradiction. For the uniqueness see the Exercise below. $\square$

EXERCISE 20.12. Prove the uniqueness part in the above Theorem.

Hint: Use induction on s. By Euclid's Lemma 20.10 since p_1 divides the product of the qs we get $p_1 = q_i$ for some i; and since q_1 divides the product of the ps we get $q_1 = p_j$ for some j. Then $p_1 = q_i \geq q_1 = p_j \geq p_1$, so $p_1 = q_1$ and one can divide by p_1.

DEFINITION 20.13. Fix an integer $m \neq 0$. Define a relation $\equiv_m$ on $\mathbb{Z}$ by $a \equiv_m b$ if and only if $m|a-b$. Say a is *congruent* to b mod m (or modulo m). Instead of $a \equiv_m b$ one usually writes (following Gauss):

$$a \equiv b \pmod{m}.$$

EXAMPLE 20.14. $3 \equiv 17 \pmod{7}$.

EXERCISE 20.15. Prove that $\equiv_m$ is an equivalence relation. Prove that the equivalence class $\overline{a}$ of a consists of all the numbers of the form $mb+a$ where $m \in \mathbb{Z}$.

EXAMPLE 20.16. If $m = 7$ then $\overline{3} = \overline{10} = \{..., -4, 3, 10, 17, ...\}$.

DEFINITION 20.17. For the equivalence relation $\equiv_m$ on $\mathbb{Z}$ the set of equivalence classes $\mathbb{Z}/\equiv_m$ is denoted by $\mathbb{Z}/m\mathbb{Z}$. The elements of $\mathbb{Z}/m\mathbb{Z}$ are called *residue classes* mod m.

EXERCISE 20.18. Prove that

$$\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, ..., \overline{m-1}\}.$$

So the residue classes mod m are: $\overline{0}, \overline{1}, ..., \overline{m-1}$.

Hint: Use Euclid division.

EXERCISE 20.19. Prove that if $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$ then $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$.

DEFINITION 20.20. Define operations $+, \times, -$ on $\mathbb{Z}/m\mathbb{Z}$ by

$$\overline{a} + \overline{b} = \overline{a+b}$$

$$\overline{a} \times \overline{b} = \overline{ab}$$

$$-\overline{a} = \overline{-1}.$$

EXERCISE 20.21. Check that the above definitions are correct, in other words that if $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$ then

$$\overline{a+b} = \overline{a'+b'}$$

$$\overline{ab} = \overline{a'b'}$$

$$\overline{-a} = \overline{-a'}.$$

Furthermore check that $(\mathbb{Z}/m\mathbb{Z}, +, \times, -, \overline{0}, \overline{1})$ is a ring.

DEFINITION 20.22. If p is a prime we write $\mathbb{F}_p$ in place of $\mathbb{Z}/p\mathbb{Z}$.

EXERCISE 20.23. Prove that $\mathbb{F}_p$ is a field.

Hint: Use Proposition 20.9.

One can generalize the above constructions as follows.

DEFINITION 20.24. Let R be a ring (always commutative and unital). A subset $I \subset R$ is called an *ideal* if for all $a, b \in I$ and all $r \in R$ we have
1) $a + b \in I$, $-a \in I$;
2) $ra \in I$.
An ideal P is called a *prime* ideal if for all $a, b \in P$ if $ab \in P$ then $a \in P$ or $b \in P$. An ideal M is called *maximal* if $M \neq R$ and for every ideal I containing M either $I = M$ or $I = R$.

EXERCISE 20.25. Let I be an ideal in a ring R. Consider the equivalent relation $\sim$ on R defined by $a \sim b$ if and only if $a - b \in I$. Write R/I for the quotient $R/\sim$ of R by this equivalence relation. If $a \sim b$ we write

$$a \equiv b \mod I.$$

Define an addition and multiplication on R/I by

$$\widehat{a} + \widehat{b} := \widehat{a+b}, \quad \widehat{a} \cdot \widehat{b} := \widehat{ab}.$$

1) Prove that R/I with these operations and $\widehat{0}, \widehat{1}$ is a ring.
2) Prove that R/I is an integral domain if and only if I is a prime ideal.
3) Prove that R/I is a field if and only if I is maximal ideal.
4) Prove that every ideal that is not equal to R is contained in at least one maximal ideal.

Hint for 4): Use Zorn's Lemma.

EXERCISE 20.26. Prove that every non-zero ideal in $\mathbb{Z}$ is of the form $a\mathbb{Z}$ for some $a \in \mathbb{N}$. Prove that the latter is maximal if and only if it is prime if and only if a is a prime.

DEFINITION 20.27. Let R be a ring. A subset $B \subset R$ is called a *subring* if for all $a, b \in B$ we have $a + b \in B$, $-a \in B$, $1 \in B$, $ab \in B$. So such a B is itself a ring with the induced operations.

DEFINITION 20.28. Let K be a field. A subset $F \subset K$ is called a *subfield* if for all $a, b \in F$ and $c \in F \setminus \{0\}$ we have $a + b \in F$, $-a \in F$, $1 \in F$, $ab \in F$, $c^{-1} \in F$. So such an F is itself a field with the induced operations.

EXERCISE 20.29. Prove that a subset of a ring R which is both an ideal and a subring is the whole of R.

EXERCISE 20.30.
1) Prove that if R is a ring then there exists a unique ring homomorphism $f_R : \mathbb{Z} \to R$.
2) Prove that $\{a \in R \mid f_R(a) = 0\}$ is an ideal hence it is of the form $c\mathbb{Z}$, $c \in \mathbb{N} \cup \{0\}$ unique; c is called the *characteristic* of R and is denoted by $char(R)$.
3) Prove that if R is an integral domain then $char(R)$ is either 0 or a prime.
4) Prove that if R is a field of characteristic 0 then R contains a subfield isomorphic to $\mathbb{Q}$; if R is a field of characteristic p then R contains a subfield isomorphic to $\mathbb{F}_p$. One usually identifies these subfields with $\mathbb{Q}$ or $\mathbb{F}_p$, respectively.

In what follows we introduce the *ring of p-adic numbers* which was invented by Hensel and plays a key role in number theory.

DEFINITION 20.31. For a prime $p \in \mathbb{Z}$ consider the set Z_p of all sequences $(a_n)_{n \geq 1}$ of integers $a_n \in \mathbb{Z}$ with the property that for all $n \in \mathbb{N}$ we have

$$a_{n+1} \equiv a_n \pmod{p^n}.$$

Define an equivalence relation $\sim$ on Z_p as follows:

$$(a_n) \sim (b_n) \leftrightarrow \forall n (a_n \equiv b_n \pmod{p^n}).$$

Define

$$\mathbb{Z}_p := Z_p/\sim$$

and let $[a_n] \in \mathbb{Z}_p$ denote the equivalence class of a sequence $(a_n) \in Z_p$. Define

$$[a_n] + [b_n] = [a_n + b_n], \quad [a_n] \cdot [b_n] = [a_n b_n], \quad -[a_n] = [-a_n],$$
$$0 = [0_n], 0_n = 0, \quad 1 = [1_n], 1_n = 1.$$

One can check that $+, -, \cdot$ above are well defined and that $(\mathbb{Z}_p, +, \cdot, -, 0, 1)$ is a ring (see the Exercise below). $\mathbb{Z}_p$ is called the *ring of p-adic integers*. Finally one defines

$$\mathbb{Q}_p = (\mathbb{Z}_p \times (\mathbb{Z}_p \setminus \{0\}))/ \simeq$$

where $\simeq$ is the equivalent relation

$$(\alpha_1, \beta_1) \simeq (\alpha_2, \beta_2) \quad \leftrightarrow \quad \alpha_1\beta_2 = \alpha_2\beta_1.$$

One can check that with the natural $+, \times$ (defined as in the case of $\mathbb{Q}$), $\mathbb{Q}_p$ is a field; it is called the field of p-adic numbers.

Note that many textbooks use the symbol $\mathbb{Z}_p$ to denote what we denoted by $\mathbb{F}_p$; this is mostly the case for textbooks that are not number theoretic. On the other hand number theoretic textbooks tend to use the notation we adopted here.

Exercise 20.32. Prove that $+, -, \cdot$ above are well defined and prove that $\mathbb{Z}_p$ is a ring and $\mathbb{Q}_p$ is a field (both of characteristic 0).

Definition 20.33. Say that p^e *divides* $\alpha = [a_n]$ if and only if there exists $\beta = [b_n]$ such that $[a_n] = [p^e b_n]$; write $p^e | \alpha$. For every $0 \neq \alpha = [a_n] \in \mathbb{Z}_p$ let $v = v(\alpha)$ be the unique integer such that $p^n | a_n$ for $n \leq v$ and $p^{v+1} \not| a_{v+1}$. Then define the *norm* of α by the formula $|\alpha| = p^{-v(\alpha)}$. We also set $|0| = 0$. Finally we define a *open disk* in $\mathbb{Z}_p$ (centered at $\alpha_0 \in \mathbb{Z}_p$, of radius $r \in \mathbb{Q}$ to be a set of the form

$$D_r(\alpha_0) := \{\alpha \in \mathbb{Z}_p \mid |\alpha - \alpha_0| < r\}.$$

We define a *closed disk* in $\mathbb{Z}_p$ (centered at $\alpha_0 \in \mathbb{Z}_p$, of radius $r \in \mathbb{Q}$ to be a set of the form

$$\overline{D_r}(\alpha_0) := \{\alpha \in \mathbb{Z}_p \mid |\alpha - \alpha_0| \leq r\}.$$

Exercise 20.34. Prove that if $\alpha = [a_n]$ and $\beta = [b_n]$ then

$$|\alpha + \beta| \leq \max\{|\alpha|, |\beta|\},$$
$$|\alpha \cdot \beta| = |\alpha| \cdot |\beta|.$$

Conclude that $\mathbb{Z}_p$ is an integral domain.

Exercise 20.35.

1) Prove that every open disk is also a closed disk (with a possibly different radius).

2) Prove that every closed disk is also an open disk (with a possibly different radius).

CHAPTER 21

Reals

Real numbers have been implicitly around throughout the history of Mathematics as an expression of the idea of continuity of magnitudes. A definition of the reals based on geometry can be found in Euclid and is due to Eudoxus. The first construction of the reals from the "discrete" (i.e., from the integers) is due to Dedekind.

DEFINITION 21.1. (Dedekind) A *real number* is a subset $u \subset \mathbb{Q}$ of the set $\mathbb{Q}$ of rational numbers with the following properties:
1) $u \neq \emptyset$ and $u \neq \mathbb{Q}$;
2) u has no minimum;
2) if $x \in u$, $y \in \mathbb{Q}$, and $x \leq y$ then $y \in u$.
Denote by $\mathbb{R}$ the set of real numbers.

EXERCISE 21.2. For every rational number $x \in \mathbb{Q}$ we can define the set

$$u_x = \{y \in \mathbb{Q} \mid x < y\}.$$

1) Prove that u_x is a real number.
2) Prove that $u_x = u_{x'}$ for $x, x' \in \mathbb{Q}$ implies $x = x'$.

From now on we will identify every rational number x with the real number u_x. So we may view $\mathbb{Q} \subset \mathbb{R}$.

DEFINITION 21.3. A real number $u \in \mathbb{R}$ is called *irrational* if $u \notin \mathbb{Q}$.

DEFINITION 21.4. If u and v are real numbers we write $u \leq v$ if and only if $v \subset u$. For $u, v \geq 0$ define

$$\begin{aligned} u+v &= \{x+y \mid x \in u, y \in v\} \\ u \times v = uv &= \{xy \mid x \in u, y \in v\}. \end{aligned}$$

EXERCISE 21.5.
1) Prove that $\leq$ is a total order on $\mathbb{R}$.
2) Prove that $u+v$ and $u \times v$ are real numbers.
3) Naturally extend the definition of addition $+$ and multiplication $\times$ of real numbers to the case when the numbers are not necessarily ≥ 0.
4) Prove $+$ and $\times$ on $\mathbb{R}$ coincide with addition and multiplication on $\mathbb{Q}$.
5) Prove that $(\mathbb{R}, +, \times, -, 0, 1)$ is a field.
6) Naturally extend the order $\leq$ on $\mathbb{Q}$ to an order on $\mathbb{R}$ and prove that $\mathbb{R}$ with $\leq$ is an ordered ring.

EXERCISE 21.6. Define the sum and the product of a family of real (or complex) numbers indexed by a finite set.

Hint: Use the already defined concept for integers (and hence for the rationals).

In the next exercises we look at square roots.

EXERCISE 21.7.

1) Let $r > 0$ be a real number, α a rational number with $\frac{1}{2} < \alpha < 1$ and $M \in \mathbb{N}$ with $M > r$. Prove that for every $n \in \mathbb{N}$ there exist positive rational numbers t_n and s_n with the following properties:

i) $0 < t_n < s_n < M$,

ii) $t_n^2 < r < s_n^2$,

iii) $s_n - t_n \leq M \times \alpha^n$.

2) Prove that for every positive rational numbers a and b with $a < b$ there exists a positive rational number c such that $a < c^2 < b$.

Hint for 1): For $n = 1$ take $t_1 = 0$ and $r < s_1 < M$. Assuming the statement is true for n one sets

$$t_{n+1} = t_n \text{ and } s_{n+1} = (t_n + s_n)/2 \text{ if } t_n^2 < r < ((t_n + s_n)/2)^2,$$

one sets

$$t_{n+1} = (t_n + s_n)/2 \text{ and } s_{n+1} = s_n \text{ if } ((t_n + s_n)/2)^2 < r < s_n^2.$$

and one sets

$$t_{n+1} = t_n \text{ and } s_{n+1} = t_n + \alpha(s_n - t_n) \text{ if } r = ((t_n + s_n)/2)^2.$$

The above method of proof based on successive divisions of intervals is used in calculus quite often.

Hint for 2): use 1) and the fact that the set $\{a^n \mid n \in \mathbb{N}\}$ is not bounded if $a \in \mathbb{Q}$, $a > 1$.

EXERCISE 21.8. Prove that for $r \in \mathbb{R}$, $r \geq 0$, the set

$$\sqrt{r} := \{x \in \mathbb{Q} \mid x \geq 0, x^2 > r\}$$

is a real number and we have $(\sqrt{r})^2 = r$.

Hint: The harder part is to show that if $z \in \mathbb{Q}$ satisfies $z > r$ then there exist $x, y \in \mathbb{Q}$ such that $x \geq 0$, $y \geq 0$, $z = xy$, $x^2 > r$, $y^2 > r$. By Exercise 21.7 there exists a positive rational number ρ with $z > \rho^2 > r$. Then we can write $z = xy$ with $x = \rho$ and $y = z/\rho > \rho$.

EXERCISE 21.9. Prove that $\sqrt{2}$ is irrational i.e., $\sqrt{2} \notin \mathbb{Q}$.

Hint: this is, of course, the same as Exercise 15.15

EXERCISE 21.10. Prove that the set

$$\{r \in \mathbb{Q} \mid r > 0, r^2 < 2\}$$

has no supremum in $\mathbb{Q}$.

EXERCISE 21.11. Prove that every subset of $\mathbb{R}$ that has an upper bound in $\mathbb{R}$ has a supremum in $\mathbb{R}$. Similarly every subset of $\mathbb{R}$ that has a lower bound in $\mathbb{R}$ has an infimum in $\mathbb{R}$.

Hint: If $A \subset \mathbb{R}$ has an upper bound in $\mathbb{R}$ then define

$$w = \bigcap_{u \in A} u \subset \mathbb{R}.$$

If w has no minimum element set $v = w$. If w has a minimum element m set $v = w \backslash \{m\}$. Prove that in both cases v is a real number and it is equal to the supremum of A.

REMARK 21.12. Later we will prove that $\mathbb{R}$ is uncountable.

REMARK 21.13. Recall from Exercise 12.38 that for every $a \in \mathbb{R}$ we let $|a|$ be a or $-a$ according as $a \geq 0$ or $a \leq 0$, respectively. We call $|a|$ the *absolute value* of a. Then, according to loc.cit.,

$$|a + b| \leq |a| + |b|, \quad |ab| = |a| \cdot |b|$$

for all $a, b \in \mathbb{R}$.

DEFINITION 21.14. For $a < b$ in $\mathbb{R}$ we define the *open interval*

$$(a, b) = \{c \in \mathbb{R} \mid a < c < b\} \subset \mathbb{R}.$$

(Not to be confused with the pair $(a, b) \in \mathbb{R} \times \mathbb{R}$ which is denoted by the same symbol.) Define the *closed interval*

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}.$$

CHAPTER 22

Imaginaries

Complex numbers (also called *imaginary numbers*) appeared in work of Cardano, Bombelli, d'Alembert, Gauss, and others, in relation to solving polynomial equations. The modern definition below is due to Hamilton.

DEFINITION 22.1. (Hamilton) A *complex number* is a pair $z = (a, b)$ where $a, b \in \mathbb{R}$. We write $a = Re\ z$, $b = Im\ z$. We denote by $\mathbb{C}$ the set of complex numbers. Hence $\mathbb{C} = \mathbb{R} \times \mathbb{R}$. Define the sum and the product of two complex numbers by

$$\begin{array}{rcl}(a,b)+(c,d) & = & (a+c, b+d) \\ (a,b)\times(c,d) & = & (ac-bd, ad+bc).\end{array}$$

REMARK 22.2. Identify every real number $a \in \mathbb{R}$ with the complex number $(a, 0) \in \mathbb{C}$; hence write $a = (a, 0)$. In particular $0 = (0, 0)$ and $1 = (1, 0)$.

EXERCISE 22.3. Prove that $\mathbb{C}$ equipped with $0, 1$ above and with the operations $+, \times$ above is a field. Also note that the operations $+$ and $\times$ on $\mathbb{C}$ restricted to $\mathbb{R}$ are the "old" operations $+$ and $\times$ on $\mathbb{R}$.

DEFINITION 22.4. We set $i = (0, 1)$.

REMARK 22.5. $i^2 = -1$. Indeed

$$i^2 = (0,1)\times(0,1) = (0\times 0 - 1\times 1, 0\times 1 + 1\times 0) = (-1, 0) = -1.$$

REMARK 22.6. For every complex number $(a, b) = a + bi$. Indeed

$$(a,b) = (a,0) + (0,b) = (a,0) + (b,0)(0,1) = a + bi.$$

DEFINITION 22.7. For every complex number $z = a + bi$ we define its *absolute value*

$$|z| = \sqrt{a^2 + b^2}.$$

EXERCISE 22.8. Prove that:

$$|z+w| \leq |z| + |w|, \quad |zw| = |z|\cdot|w|$$

for all $z, w \in \mathbb{C}$.

DEFINITION 22.9. For every complex number $z = a + bi$ we define its *conjugate*

$$\overline{z} = a - bi.$$

(The upper bar is not to be confused with the notation used in the chapter on residues.)

EXERCISE 22.10. Prove that for every $z, w \in \mathbb{C}$ we have
1) $\overline{z+w} = \overline{z} + \overline{w}$;
2) $\overline{z\cdot w} = \overline{z}\cdot\overline{w}$;

3) $\overline{z^{-1}} = \overline{z}^{-1}$ for $z \neq 0$;
4) $z \cdot \overline{z} = |z|^2$.

Definition 22.11. For every complex number $z_0 = a + bi \in \mathbb{C}$ and every real number $r > 0$ we define the *open disk* with *center* z_0 and *radius* r,

$$D_r(z_0) = \{z \in \mathbb{C} \mid |z - z_0| < r\} \subset \mathbb{C}$$

and the *closed disk* with *center* z_0 and *radius* r,

$$\overline{D_r}(z_0) = \{z \in \mathbb{C} \mid |z - z_0| \leq r\} \subset \mathbb{C}.$$

Exercise 22.12. Prove that $\mathbb{C}$ cannot be given the structure of an ordered ring.

The idea of constructing $\mathbb{C}$ from $\mathbb{R}$ can be applied to constructing finite fields. This idea originates with Galois. Here is an example of that.

Exercise 22.13. Let $\mathbb{F}_4$ be the set $\{0,1\} \times \{0,1\}$ and write $0 = (0,0)$, $1 = (1,0)$, $\iota = (0,1)$. Define addition and multiplication on $\mathbb{F}_4$ by

$$\begin{array}{rcl} (a,b) + (c,d) & = & (a+c, b+d) \\ (a,b) \times (c,d) & = & (ac+bd, ad+bc+bd) \end{array}$$

1) Prove that $(a,b) = a + b\iota$.
2) Prove that $\iota^2 = 1 + \iota$.
3) Prove that $\mathbb{F}_4$ is a field containing the field $\mathbb{F}_2 = \{0,1\}$.
4) Prove that if instead of the above multiplication one considers the multiplication given by

$$(a,b) \times (c,d) = (ac - bd, ad + bc)$$

(similar to the one on $\mathbb{C}$) then one obtains a ring which is not a field.

CHAPTER 23

Topology

Topology is concerned with geometric properties that are invariant under *continuous* deformation. From the viewepoint of topology a sphere "is the same" as an ellipsoid or a cube, but not "the same" as a doughnut. An early topological result is the formula of Descartes-Euler relating the number of vertices, edges, and faces of a convex polyhedron; cf. Remark 14.28. (We will not discuss this here as it is surprisingly difficult to present things rigorously.) After Riemann's work on surfaces defined by algebraic functions, topology became a key feature in geometry and calculus and nowadays topological ideas are to be found everywhere in Mathematics, including number theory and functional analysis (where the points of "space" are functions). Here we will restrict ourselves to explaining the basic idea of continuity.

Definition 23.1. A *topology* on a set X is a subset $\mathcal{T} \subset \mathcal{P}(X)$ of the power set of X with the following properties:

1) $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$;
2) If $U, V \in \mathcal{T}$ then $U \cap V \in \mathcal{T}$;
3) If $(U_i)_{i \in I}$ is a family of subsets $U_i \subset X$ and if for all $i \in I$ we have $U_i \in \mathcal{T}$ then $\bigcup_{i \in I} U_i \in \mathcal{T}$.

The elements of X are called *points* of X. A subset $U \subset X$ is called *open* if $U \in \mathcal{T}$. A subset $Z \subset X$ is called *closed* if $X \backslash Z$ is open. A subset $Y \subset X$ is called *dense* if every open set of X has a non-empty intersection with Y. A subset $S \subset X$ is called *discrete* if for every point $x \in X$ there exists an open set U in X such that $U \cap S = \{x\}$.

Example 23.2. $\mathcal{T} = \mathcal{P}(X)$ is a topology on X.

Example 23.3. $\mathcal{T} = \{\emptyset, X\} \subset \mathcal{P}(X)$ is a topology on X.

Definition 23.4. If X is a topological space and $Y \subset X$ is a subset then the set of all subsets of Y of the form $U \cap Y$ with U open in X form a topology on Y called the *induced* topology.

Example 23.5. A subset $U \subset \mathbb{R}$ is called *open* if for every $x \in U$ there exists an open interval containing x and contained in U, $x \in (a, b) \subset U$. Let $\mathcal{T} \subset \mathcal{P}(\mathbb{R})$ be the set of all open sets of $\mathbb{R}$. Then $\mathcal{T}$ is a topology on $\mathbb{R}$; we call this the *Euclidean topology*. The induced topology on a subset A of $\mathbb{R}$ will be called the Euclidean topology on A.

Exercise 23.6. Prove that $\mathcal{T}$ in Example 23.5 is a topology. Prove that every open interval (a, b) is open in $\mathbb{R}$ in the Euclidean topology and every closed interval $[a, b]$ is closed in $\mathbb{R}$ in the Euclidean topology.

Example 23.7. A subset $U \subset \mathbb{C}$ is called *open* if for every $z_0 \in U$ there exists an open disk (Cf. Definition 22.11) centered at z_0 and contained in U. Let $\mathcal{T} \subset \mathcal{P}(\mathbb{V})$ be the set of all open sets of $\mathbb{C}$. Then $\mathcal{T}$ is a topology on $\mathbb{C}$; we call this the *Euclidean topology*. The induced topology on a subset A of $\mathbb{C}$ will be called the Euclidean topology on A.

Exercise 23.8. Prove that the set $\mathcal{T}$ of open subsets of $\mathbb{C}$ is a topology on $\mathbb{C}$. Prove that every open disk is open in the Euclidean topology. Prove that every closed disk is closed in the Euclidean topology.

Example 23.9. A subset $U \subset \mathbb{Z}_p$ is called *open* if for every $\alpha_0 \in U$ there exists an open disk centered at α_0 and contained in U. Let $\mathcal{T} \subset \mathcal{P}(\mathbb{V})$ be the set of all open sets of $\mathbb{Z}_p$. Then $\mathcal{T}$ is a topology on $\mathbb{Z}_p$; we call this the *p-adic topology*.

Exercise 23.10. Prove that the set $\mathcal{T}$ of open subsets of $\mathbb{Z}_p$ is a topology on $\mathbb{Z}_p$. Prove that every open disk in $\mathbb{Z}_p$ is both open and closed in the p-adic topology. (Since closed disks are automatically open disks, of possibly different radii, it follows that every closed disk in $\mathbb{Z}_p$ is both open and closed in the p-adic topology.)

Exercise 23.11. Prove that if $U, V \subset X$ are two arbitrary subsets of a set X then

$$\mathcal{T} = \{\emptyset, U, V, U \cup V, U \cap V, X\}$$

is a topology on X. Find the closed sets of X.

Exercise 23.12. Prove that if $(\mathcal{T}_j)_{j \in J}$ is a family of topologies $\mathcal{T}_j \subset \mathcal{P}(X)$ on X then $\bigcap_{j \in J} \mathcal{T}_j$ is a topology on X.

Definition 23.13. If $\mathcal{T}_0 \subset \mathcal{P}(X)$ is a subset of the power set then the intersection

$$\mathcal{T} = \bigcap_{\mathcal{T}' \supset \mathcal{T}_0} \mathcal{T}'$$

of all topologies $\mathcal{T}'$ containing $\mathcal{T}_0$ is called the topology *generated* by $\mathcal{T}_0$.

Exercise 23.14. Let $\mathcal{T}_0 = \{U, V, W\} \subset \mathcal{P}(X)$. Find explicitly the topology generated by $\mathcal{T}_0$. Find all the closed sets in that topology.

Definition 23.15. A *topological space* is a pair $(X, \mathcal{T})$ consisting of a set X and a topology $\mathcal{T} \subset \mathcal{P}(X)$ on X. Sometimes one writes X instead of $(X, \mathcal{T})$ if $\mathcal{T}$ is understood from context.

Definition 23.16. Let X and X' be two topological spaces. A map $F : X \to X'$ is *continuous* if and only if for all open $V \subset X'$ the set $F^{-1}(V) \subset X$ is open.

Exercise 23.17. If $\mathcal{T}$ is a topology on X and $\mathcal{T}'$ is the topology on X' defined by $\mathcal{T}' = \{\emptyset, Y\}$ then every map $F : X \to X'$ is continuous.

Exercise 23.18. If $\mathcal{T}$ is the topology $\mathcal{T} = \mathcal{P}(X)$ on X and $\mathcal{T}'$ is any topology on X' then every map $F : X \to X'$ is continuous.

Exercise 23.19. Prove that if X, X', X'' are topological spaces and $G : X \to X'$, $F : X' \to X''$ are continuous maps then the composition $F \circ G : X \to X''$ is continuous.

Exercise 23.20. Give an example of two topological spaces X, X' and of a bijection $F : X \to X'$ such that F is continuous but F^{-1} is not continuous. (This is to be contrasted with the situation of algebraic structures to be discussed later. See Exercise 12.27.)

Motivated by the above phenomenon, one gives the following

Definition 23.21. A *homeomorphism* between two topological spaces is a continuous bijection whose inverse is also continuous.

Exercise 23.22. Prove that if X is a topological space and $Y \subset X$ is open then the induced topology on Y consists of all open sets of X that are contained in Y.

Definition 23.23. Let X be a topological space and let $A \subset X$ be a subset. We say that A is *connected* if and only if whenever U and V are open in X with $U \cap V \cap A = \emptyset$ and $A \subset U \cup V$ it follows that $U \cap A = \emptyset$ or $V \cap A = \emptyset$.

Exercise 23.24. Prove that if $F : X \to X'$ is continuous and $A \subset X$ is connected then $F(A) \subset X'$ is connected.

Definition 23.25. Let X be a topological space and let $A \subset X$ be a subset. A point $x \in X$ is called an *accumulation point* of A if and only if for every open set U in X containing x the set $U \backslash \{x\}$ contains a point of A.

Exercise 23.26. Let X be a topological space and let $A \subset X$ be a subset. Prove that A is closed if and only if A contains all its accumulation points.

Definition 23.27. A topological space X is a *Hausdorff* space if and only if for every two points $x, y \in X$ there exist open sets $U \subset X$ and $V \subset X$ such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

Exercise 23.28. Prove that $\mathbb{R}$ with the Euclidean topology is a Hausdorff space.

Definition 23.29. Let X be a topological space and $A \subset X$. We say A is *compact* if and only if whenever

$$A \subset \bigcup_{i \in I} U_i$$

with $(U_i)_{i \in I}$ a family of open sets in X indexed by some set I there exists a finite subset $J \subset I$ such that

$$A \subset \bigcup_{j \in J} U_j.$$

We sometimes refer to $(U_i)_{i \in I}$ as an *open cover* of A and to $(U_j)_{j \in J}$ as a *finite open subcover*. So A is compact if and only if every open cover of A has a finite open subcover.

Exercise 23.30. Prove that $\mathbb{R}$ is not compact in the Euclidean topology.

Hint: Consider the open cover

$$\mathbb{R} = \bigcup_{n \in \mathbb{N}} (-n, n)$$

and show it has no finite open subcover.

Exercise 23.31. Prove that no open interval (a, b) in $\mathbb{R}$ is compact ($a < b$).

Exercise 23.32. Prove that if $F : X \to X'$ is a continuous map of topological spaces and $A \subset X$ is compact then $F(A) \subset X'$ is compact.

Exercise 23.33. Prove that if X is a Hausdorff space, $A \subset X$ is compact, and $x \in X \backslash A$ then there exist open sets $U \subset X$ and $V \subset X$ such that $x \in U$, $A \subset V$, and $U \cap V = \emptyset$. In particular every compact subset of a Hausdorff space is closed.

Hint: For every $a \in A$ let $U_a \subset X$ and $V_a \subset X$ be open sets such that $x \in U_a$, $a \in V_a$, $U_a \cap V_a = \emptyset$. Then $(V_a)_{a \in A}$ is an open covering of A. Select $(V_b)_{b \in B}$ a finite subcover of A where $B \subset A$ is a finite set, $B = \{b_1, ..., b_n\}$. Then let

$$U = U_{b_1} \cap ... \cap U_{b_n}$$
$$V = V_{b_1} \cup ... \cup V_{b_n}.$$

Definition 23.34. Let X, X' be topological spaces. Then the set $X \times X'$ may be equipped with the topology generated by the family of all sets of the form $U \times U'$ where U and U' are open in X and X' respectively. This is called the *product topology* on $X \times X'$. Iterating this we get a product topology on a product $X_1 \times ... \times X_n$ of n topological spaces.

Exercise 23.35. Prove that the Euclidean topology on $\mathbb{C}$ coincides with the product topology on $\mathbb{R}^2$ (when $\mathbb{R}$ is given the Euclidean topology).

Definition 23.36. A *topological manifold* is a topological space X such that for every point $x \in X$ there exists an open set $U \subset X$ containing x and a homeomorphism $F : U \to V$ where $V \subset \mathbb{R}^n$ is an open set in $\mathbb{R}^n$ for the Euclidean topology. (Here U and V are viewed as topological spaces with the topologies induced from X and $\mathbb{R}^n$, respectively.)

Remark 23.37. If $\mathcal{X}$ is a set of topological manifolds then one can consider the following relation $\sim$ on $\mathcal{X}$: for $X, X' \in \mathcal{X}$ we let $X \sim X'$ if and only if there exists a homeomorphism $X \to X'$. Then $\sim$ is an equivalence relation on $\mathcal{X}$ and one of the basic problems of topology is to "describe" the set $\mathcal{X}/\sim$ of equivalence classes in various specific cases.

More properties of the Euclidean topology of $\mathbb{R}$ and $\mathbb{C}$ will be examined in the chapter on limits.

CHAPTER 24

Groups

Our next Chapters investigate a few topics in algebra. Recall that algebra is the study of algebraic structures, i.e., sets with operations on them. We already introduced, and constructed, some elementary examples of algebraic structures such as rings and, in particular, fields. With rings/fields at our disposal one can study some other fundamental algebraic objects such as groups, vector spaces, polynomials. In what follows we briefly survey some of these. We begin with groups. In some sense groups are more fundamental than rings and fields; but in order to be able to look at more interesting examples we found it convenient to postpone the discussion of groups until this point. Groups appeared in Mathematics in the context of symmetries of roots of polynomial equations; cf. the work of Galois that involved finite groups. Galois' work inspired Lie who investigated differential equations in place of polynomial equations; this led to (continuous) Lie groups, in particular groups of matrices. Groups eventually penetrated most of Mathematics and Physics (Klein, Poincaré, Einstein, Cartan, Weyl).

Recall the following:

Definition 24.1. A *group* is a tuple $(G, \star, e)$ consisting of a set G, a binary operation $\star$ on G, and an element $e \in G$ (called the *identity* element) such that the following conditions are satisfied:

1) For every $x, y, z \in G$, $x \star (y \star z) = (x \star y) \star z$;
2) For every $x \in G$, $x \star e = e \star x = x$;
3) For every $x \in G$ there exists $x' \in G$ such that $x \star x' = x' \star x = e$.

If in addition $x \star y = y \star x$ for all $x, y \in G$ we say G is commutative (or Abelian in honor of Abel).

Remark 24.2. The element x' in 3) above is easily proved to be unique.

Exercise 24.3. Check the uniqueness of x' in 3).

Sometimes one writes $e = 1$, $x \star y = xy$, $x' = x^{-1}$, $x^n = x \star ... \star x$ (n times). In the Abelian case one sometimes writes $e = 0$, $x \star y = x + y$, $x' = -x$, $x \star ... \star x = nx$ (n times). These notations depend on the context and are justified by the following examples.

Example 24.4. If R is a ring then R is an Abelian group with the following operations: $e = 0$, $x \star y = x + y$, $x' = -x$. Hence $\mathbb{Z}, \mathbb{Z}/m\mathbb{Z}, \mathbb{F}_p, \mathbb{Z}_p, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are groups "with respect to addition."

Example 24.5. If K is a field then $K^{\times} = K \backslash \{0\}$ is an Abelian group with $e = 1$, $x \star y = xy$, $x' = x^{-1}$. Hence $\mathbb{Q}^{\times}, \mathbb{R}^{\times}, \mathbb{C}^{\times}, \mathbb{F}_p^{\times}$ are groups "with respect to multiplication."

EXAMPLE 24.6. The set $\mathrm{Perm}(X)$ of bijections $\sigma : X \to X$ from a set X into itself is a group with $e = 1_X$ (the identity map), $\sigma \star \tau = \sigma \circ \tau$ (composition), σ^{-1} =inverse map. If $X = \{1, ..., n\}$ then one writes $S_n = \mathrm{Perm}(X)$ and calls this group the *symmetric group*. Its elements are called *permutations*. If $\sigma(1) = i_1, ..., \sigma(n) = i_n$ one usually writes

$$\sigma = \begin{pmatrix} 1 & 2 & ... & n \\ i_1 & 1_2 & ... & i_n \end{pmatrix}.$$

If $\sigma(i) = j \neq i$, $\sigma(j) = i$, and $\sigma(k) = k$ for $k \notin \{i, j\}$ we call σ a *transposition*. Every permutation is a product of transpositions (as one can see by induction).

EXERCISE 24.7. Compute

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 4 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 2 & 1 & 3 \end{pmatrix}.$$

Also compute

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 4 & 3 \end{pmatrix}^{-1}.$$

EXAMPLE 24.8. A 2×2 matrix with coefficients in a field K is a map

$$A : \{1, 2\} \times \{1, 2\} \to K.$$

If the map is given by

$$\begin{aligned} A(1,1) &= a \\ A(1,2) &= b \\ A(2,1) &= c \\ A(2,2) &= d \end{aligned}$$

we write A as

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Define the sum and the product of two matrices by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix},$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

Define the product of an element $r \in K$ with a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by

$$r \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ra & rb \\ rc & rd \end{pmatrix}.$$

For a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ define its determinant by

$$\det(A) = ad - bc.$$

Say that A is invertible if $\det(A) \neq 0$ and setting $\delta = \det(A)$ define the inverse of A by

$$A^{-1} = \delta^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Define the identity matrix by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the zero matrix by

$$O = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Let $M_2(K)$ be the set of all matrices and $GL_2(K)$ be the set of all invertible matrices. Then the following are true:

1) $M_2(K)$ is a group with respect to addition of matrices;

2) $GL_2(K)$ is a group with respect to multiplication of matrices; it is called the *general linear group* of 2×2 matrices;

3) $(A+B)C = AC+BC$ and $C(A+B) = CA+CB$ for every matrices A, B, C;

4) There exist matrices A, B such that $AB \neq BA$;

5) $\det(AB) = \det(A) \cdot \det(B)$.

EXERCISE 24.9. Prove 1), 2), 3), 4), 5) above.

EXAMPLE 24.10. Groups are examples of algebraic structures so there is a well-defined notion of homomorphism of groups (or group homomorphism). According to the general definition a *group homomorphism* is a map between the two groups $F : G \to G'$ such that for all $a, b \in G$:

1) $F(a \star b) = F(a) \star' F(b)$,

2) $F(a^{-1}) = F(a)^{-1}$ (this is automatic !),

3) $F(e) = e'$ (this is, again, automatic !).

Here $\star$ and $\star'$ are the operations on G and G'; similarly e and e' are the corresponding identity elements. A bijective group homomorphism is a group isomorphism.

DEFINITION 24.11. A subset H of a group G is called a *subgroup* if

1) For all $a, b \in H$ we have $a \star b \in H$.

2) For all $a \in H$ we have $a^{-1} \in H$.

3) $e \in H$.

EXERCISE 24.12. Show that if H is a subgroup of G then H, with the natural operation induced from G, is a group.

EXERCISE 24.13.

1) $\mathbb{Z}$ is a subgroup of $\mathbb{Q}$.

2) $\mathbb{Q}$ is a subgroup of $\mathbb{R}$.

3) $\mathbb{R}$ is a subgroup of $\mathbb{C}$.

4) If K is a field then the set

$$SL_2(K) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b \in K,\ ad - bc = 1\}$$

is a subgroup of $GL_2(K)$; it is called the *special linear group*.

5) If K is a field then the set

$$SO_2(K) = \{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a, b \in K,\ a^2 + b^2 = 1\}$$

is a subgroup of $SL_2(K)$; it is called the *special orthogonal group*.

DEFINITION 24.14. If $F : G \to G'$ is a group homomorphism define the *kernel* of F,

$$Ker\ F = \{a \in G \mid F(a) = e'\}$$

and the *image* of F:

$$Im\ F = \{b \in G' \mid \exists a \in G, F(a) = b\}.$$

EXERCISE 24.15. Prove that $Ker\ F$ is a subgroup of G and $Im\ F$ is a subgroup of G'.

We continue our investigation of groups and introduce the concept of *order* of elements in a group. This has nothing to do with the word "order" used in the phrases "first order logic" or "order relations" or "order of a graph."

DEFINITION 24.16. Let G be a group and $g \in G$; we denote by $\langle g \rangle$ the set of all elements $a \in G$ for which there exists $n \in \mathbb{Z}$ such that $a = g^n$.

EXERCISE 24.17. Prove that $\langle g \rangle$ is a subgroup of G. We call $\langle g \rangle$ the subgroup *generated* by g.

DEFINITION 24.18. We say that a group G is *cyclic* if and only if there exists $g \in G$ such that $G = \langle g \rangle$; g is called a *generator* of G.

EXAMPLE 24.19. $\mathbb{Z}$ is cyclic. 1 is a generator of $\mathbb{Z}$; -1 is also a generator of $\mathbb{Z}$.

EXERCISE 24.20. Prove that $\mathbb{Q}$ is not cyclic.

DEFINITION 24.21. Let G be a group and $g \in G$. We say the order of g is *infinite* if $g^n \neq e$ for all $n \in \mathbb{N}$. We say the order of g is $n \in \mathbb{N}$ if:
1) $g^n = e$;
2) $g^k \neq e$ for all $k \in \mathbb{N}$ with $k < n$.
We denote by $o(g)$ the order of g.

DEFINITION 24.22. The *order* of a finite group G is the cardinality $|G|$ of G.

EXERCISE 24.23. The order $o(g)$ of g equals the order $|\langle g \rangle|$ of $\langle g \rangle$.

EXERCISE 24.24. g has order $n \in \mathbb{N}$ if and only if:
1') $g^n = e$;
2') If $g^N = e$ for some $N \in \mathbb{N}$ then $n|N$.

Hint: If 1') and 2') above hold then clearly g has order n. Conversely if g has order n then 1') clearly holds. To check that 2') holds use Euclidean division to write $N = nq + r$ with $0 \leq r < n$. Then $g^r = (g^n)^q g^r = g^N = e$. By condition 2) in the definition of order $r = 0$ hence $n|N$.

In what follows we say that two integers are coprime if they have no common divisor > 1.

PROPOSITION 24.25. *Assume a, b are two elements in a group such that $ab = ba$ and assume $o(a)$ and $o(b)$ are coprime. Then*

$$o(ab) = o(a)o(b).$$

Proof. Set $k = o(a)$, $l = o(b)$. Clearly, since $ab = ba$ we have

$$(ab)^{kl} \equiv (a^k)^l (b^l)^k = e.$$

Now assume $(ab)^N = e$. Raising to power l we get $a^{Nl}b^{Nl} = e$, hence $a^{Nl} = e$, hence, by Exercise 24.24, $k|Nl$. Since k and l are coprime $k|N$ (by the Fundamental Theorem of Arithmetic). In a similar way raising $(ab)^N = e$ to power k we get $a^{Nk}b^{Nk} = e$, hence $b^{Nk} = e$, hence $l|Nk$, hence $l|N$. Again, since k and l are coprime, $l|N$ and $k|N$ imply $kl|N$ and we are done. $\square$

Exercise 24.26. Prove that if $o(a) = kl$ then $o(a^k) = l$.

Theorem 24.27. *(Lagrange) If H is a subgroup of a finite group G then the order of H divides the order of G: if $n = |H|$, $m = |G|$ then $n|m$. In particular if $a \in G$ then the order $o(a)$ of a divides the order $|G|$ of the group. So if $n = |G|$ then $a^n = e$.*

Proof. For each $g \in G$ we let gH be the set of all elements of G of the form gh with $h \in H$. Let $\pi : G \to \mathcal{P}(G)$ be the map $\pi(g) = gH \in \mathcal{P}(G)$. Let $\mathcal{X} = \pi(G)$ and let $\sigma : \mathcal{X} \to G$ be any map such that $\pi \circ \sigma$ is the identity of $\mathcal{X}$. (The existence of σ follows by induction.) We claim that the map

$$\mathcal{X} \times H \to G, \quad (X, h) \mapsto \sigma(X)h, \quad X \in \mathcal{X}, \quad h \in H \tag{24.1}$$

is a bijection. Assuming the claim for a moment note that the claim implies

$$|\mathcal{X}| \times |H| = |G|,$$

from which the theorem follows. Let us check the claim. To prove that 24.1 is surjective let $g \in G$. Let $g' = \sigma(gH)$. Then $g'H = \pi(g') = \pi(\sigma(gH)) = gH$. So there exists $h \in H$ such that $g'h = ge = g$; hence $g = \sigma(gH)h$ which ends the proof of surjectivity. We leave the proof of injectivity to the reader. $\square$

Exercise 24.28. Check the injectivity of 24.1.

Theorem 24.29. *(Fermat's Little Theorem) For every $a \in \mathbb{Z}$ and every prime p we have*

$$a^p \equiv a \pmod{p}.$$

Proof. If $p|a$ this is clear. If $p \not| a$ let $\overline{a}$ be the image of a in $\mathbb{F}_p^\times$. By Lagrange's theorem applied to the group $\mathbb{F}_p^\times$ we have $\overline{a}^{p-1} = \overline{1}$. Hence $a^{p-1} \equiv 1 \pmod{p}$. So $a^p \equiv a \pmod{p}$. $\square$

Definition 24.30. A subgroup H of a group G is called *normal* if for all $h \in H$ and all $g \in G$ we have $ghg^{-1} \in H$.

Exercise 24.31. Let H be a normal subgroup of a group G. Define an equivalence relation $\sim$ on G as follows: $g \sim k$ if and only $g^{-1}k \in H$; write this as

$$g \equiv k \mod H.$$

Denote by G/H the quotient $G/\sim$ of G by $\sim$. Define an operation on G/H by the formula

$$\widehat{g}\widehat{k} = \widehat{gk}.$$

Prove that the definition is correct and that G/H becomes a group with this operation and with identity element $\widehat{e}$ (with e the identity in G).

CHAPTER 25

Vectors

Vectors implicitly appeared in a number of contexts such as mechanics (Galileo, Newton, etc.), hypercomplex numbers (Hamilton, Cayley, etc.), algebraic number theory (Dirichlet, Kummer, Eisenstein, Kronecker, etc.), and analysis (Hilbert, Banach, etc.). They are now a basic concept in linear algebra which is itself part of abstract algebra.

Definition 25.1. Let R be a (commutative unital) ring. A *module* over R (or R-*module*) is an Abelian group $(V, +, -, 0)$ together with a map $R \times V \to V$, $(a, v) \mapsto av$ satisfying the following conditions for all $a, b \in R$ and all $u, v \in V$:

1) $(a + b)v = av + bv$;
2) $a(u + v) = au + av$;
3) $a(bv) = (ab)v$;
4) $1v = v$.

For a field K modules over K are called *vector spaces* (or *linear spaces*) over K.

Example 25.2. R^n is a module over R viewed with the operations

$$(a_1, ..., a_n) + (b_1, ..., b_n) = (a_1 + b_1, ..., a_n + b_n),$$
$$-(a_1, ..., a_n) = (-a_1, ..., -a_n),$$
$$c(a_1, ..., a_n) = (ca_1, ..., ca_n).$$

The elements of R^n are referred to as *vectors*.

Definition 25.3. The elements $u_1, ..., u_n \in V$ are *linearly independent* if and only if whenever $a_1, ..., a_n \in R$ satisfies $(a_1, ..., a_n) \neq (0, ..., 0)$ it follows that

$$a_1u_1 + ... + a_nu_n \neq 0.$$

Definition 25.4. The elements $u_1, ..., u_n \in V$ *generate* V if and only if for every $u \in V$ there exist $a_1, ..., a_n \in R$ such that $u = a_1u_1 + ..., a_nu_n$. (We also say that u is a linear combination of $u_1, ..., u_n$.)

Definition 25.5. The elements $u_1, ..., u_n \in V$ are a *basis* of V if and only if they are linearly independent and generate V.

Exercise 25.6.

1) Show that $(-1, 1, 0)$ and $(0, 1, -1)$ are linearly independent in R^3 but they do not generate R^3.

2) Show that $(-1, 1, 0), (0, 1, -1), (1, 0, 1), (0, 2, -1)$ generate R^3 but are not linearly independent in R^3.

3) Show that $(-1, 1, 0), (0, 1, -1), (1, 0, 1)$ is a basis in R^3.

Exercise 25.7. If V has a basis $u_1, ..., u_n$ then the map $R^n \to V$, $(a_1, ..., a_n) \mapsto a_1u_1 + ... + a_nu_n$ is bijective.

Hint: Directly from definitions.

Definition 25.8. A module over a ring is called *free of finite rank* if it has a finite basis.

Exercise 25.9. Give examples of modules over rings that are generated by finitely many elements but are not free of finite rank.

Exercise 25.10. If K is a field and V is generated by n elements then V has a basis consisting of at most n elements (hence it is free of finite rank).

Hint: Considering a subset of $\{u_1, ..., u_n\}$ minimal with the property that it generates V we may assume that every subset obtained from $\{u_1, ..., u_n\}$ does not generate V. We claim that $u_1, ..., u_n$ are linearly independent. Assume not. Hence there exists $(a_1, ..., a_n) \neq (0, ..., 0)$ such that $a_1u_1 + ... + a_nu_n = 0$. We may assume $a_1 = 1$. Then one checks that $u_2, ..., u_n$ generate V, contradicting minimality.

Exercise 25.11. Assume $K = \mathbb{F}_p$ and V has a basis with n elements. Then $|V| = p^n$.

Theorem 25.12. *If K is a field and V has a basis $u_1, ..., u_n$ and a basis $v_1, ..., v_m$ then $n = m$.*

Proof. We prove $m \leq n$; similarly one has $n \leq m$. Assume $m > n$ and seek a contradiction. Since $u_1, ..., u_n$ generate V we may write $v_1 = a_1u_1 + ... + a_nu_n$ with not all $a_1, ..., a_n$ zero. Renumbering $u_1, ..., u_n$ we may assume $a_1 \neq 0$. Hence $v_1, u_2, ..., u_n$ generates V. Hence $v_2 = b_1v_1 + b_2u_2 + ... + b_nu_n$. But not all $b_2, ..., b_n$ can be zero because v_1, v_2 are linearly independent. So renumbering $u_2, ..., u_n$ we may assume $b_2 \neq 0$. So $v_1, v_2, u_3, ..., u_n$ generates V. Continuing (one needs induction) we get that $v_1, v_2, ..., v_n$ generates V. So $v_{n+1} = d_1v_n + ... + d_nv_n$. But this contradicts the fact that $v_1, ..., v_m$ are linearly independent. $\square$

Exercise 25.13. Give a quick proof of the above theorem in case $K = \mathbb{F}_p$.

Hint: We have $p^n = p^m$ hence $n = m$.

Remark 25.14. Theorem 25.12 remains true for an arbitrary (commutative unital) ring instead of a field; we will not need this in what follows.

Definition 25.15. Assume K is a field. We say that a vector space V over K is *finite dimensional* (or that it has a finite basis) if and only if there exists a basis $u_1, ..., u_n$ of V. Then we define the dimension of V to be n; write $\dim V = n$. (The definition is correct due to Theorem 25.12.)

Exercise 25.16. Prove that if V is a finite dimensional space over a field K and $W \subset V$ is a *subspace* (i.e., a subset closed under addition and multiplication by elements in K) then $\dim W \leq \dim V$ with equality holding if and only if $W = V$.

Definition 25.17. If V and W are R-modules a map $F : V \to W$ is called *linear* (or *R-linear* or an *R-module homomorphism*) if and only if for all $a \in R$, $u, v \in V$ we have:
1) $F(au) = aF(u)$,
2) $F(u + v) = F(u) + F(v)$.

Example 25.18. If $a, b, c, d, e, f \in R$ then the map $F : R^3 \to R^2$ given by

$$F(u, v, w) = (au + bv + cw, du + ev + fw)$$

is a linear map.

EXERCISE 25.19. Prove that if $F : V \to W$ is a linear map of R-modules then $V' = F^{-1}(0)$ and $V'' = F(V)$ are R-modules (with respect to the obvious operations). If in addition V and W are finite dimensional vector spaces over a field then V' and V'' are finite dimensional and

$$\dim V = \dim V' + \dim V''.$$

Hint: Construct corresponding bases.

EXERCISE 25.20. Give an example of a vector space that is not finite dimensional.

DEFINITION 25.21. A subset M' of a module M over a ring R is called a *submodule* if for all $a, b \in M'$ and $r \in R$ we have $a + b \in M'$, $-a \in M'$, $ra \in M'$. Such an M' becomes then an R-module.

EXERCISE 25.22. Let M' be a submodule of a module M. Define an equivalence relation $\sim$ on M as follows: $a \sim b$ if and only $a - b \in M'$; write this as

$$a \equiv b \mod M'.$$

Denote by M/M' the quotient $M/\sim$ of M by $\sim$. Define operations on M/M' by the formulae

$$\widehat{a} + \widehat{b} := \widehat{a+b}, \quad r \cdot \widehat{a} := \widehat{ra}$$

Prove that the definition is correct and that M/M' becomes a module over R with these operations. If in addition R is a field K and M has finite dimension then prove that

$$\dim M' + \dim M/M' = \dim M.$$

CHAPTER 26

Matrices

Matrices appeared in the context of linear systems of equations and were studied in the work of Leibniz, Cramer, Cayley, Eisenstein, Hamilton, Sylvester, Jordan, etc. They were later rediscovered and applied in the context of Heisenberg's matrix mechanics. Nowadays they are a standard concept in linear algebra courses.

DEFINITION 26.1. Let $m, n \in \mathbb{N}$. An $m \times n$ *matrix* with coefficients in a ring R is a map

$$A : \{1, ..., m\} \times \{1, ..., n\} \to R.$$

If $A(i, j) = a_{ij}$ for $1 \leq i \leq m$, $1 \leq j \leq n$ then we write

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{m1} & ... & a_{mn} \end{pmatrix}.$$

We denote by

$$R^{m\times n} = M_{m\times n}(R)$$

the set of all $m \times n$ matrices. We also write $M_n(R) = M_{n\times n}(R)$. Note that $R^{1\times n}$ identifies with R^n; its elements are of the form

$$(a_1, ..., a_n)$$

and are called *row* matrices (row vectors). Similarly the elements of $R^{m\times 1}$ are of the form

$$\begin{pmatrix} a_1 \\ ... \\ ... \\ a_m \end{pmatrix}$$

and are called *column* matrices (column vectors). If $A = (a_{ij}) \in R^{m\times n}$ then

$$u^1 = \begin{pmatrix} a_{11} \\ ... \\ ... \\ a_{m1} \end{pmatrix}, ..., u^n = \begin{pmatrix} a_{1n} \\ ... \\ ... \\ a_{mn} \end{pmatrix}$$

are called the *columns* of A and we also write

$$A = (u^1, ..., u^n).$$

Similarly

$$(a_{11}, ..., a_{1n}), ..., (a_{m1}, ..., a_{mn})$$

are called the *rows* of A.

DEFINITION 26.2. Let $0 \in R^{m\times n}$ the matrix $0 = (z_{ij})$ with $z_{ij} = 0$ for all i, j; 0 is called the *zero* matrix. Let $I \in R^{n\times n}$ the matrix $I = (\delta_{ij})$ where $\delta_{ii} = 1$ for all i and $\delta_{ij} = 0$ for all $i \neq j$; I is called the *identity* matrix and δ_{ij} is called the *Kronecker symbol*.

DEFINITION 26.3. If $A = (a_{ij}), B = (b_{ij}) \in R^{m\times n}$ we define the sum

$$C = A + B \in R^{m\times n}$$

as

$$C = (c_{ij}), \quad c_{ij} = a_{ij} + b_{ij}.$$

If $A = (a_{is}) \in R^{m\times k}$, $B = (b_{sj}) \in R^{k\times n}$, we define the product

$$C = AB \in R^{m\times n}$$

as

$$C = (c_{ij}), \quad c_{ij} = \sum_{s=1}^{k} a_{is} b_{sj}.$$

If $A = (a_{ij}) \in R^{n\times m}$ and $\lambda \in R$ define the matrix

$$\lambda A = (\lambda a_{ij}) \in R^{n\times m}.$$

EXERCISE 26.4. Prove that:
1) $R^{m\times n}$ is a group with respect to $+$.
2) $A(BC) = (AB)C$ for all $A \in R^{m\times k}$, $B \in R^{k\times l}$, $C \in R^{l\times n}$.
3) $A(B + C) = AB + AC$ for all $A \in R^{m\times k}$ and $B, C \in R^{k\times n}$.
4) $(B + C)A = BA + CA$ for all $B, C \in R^{m\times k}$ and $A \in R^{k\times n}$.
5) $AI = IA$ for all $A \in R^{n\times n}$.
6) $\lambda(A + B) = \lambda A + \lambda B$ and $(\lambda A)B = A(\lambda B) = \lambda(AB)$ for all $\lambda \in R$.

DEFINITION 26.5. For any two square matrices $A, B \in R^{n\times n}$ one defines the *commutator* $[A, B] \in R^{n\times}$ by the formula

$$[A, B] := AB - BA.$$

EXERCISE 26.6. Prove that for every square matrices A, B, C we have the equality (called the Jacobi identity):

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

EXERCISE 26.7. If $A \in R^{m\times k}$, $B \in R^{k\times n}$, and the columns of B are $b^1, ..., b^n \in R^{k\times 1}$ then the columns of AB are $Ab^1, ..., Ab^n \in R^{m\times 1}$ (where Ab^i is the product of the matrices A and b^i). In other words

$$B = (b^1, ..., b^n) \quad \rightarrow \quad AB = (Ab^1, ..., Ab^n).$$

DEFINITION 26.8.

$$\begin{pmatrix} 1 \\ 0 \\ ... \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ ... \\ 0 \end{pmatrix}, \quad ..., \quad \begin{pmatrix} 0 \\ 0 \\ ... \\ 1 \end{pmatrix}$$

is called the *standard basis* of $R^{m\times 1}$. These vectors are usually denoted by $e^1, ..., e^n$.

EXERCISE 26.9. Prove that the above is indeed a basis of $R^{m\times 1}$.

Here is the link between linear maps and matrices:

DEFINITION 26.10. If $F : V \to W$ is a linear map of R-modules and $v_1, ..., v_n$ and $w_1, ..., w_m$ are bases of V and W, respectively, then for $j = 1, ..., n$ one can write uniquely

$$F(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

The matrix $A = (a_{ij}) \in R^{m\times n}$ is called the *matrix of* F with respect to the bases $v_1, ..., v_n$ and $w_1, ..., w_m$.

EXERCISE 26.11. Consider a matrix $A = (a_{ij}) \in R^{m\times n}$ and consider the map

$$F : R^{n\times 1} \to R^{m\times 1}, \quad F(u) = Au \ \ \text{(product of matrices)}.$$

Then the matrix of F with respect to the canonical bases of $R^{n\times 1}$ and $R^{m\times 1}$ is A itself.

Hint: Let $e^1, ..., e^n$ be the standard basis of $R^{n\times 1}$ and let $f^1, ..., f^m$ be the standard basis of $R^{m\times 1}$. Then

$$F(e^1) = Ae^1 = \left(\begin{array}{ccc} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{m1} & ... & a_{mn} \end{array}\right) \left(\begin{array}{c} 1 \\ 0 \\ ... \\ 0 \end{array}\right) = \left(\begin{array}{c} a_{11} \\ a_{21} \\ ... \\ a_{m1} \end{array}\right) = a_{11}f^1 + ... + a_{m1}f^m.$$

A similar computation can be done for $e^2, ..., e^n$.

EXERCISE 26.12. Let $F : R^{n\times 1} \to R^{m\times 1}$ be a linear map and let $A \in R^{m\times n}$ be the matrix of F with respect to the standard bases. Then for all $u \in R^{n\times 1}$ we have $F(u) = Au$.

EXERCISE 26.13. Let $G : R^{n\times 1} \to R^{k\times 1}$ and let $F : R^{k\times 1} \to R^{m\times 1}$ be linear maps. Let A be the matrix of F with respect to standard bases and let B be the matrix of G with respect to the standard bases. Then the matrix of $F \circ G$ with respect to the standard bases is AB (product of matrices).

Hint: $F(G(u)) = A(Bu) = (AB)u$.

DEFINITION 26.14. For every matrix $A = (a_{ij}) \in R^{n\times n}$ define its *trace*

$$Tr(A) = \sum_{i=1}^{n} a_{ii}.$$

EXERCISE 26.15. Let $A, B \in R^{n\times n}$ and $c \in R$.
1) Prove that $Tr(A + B) = Tr(A) + Tr(B)$ and $Tr(cA) = c \cdot Tr(A)$.
2) Prove that $Tr(AB) = Tr(BA)$.
3) Prove that if R contains $\mathbb{Z}$ then $AB - BA \neq I$.
4) If $F : V \to V$ is a linear map on a free module of finite rank V and A, B are the matrices of F with respect to two bases of V then $Tr(A) = Tr(B)$. We denote any of these traces by $Tr(F)$.

REMARK 26.16. Part 3) in the previous Exercise is relevant for the foundations of quantum mechanics. Indeed, a relation such as $AB - BA = I$ is necessary in quantum mechanics to express the "Heisenberg uncertainty principle." Since, by 3) above, this relation is impossible the "baby model" for quantum mechanics in Example 26.36 below needs to be replaced with one that uses infinite dimensional vector spaces and infinite matrices.

Definition 26.17. Let R be a ring. If $A = (a_{ij}) \in R^{m\times n}$ is a matrix one defines the *transpose* of A as the matrix $A^t = (a'_{ij}) \in R^{n\times m}$ where $a'_{ij} = a_{ji}$.

Example 26.18.

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}^t = \begin{pmatrix} a & d \\ b & e \\ c & f \end{pmatrix}.$$

Exercise 26.19. Prove that:
1) $(A+B)^t = A^t + B^t$ for $A, B \in R^{n\times m}$.
2) $(AB)^t = B^tA^t$ for $A \in R^{n\times m}$ and $B \in R^{m\times k}$.
3) $I^t = I$.

Definition 26.20. Let $F : V \to V$ be a linear map with V a vector space over a field K.

1) We say $\lambda \in K$ is an *eigenvalue* of F if and only if there exists $v \in V$, $v \neq 0$, such that $F(v) = \lambda v$. We say that v belongs to λ. The vector space $V_\lambda = \{v \in V \mid F(v) = \lambda v\}$ is called the *eigenspace* of F corresponding to λ.

2) We say that $v \in V$ is an eigenvector of F if and only if $v \neq 0$ there exists $\lambda \in K$ such that $F(v) = \lambda v$. We say that λ corresponds to v.

3) If $A \in K^{n\times n}$ and $V = K^{n\times 1}$ then the eigenvalues and the eigenvectors of the map $F : V \to V$, $F(v) = Av$, are referred to as the eigenvalues (resp. eigenvectors) of A.

Remark 26.21. Later we will show that A and A^t have the same eigenvalues.

Exercise 26.22. Let $F : V \to V$ be a linear map with V a vector space over a field K. Let $\lambda_1, ..., \lambda_n \in K$ be distinct eigevalues of F and $v_1, ..., v_n$ be eigenvectors belonging to them, respectively. Then $v_1, ..., v_n$ are linearly independent.

Hint: Assume the contrary, assume n is minimal with the property that the statement is false, and consider a linear combination

$$c_1v_1 + ... + c_nv_n = 0$$

with not all c_i zero. Then by minimality none of the c_is is 0. Applying F one gets

$$\lambda_1c_1v_1 + ... + \lambda_nc_nv_n = 0.$$

Multiply the first equation by λ_1 and subtract from the second equation. One gets

$$(\lambda_2 - \lambda_1)c_2v_2 + ... = 0,$$

a contradiction with minimality.

Definition 26.23. For every $A = (a_{ij}) \in \mathbb{C}^{n\times m}$ we define $\overline{A} = (\overline{a_{ij}}) \in \mathbb{C}^{n\times m}$ and $A^* = \overline{A}^t \in \mathbb{C}^{m\times n}$.

Exercise 26.24. Prove that
1) $(A+B)^* = A^* + B^*$ for $A, B \in \mathbb{C}^{n\times m}$.
2) $(AB)^* = B^*A^*$ for $A \in \mathbb{C}^{n\times m}$ and $B \in \mathbb{C}^{m\times k}$.
3) $I^* = I$.

Definition 26.25.
A matrix $A \in R^{n\times n}$ is called *orthogonal* if $AA^t = I$ (equivalently $A^tA = I$).
A matrix $A \in R^{n\times n}$ is called *symmetric* if $A^t = A$.
A matrix $A \in R^{n\times n}$ is called *anti-symmetric* if $A^t + A = 0$.

A matrix $A \in \mathbb{C}^{n\times n}$ is called *unitary* if $AA^* = I$ (equivalently $A^*A = I$).
A matrix $A \in \mathbb{C}^{n\times n}$ is called *Hermitian* if $A = A^*$.
A matrix $A \in \mathbb{C}^{n\times n}$ is called *anti-Hermitian* if $A + A^* = 0$.

EXERCISE 26.26. For $\varphi, \psi \in V := \mathbb{C}^n = \mathbb{C}^{n\times 1}$ define their Hermitian product

$$\langle \varphi, \psi \rangle := \psi^* \varphi \in \mathbb{C} = \mathbb{C}^{1\times 1}.$$

1) Prove that $\langle \varphi, \varphi \rangle \in \mathbb{R}_{>0}$ for $\varphi \neq 0$.
2) Prove that $\overline{\langle \varphi, \psi \rangle} = \langle \psi, \varphi \rangle$.
3) Prove that $\langle \lambda\varphi, \psi \rangle = \lambda \langle \varphi, \psi \rangle$ and $\langle \varphi, \lambda\psi \rangle = \overline{\lambda} \langle \varphi, \psi \rangle$ for $\lambda \in \mathbb{C}$.
4) Prove that for all non-zero $\varphi, \psi \in \mathbb{C}^n$ we have the following inequality (of Cauchy-Schwarz):

$$|\langle \varphi, \psi \rangle|^2 \leq \langle \varphi, \varphi \rangle \langle \psi, \psi \rangle$$

with equality if and only if $\varphi + \lambda\psi = 0$ for some $\lambda \in \mathbb{C}$. Conclude that if $||\varphi|| := \sqrt{\langle \varphi, \varphi \rangle}$, and similarly for ψ then for all $z \in \mathbb{C}$,

$$||\varphi + \psi|| \leq ||\varphi|| + ||\psi||, \quad ||z\varphi|| = |z| \cdot ||\varphi||.$$

5) Let us say that a basis $\Psi := (\psi_1, ..., \psi_n) \in \mathbb{C}^n \times ... \times \mathbb{C}^n = \mathbb{C}^{n\times n}$ of V is orthonormal if $\langle \psi_i, \psi_j \rangle = \delta_{ij}$ for all i, j. Prove that if $\psi = \sum_i u_i \psi_i$ with $\psi_1, ..., \psi_n$ orthonormal and $u_i \in \mathbb{C}$ then $u_i = \langle \psi, \psi_i \rangle$. Prove that $||\psi|| = 1$ if and only if $\sum_i |u_i|^2 = 1$; in particular if $||\psi|| = 1$ then the family $(|u_i|^2)$ is the probability distribution of a probability measure $\mu = \mu_{\psi,\Psi}$ on $\Omega = \{1, ..., n\}$, $\mu(\{i\}) = |u_i|^2$.
6) Prove that for every matrix $A \in \mathbb{C}^{n\times n}$ and all $\varphi, \psi \in V$ one has $\langle A\varphi, \psi \rangle = \langle \varphi, A^*\psi \rangle$.
7) Let $\psi_1, ..., \psi_n$ be a basis of V. Prove that a matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian if and only if $\langle A\psi_i, \psi_j \rangle = \langle \psi_i, A\psi_j \rangle$.
8) Let $\varphi_1, ..., \varphi_n$ and $\psi_1, ..., \psi_n$ be two orthonormal bases. Then the matrix $U = (\langle \varphi_i, \psi_j \rangle)$ is unitary. Moreover if Φ and Ψ are the matrices with columns the two orthonormal bases then $U\Psi^t = \Phi^t$; equivalently $\Psi U^t = \Phi$ or $\Psi = \Phi \overline{U}$.
9) Let $\psi = (\psi_1, ..., \psi_n)$ be an orthonormal basis of V and $\lambda_1, ..., \lambda_n \in \mathbb{R}$. Let $A \in \mathbb{C}^{n\times n}$ be the unique matrix such that $A\psi_i = \lambda_i \psi_i$. Prove that A is Hermitian and that for all ψ the following holds:

$$\langle A\psi, \psi \rangle = \sum_i \lambda_i |\langle \psi, \psi_i \rangle|^2 \in \mathbb{R}.$$

10) Let $A = (a_{ij}) \in \mathbb{C}^{n\times n}$. Prove that the set

$$\{\frac{||A\psi||}{||\psi||} \mid \psi \neq 0\}$$

is bounded and let $||A||$ be the supremum of this set. Prove that $|a_{ij}| \leq ||A||$ for all i, j. Prove that for all $A, B \in \mathbb{C}^{n\times n}$,

$$||A + B|| \leq ||A|| + ||B||, \quad ||AB|| \leq ||A|| \cdot ||B||.$$

Hint for 4): We have

$$\langle \varphi + \lambda\psi, \varphi + \lambda\psi \rangle \in \mathbb{R}_{>0}$$

for all $\lambda \in \mathbb{C}$ with $\varphi + \lambda\psi \neq 0$. Set $\lambda = r\langle \psi, \varphi \rangle$ with $r \in \mathbb{R}$ arbitrary and use the fact that if $a, b, c \in \mathbb{R}$, $a \neq 0$ are such that $ar^2 + br + c > 0$ for all $r \in \mathbb{R}$ then $b^2 - 4ac < 0$; for the latter take $r = -\frac{b}{2a}$.

Hint for 7): Prove that $\langle A\varphi, \psi\rangle = \langle \varphi, A\psi\rangle$ for all φ, ψ and specialize to the case where the latter are members e^i, e^j of the standard basis.

Definition 26.27. Let A be a Hermitian matrix and $\psi \in \mathbb{C}^n$ with $||\psi|| = 1$.
1) One defines the *expectation value* $\mathbb{E}_\psi(A)$ of A with respect to ψ by:

$$\mathbb{E}_\psi(A) := \langle A\psi, \psi\rangle.$$

2) One defines the *variance* $\mathbb{V}_\psi(A)$ of A with respect to ψ by:

$$\mathbb{V}_\psi(A) := \mathbb{E}_\psi((A - \mathbb{E}_\psi(A)I)^2).$$

Remark 26.28. Note that we have the following computation:

$$\begin{aligned} \mathbb{V}_\psi(A) &= \langle (A - E_\psi(A)I)\psi, A - E_\psi(A)I)\psi\rangle \\ &= \langle A\psi, A\psi\rangle - (\langle \psi, A\psi\rangle)^2. \end{aligned}$$

Exercise 26.29. Prove that if A is Hermitian and $||\psi|| = 1$ then $\mathbb{V}_\psi(A) = 0$ if and only if ψ is an eigenvector of A.

Hint: If $\mathbb{V}_\psi(A) = 0$ then, by Remark 26.28 we get $(A - E_\psi(A))\psi = 0$ so ψ is an eigenvector. The converse follows again from Remark 26.28.

Remark 26.30. Consider the situation in 9) of Exercise 26.26 with $||\psi|| = 1$. Consider the probability measure $\mu_{\psi,\Psi}$ on $\Omega := \{1, ..., n\}$ defined by $\mu_{\psi,\Psi}(\{i\}) = |\langle \psi, \psi_i\rangle|^2$ and consider the random variable $\lambda : \Omega \to \mathbb{R}$ defined by $\lambda(i) := \lambda_i$. Then by loc.cit. the real number $\mathbb{E}_\psi(A)$ equals the expectation value $\mathbb{E}_{\mu_{\psi,\Psi}}(\lambda)$ of λ with respect to $\mu_{\psi,\Psi}$:

$$\mathbb{E}_\psi(A) = \mathbb{E}_{\mu_{\psi,\Psi}}(\lambda).$$

Since the eigenvalues of $A - \mathbb{E}_\psi(A)I$ corresponding to ψ_i are $\lambda_i - \mathbb{E}_\psi(A)$ we also have that the variance $\mathbb{V}_\psi(A)$ equals the variance of λ with respect to $\mu_{\psi,\Psi}$:

$$\mathbb{V}_\psi(A) = \mathbb{V}_{\mu_{\psi,\Psi}}(\lambda).$$

Proposition 26.31. *(Heisenberg Uncertainty Principle). If A and B are Hermitian and $||\psi|| = 1$ then we have*

$$\mathbb{V}_\psi(A) \cdot \mathbb{V}_\psi(B) \geq \frac{1}{4}|\mathbb{E}_\psi([A, B])|^2.$$

Proof. Let $A_0 := A - \mathbb{E}_\psi(A)I = A - \langle \psi, A\psi\rangle I$ and define B_0 similarly. Then using the Cauchy-Schwarz inequality and the fact that A_0, B_0 are Hermitian, we have:

$$\begin{aligned}
\mathbb{V}_\psi(A)\cdot\mathbb{V}_\psi(B) &= \langle A_0\psi, A_0\psi\rangle\cdot\langle B_0\psi, B_0\psi\rangle \\
&\geq |\langle A_0\psi, B_0\psi\rangle|^2 \\
&\geq |Im\langle A_0\psi, B_0\psi\rangle|^2 \\
&= \tfrac{1}{4}|\langle A_0\psi, B_0\psi\rangle - \langle B_0\psi, A_0\psi\rangle|^2 \\
&= \tfrac{1}{4}|\langle \psi, A_0B_0\psi\rangle - \langle \psi, B_0A_0\psi\rangle|^2 \\
&= \tfrac{1}{4}|\langle \psi, [A_0, B_0]\psi\rangle|^2 \\
&= \tfrac{1}{4}|\langle \psi, [A, B]\psi\rangle|^2 \\
&= \tfrac{1}{4}|\mathbb{E}_\psi([A,B])|^2.
\end{aligned}$$

$\square$

We discuss next some applications of matrices to classical (Galilean, relativistic) and quantum mechanics.

EXAMPLE 26.32. (Free classical particles in $1+1$ dimensions). The discussion in this Example will be informal. We are interested in showing how matrices can be used to introduce some concepts from classical and relativistic (non-quantum) mechanics.

Let M be a set which we refer to as *space-time*; the elements of M will be referred to as *space-times points*. By a *chart* (or frame) on M we mean a bijection $\Phi : M \to \mathbb{R}^2 = \mathbb{R}^{2\times 1}$. Let G be a subgroup of $SL_2(\mathbb{R})$ identified with a subset of the group of bijections $\mathbb{R}^2 \to \mathbb{R}^2$; for $A \in G$ we identify A with the bijection $v \mapsto Av$. By an *atlas* (or G-atlas) on M we mean a set of charts such that for every two charts Φ, Φ' we have that $\Phi' \circ \Phi^{-1} \in G$. By a *maximal atlas* on M we mean an atlas such that for all chart Φ and all $A \in G$ we have that $A \circ \Phi$ is a chart. For every chart there is a unique maximal atlas containing that chart (prove this!). Fix, in what follows, a maximal atlas. There is a unique point $P_0 \in M$ such that $\Phi(P_0) = 0$ for all charts Φ (prove this!). For a vector $\xi = \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^2$ with $t \neq 0$ we define the *velocity* of v to be the real number

$$\mathrm{vel}(\xi) = \frac{x}{t}.$$

(The interpretation in "physical Argot" of this is the following. We are looking at free particles i.e., particles that are not acted upon by any "force." If a free particle is present at both space-time points P_0 and P then the number $\mathrm{vel}(\Phi(P))$ represents the velocity of the particle measured in the frame/chart Φ. Since "free" means "not subjected to any force," the theory requires that the velocity is constant.) So if

$$A = \Phi' \circ \Phi^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

and $\xi = \Phi(P) = \begin{pmatrix} x \\ t \end{pmatrix}$, $\xi' = \Phi'(P)$ then

$$\xi' = \begin{pmatrix} x' \\ t' \end{pmatrix} = A\xi = \begin{pmatrix} \alpha x + \beta t \\ \gamma x + \delta t \end{pmatrix}$$

hence if $v = \mathrm{vel}(\xi)$ and $v' = \mathrm{vel}(\xi')$ then

$$v' = \frac{\alpha x + \beta t}{\gamma x + \delta t} = \frac{\alpha v + \beta}{\gamma v + \delta} =: A \star v$$

when the latter denominator is non-zero. (The map $v \mapsto A \star v$ is called the *Mobius transformation* defined by A.) A *momentum-energy vector* is a vector $\begin{pmatrix} \pi \\ \epsilon \end{pmatrix}$ of functions $\pi, \epsilon : (-v_0, v_0) \to \mathbb{R}$, where $(-v_0, v_0)$ is an open interval in $\mathbb{R}$ containing 0, such that $\pi(0) = 0$. For $m \in \mathbb{R}_{\geq 0}$ the expression $p = m\pi(v)$ is interpreted in physical Argot as the "momentum' of a "free particle" with "rest mass" m and "velocity" v, the expression $E = m\epsilon(v)$ is interpreted as the "energy" of such a particle, the quantity $U = m\epsilon(0)$ is interpreted as "potential energy," the quantity $T = E - U$ is interpreted as 'kinetic energy,", and (following Galileo) the expression $\frac{p}{v}$ is interpreted as the "inertial mass" of the particle. (However Galileo postulated that the inertial mass coincides with the rest mass, i.e. $\frac{p}{v} = m$, which in relativity theory ceases to be the case). We say that $\begin{pmatrix} \pi \\ \epsilon \end{pmatrix}$ is *conservative* if for all $m_1, m_2 \in \mathbb{R}$, all $v_1, v_2, w_1, w_2 \in \mathbb{R}$, and all $A \in G$ if

$$\sum_{i=1}^{2} m_i \begin{pmatrix} \pi(v_i) \\ \epsilon(v_i) \end{pmatrix} = \sum_{i=1}^{2} m_i \begin{pmatrix} \pi(w_i) \\ \epsilon(w_i) \end{pmatrix}$$

then

$$\sum_{i=1}^{2} m_i \begin{pmatrix} \pi(A \star v_i) \\ \epsilon(A \star v_i) \end{pmatrix} = \sum_{i=1}^{2} m_i \begin{pmatrix} \pi(A \star w_i) \\ \epsilon(A \star w_i) \end{pmatrix}.$$

Here m_i are physically interpreted as the "rest masses" of two particles, v_i are the "velocities of the particles before an elastic collision," and w_i are interpreted as the "velocities of the particles after the elastic collision." So the condition of being conservative says that "if the conservation of the total momentum and energy after an elastic collision holds in one frame then it holds in any other frame." Note that if β is a momentum-energy vector which is conservative then for all matrices $\Lambda \in GL_2(\mathbb{R})$ of the form $\Lambda = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$ and all $\tau = \begin{pmatrix} 0 \\ e \end{pmatrix} \in \mathbb{R}^{2\times 1}$ the vector $\Lambda\beta + \tau$ is a momentum-energy vector that is conservative.

Let us say that a momentum-energy vector $\begin{pmatrix} \pi \\ \epsilon \end{pmatrix}$ is *covariant* if for all $A \in G$ and all $v \in (-v_0, v_0)$ we have that

$$A \begin{pmatrix} \pi(v) \\ \epsilon(v) \end{pmatrix} = \begin{pmatrix} \pi(A \star v) \\ \epsilon(A \star v) \end{pmatrix}.$$

Clearly if a momentum-energy vector is covariant then it is conservative. As we shall see covariance is one of the novelties brought in by relativity theory (in particular it "mixes/bundles" the components of the momentum-energy vector in the same was it "mixes/bundles" the components of the "space-time" vector) and is violated in classical mechanics.

EXERCISE 26.33. (Galilean mechanics). Let $M = \mathbb{R}^2$ and consider the unique maximal atlas containing the chart given by the identity. Let G be the group (called the *Galilean group* in $1+1$ dimensions) of all matrices of the form

$$A = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \quad u \in \mathbb{R}.$$

1) Prove that if $v = \mathrm{vel}(\xi)$, $\xi' = A\xi$, and $v' = \mathrm{vel}(\xi')$. then

$$v' = v + u.$$

(This is interpreted in physical Argot as the "law of addition of velocities.")

2) Prove that for all $\epsilon_0 \in \mathbb{R}$ the vector $\begin{pmatrix} \pi \\ \epsilon \end{pmatrix}$ where $(-v_0, v_0) = \mathbb{R}$, $\pi(v) = v$ and $\epsilon(v) = v^2/2 + \epsilon_0$ is a momentum-energy vector which is conservative but not covariant. (This vector corresponds to the classical momentum and energy for a free particle.)

EXERCISE 26.34. (Relativistic mechanics). Let $M = \mathbb{R}^2$ and consider the unique maximal atlas containing the chart given by the identity. Let G be the group (called the *Lorenz* group in $1+1$ dimensions) of all matrices in $GL_2(\mathbb{R})$ of the form

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}, \quad \alpha^2 - \beta^2 = 1.$$

1) Let $q : \mathbb{R}^{2\times 1} \to \mathbb{R}$ be defined by

$$q(\xi) = \xi^t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \xi, \quad q\begin{pmatrix} x \\ t \end{pmatrix} = x^2 - t^2.$$

Prove that G coincides with the group $O_{1,1}(\mathbb{R})$ of all matrices $A \in SL_2(\mathbb{R})$ such that $q(A\xi) = q(\xi)$ for all ξ.

2) Prove that a matrix $A \in SL_2(\mathbb{R})$ belongs to G if and only if the following conditions hold:

i) If $\mathrm{vel}(\xi) = 1$ then $\mathrm{vel}(A\xi) = 1$. (This is interpreted in physical Argot as saying that "if a particle has velocity 1 in one frame then it has velocity 1 in any other frame"; velocity 1 is interpreted as the "velocity of light" so "the velocity of light is independent of the frame" which is one of the axioms of relativity theory. This is symmetry is a physically natural condition.)

ii) If $\mathrm{vel}(\xi) = 0$, $\mathrm{vel}(A\xi) = u$, $\mathrm{vel}(A\eta) = 0$ then $\mathrm{vel}(\eta) = -u$. (This is interpreted as saying that if a particle that appears at rest with respect to Φ appears to have velocity u with respect to Φ' then every particle that appears to be at rest with respect to Φ' appears to have velocity $-u$ with respect to Φ.)

3) For every $u \in \mathbb{R}$ with $-1 < u < 1$ let

$$\alpha_u := \frac{1}{\sqrt{1-u^2}}, \quad A_u := \alpha_u \cdot \begin{pmatrix} 1 & -u \\ -u & 1 \end{pmatrix}.$$

Prove that for all such u we have that $A_u \in G$ and that for all $A \in G$ there exists u such that $A = A_u$.

4) Prove that for all $-1 < u, v < 1$ we have

$$\alpha_{\frac{v-u}{1-vu}} = \alpha_u \alpha_v (1 - uv).$$

5) Prove that if $\mathrm{vel}(\xi) = v$ with $-1 < v < 1$ then

$$\mathrm{vel}(A_u \xi) = \frac{v-u}{1-vu}.$$

(This is interpreted as the "law of composition of velocities" in relativistic mechanics.)

6) Prove that for $u, v \in (-1, 1)$ we have

$$A_u \begin{pmatrix} \alpha_v v \\ \\ \alpha_v \end{pmatrix} = \begin{pmatrix} \alpha_{\frac{v-u}{1-vu}} \cdot \frac{v-u}{1-vu} \\ \\ \alpha_{\frac{v-u}{1-vu}} \end{pmatrix}.$$

In other words for $\pi(v) = \alpha_v \cdot v$, $\epsilon(v) = \alpha_v$, $(-v_0, v_0) = (-1, 1)$, the momentum-energy vector $\begin{pmatrix} \pi \\ \epsilon \end{pmatrix}$ is covariant, hence conservative.

REMARK 26.35. A consequence of 6) above is that the "inertial mass" $\frac{p}{v} = \frac{m\pi(v)}{v} = m\alpha_v$ is equal to the "energy" $E = m\epsilon(v) = m\alpha_v$:

$$\frac{p}{v} = E.$$

In particular we get $E = m$ for $v = 0$ which is Einstein's celebrated formula. However the above formula fails if one replaces π and ϵ by $\pi^* = a\pi$ and $\epsilon^* = c\pi + d\epsilon + e$, respectively (with a, c, d, e constants) and one replaces E and p by the corresponding E^* and p^*. This is not satisfactory because E^* and p^* are equally plausible candidates for the energy and momentum so it may appear as if the formulae above are an artifact of "convenient/ad hoc" normalizations. Hence one needs arguments supporting the fact that it is reasonable to take $a = d = 1$, $c = e = 0$. This will be dealt with below.

We start with noting that

$$\frac{1}{\sqrt{1-v^2}} = 1 + \frac{v^2}{2} + O(v^4)$$

where $O(v^n)$ stands for any function that can written as v^n times some function which stays bounded when $v \to 0$. Let E and E^* be the corresponding energies and T and T^* the corresponding kinetic energies. We get that

$$\begin{aligned} p &= mv + O(v^3) \\ E &= m + \tfrac{mv^2}{2} + O(v^4) \\ T &= \tfrac{mv^2}{2} + O(v^4). \end{aligned}$$

We deduce that

$$\begin{aligned} p^* &= amv + O(v^3) \\ E^* &= dm + e + cmv + \tfrac{dmv^2}{2} + O(v^3) \\ T^* &= cmv + \tfrac{dmv^2}{2} + O(v^3). \end{aligned}$$

We now impose the condition that p^* and T^* coincide with the Galilean values mv and $\frac{mv^2}{2}$ "up to $O(v^3)$ terms." We deduce that $a = 1$, $d = 1$, $c = 0$. Note at this point the value of e is still undetermined. It is a remarkable fact (which does not hold in Galilean mechanics) that one can pin down the value of e by imposing covariance. Indeed, the vector $\begin{pmatrix} \pi \\ \epsilon + e \end{pmatrix}$ is covariant if and only if $e = 0$.

EXAMPLE 26.36. (Quantum mechanics measurement). The discussion in this Remark will be informal. We are interested in using the language of Hermitian products to introduce some concepts from quantum mechanics. There are two types of behavior for quantum systems:

1) in between two measurements;

2) during a measurement.

The two types of behavior are mathematically (and conceptually) described in completely different ways and this is one of the theoretically unsatisfactory aspects of quantum mechanics. Type 1) is "deterministic" (described by Schrödinger differential equation, cf. Exercise 35.35) and we will ignore here. Type 2) is "probabilistic" and the following is a description of a "baby model" for it; we restrict ourselves to the "finite dimensional case" (which should only be viewed as an approximation of the full theory and, in particular, does not fully accommodate the Heisenberg uncertainty principle, cf. Remark 26.16).

One starts with a finite dimensional vector space V over $\mathbb{C}$ which we identify with $Fun(\Omega, \mathbb{C}) = \mathbb{C}^n = \mathbb{C}^{n\times 1}$, $\Omega := \{1, ..., n\}$. The vectors $\psi \in V$ with $\langle \psi, \psi \rangle = 1$ are called *states* (or state vectors or wave vectors or wave functions on the "discrete space" $\{1, ..., n\}$). We denote by S the set of states. Two states $\psi, \psi' \in S$ are called *equivalent* if $\psi' = z \cdot \psi$ for some $z \in \mathbb{C}$ (which necessarily satisfies $|z| = 1$). Assume we are given a thing called (physical) *system*. The system can be in any of the states in S. One also considers a set $\mathcal{O}$ whose elements are called *observables*. For each observable $A \in \mathcal{O}$ one postulates a measuring device that comes with the following data: a finite subset $\sigma(A) = \{\lambda_1^A, ..., \lambda_n^A\}$ of $\mathbb{R}$ of cardinality n (called the *spectrum* of A) and a collection $L_1^A, ..., L_n^A$ of one dimensional subspaces of V with $\langle L_i^A, L_j^A \rangle = 0$ for $i \neq j$. The device D^A receives inputs and produces outputs. The inputs of D^A can be any of the states $\psi \in S$. Regardless of the input, the outputs of D^A can be any states in the union

$$\bigcup_{i=1}^{n} L_i^A.$$

If the output is in L_i^A we denote it by ψ_i^A. The process of passing from input to output is called the *measurement* of A by D^A. The output ψ_i^A is not empirically knowable and not determined by ψ and A. (This is one of the unsatisfactory aspects of the phenomenology of quantum mechanics.) The value λ_i is not determined by ψ and A but it is known after measurement and it is called the *result* of the measurement of A by our device. One calls $\langle \psi, \psi_i^A \rangle$ the *probability amplitude* of obtaining the output ψ_i^A if the state of the system before the measurement was ψ. (So this complex number is not empirically knowable.) One calls the real number

$$p_i := |\langle \psi, \psi_i^A \rangle|^2$$

the *probability* of obtaining the output ψ_i^A if the state of the system before the measurement was ψ. (This real number can be compared with experimental data.) This latter number is between 0 and 1 in view of 4) in Exercise 26.26 and only depends of the equivalence class of ψ and on L_i^A; also by 5) in Exercise 26.26, the sum of these probabilities as i runs through $\{1,...,n\}$ is 1 so (p_i) is a probability distribution for a probability measure μ_{ψ,Ψ^A} on the set $\Omega := \{1,...,n\}$ (where $\Psi^A = (\psi_1^A,...,\psi_n^A)$). The above definitions/postulates agree with the assumption (without which no theory can take off) that if a measurement of A is performed with result λ_i^A and then a second measurement of A is performed immediately after (so that the Schrödinger equation evolution does not have time to "kick in") then the new result is always, again, λ_i^A: indeed this is compatible with the postulate that $\psi_i^A \in L_i^A$ so the second measurement is certain to yield, again, the value λ_i^A because $|\langle\psi_i^A,\psi_i^A\rangle|^2 = 1$.

Given the real numbers $\{\lambda_1^A,...,\lambda_n^A\}$ and the spaces $\{L_1^A,...,L_n^A\}$ one can consider the unique matrix, for simplicity still denoted by A, such that $A\psi = \lambda_i^A\psi$ for $\psi \in L_i^A$; hence $A\psi_i^A = \lambda_i^A\psi_i^A$ for all i. One can see the matrix A as a mathematical artifact without physical significance: the physically significant objects are, rather, the lines L_i^A and the real values λ_i^A. By 9) in Exercise 26.26 the matrix A is Hermitian and the expectation value $\mathbb{E}_\psi(A) := \langle A\psi,\psi\rangle$ equals the expectation value $\mathbb{E}_{\mu_{\psi,\Psi^A}}(\lambda_A)$, with respect to the probability measure μ_{ψ,Ψ^A} on $\Omega = \{1,...,n\}$, of the random variable $\lambda_A : \Omega \to \mathbb{R}$ given by $\lambda_A(i) := \lambda_i^A$. It corresponds to the average value, of the measurement of A for a system which was in state ψ before the measurement. Recall that we further defined the variance of A with respect to ψ by $\mathbb{V}_\psi(A) := \mathbb{E}_\psi((A - \mathbb{E}_\psi(A)I)^2)$. It is interpreted as the "uncertainty" for the measurement of A when the state of the particle before the measurement is ψ; cf. Remark 26.29. The eigenvalues of $A - \mathbb{E}_\psi(A)I$ with respect to ψ_i^A are $\lambda_i^A - \mathbb{E}_\psi(A)$ so we have that the variance $\mathbb{V}_\psi(A)$ equals the variance, with respect to μ_{ψ,Ψ^A}, of the random variable $\lambda_A : \Omega \to \mathbb{R}$; i.e., $\mathbb{V}_\psi(A) = \mathbb{V}_{\mu_{\psi,\Psi^A}}(\lambda_A)$.

Note that $V = \mathbb{C}^n$ has a distinguished family of one dimensional subspaces, with bases the elements of the standard basis e^i, coming from the identification of V with $Fun(\{1,...,n\},\mathbb{C})$. This family of subspaces, together with a collection $\lambda_1,...,\lambda_n$ of real numbers, are considered to correspond to the observable X called *position*: so $L_i^X = \{ze^i \mid z \in \mathbb{C}\}$ and $\lambda_i^X = \lambda_i$ is the coordinate of a point on the real axis that stands for the position of a particle. (So the particle can occupy only n positions.) If $\psi = (z_1,...,z_n)^t \in \mathbb{C}^n$ then the probability that the particle is at the point with coordinate λ_i is $|z_i|^2$.

A remark on successive measurements. One postulates that the probability amplitude satisfies properties similar to those of probability relative to independent events; we will not review these here. Assume now that A, B, C, D are observables measured for a system, in this order. Assume we know that after the measurement of A the result was λ_i^A. Then, under this assumption, the probability amplitude that after the measurement of D the output is λ_l^D when the input for the measuement of A was ψ_i^A is

$$\sum_{j,k=1}^{n} \langle\psi_i^A,\psi_j^B\rangle\langle\psi_j^B,\psi_k^C\rangle\langle\psi_k^C,\psi_l^D\rangle$$

which is the (i,l)-element of the matrix $S_{AB}S_{BC}S_{CD}$ where

$$S_{AB} = (\langle\psi_i^A,\psi_j^B\rangle)$$

and similarly for S_{BC}, S_{CD}. (Note that S_{AB}, etc. are unitary by Exercise 26.26, 8).) This probability amplitude only depends on

$$\psi_i^A, L_j^B, L_k^C, \psi_l^D$$

(and not on ψ_j^B, ψ_k^C.) The corresponding probability only depends on

$$L_i^A, L_j^B, L_k^C, L_l^D.$$

This probability is usually different from that obtained by measuring these observables in the following order: A, C, B, D. For this new probability amplitude is the (i,l)-element of the matrix $S_{AC}S_{CB}S_{BD}$ and, generally, we have

$$S_{AB}S_{BC}S_{CD} \neq S_{AC}S_{CB}S_{BD}.$$

However if we assume that B and C have the same eigenvectors, i.e., $L_i^B = L_i^C$ then we have

$$S_{AB}S_{BC}S_{CD} = S_{AB}S_{BD} = S_{AC}S_{CD} = S_{AC}S_{CB}S_{BD}.$$

In this case we also have $BC = CB$. Note that, conversely, by Exercise 27.17, if $BC = CB$ then each of $L_1^B, ..., L_n^B$ is one of $L_1^C, ..., L_n^C$.

CHAPTER 27

Determinants

A fundamental concept in the theory of matrices is that of determinant of a matrix. The main results are due to Cauchy, Kronecker, and Weierstrass. In spite of the computational aspect of this concept the best way to approach it is via an "axiomatic" method which is what we shall do. We will add some applications to groups that play a role in classical and quantum mechanics.

DEFINITION 27.1. Let V and W be modules over a ring R and let

$$f : V^n = V \times ... \times V \to W$$

be a map. We say f is *multilinear* (or *R-multilinear*) if and only if for every $v_1, ..., v_n \in V$ and every $i \in \{1, ..., n\}$ we have:

1) If $v_i = v_i' + v_i''$ then

$$f(v_1, ..., v_n) = f(v_1, ..., v_i', ..., v_n) + f(v_1, ..., v_i'', ..., v_n).$$

2) If $v_i = cv_i'$ with $c \in R$ then

$$f(v_1, ..., v_n) = cf(v_1, ..., v_i', ..., v_n).$$

For $n = 2, 3, 4$ we say *bilinear, trilinear, quadrilinear*, respectively.

EXAMPLE 27.2. $f : R^{3\times 1} \times R^{3\times 1} \to R$ defined by

$$f\left(\left(\begin{array}{c} a \\ b \\ c \end{array}\right), \left(\begin{array}{c} d \\ e \\ f \end{array}\right)\right) = ad + 3bf - ce$$

is bilinear.

DEFINITION 27.3. A multilinear map $f : V^n = V \times ... \times V \to W$ is called *alternating* if and only if whenever $v_1, ..., v_n \in V$ and there exist indices $i \neq j$ such that $v_i = v_j$ we have $f(v_1, ..., v_n) = 0$.

EXAMPLE 27.4. f in Example 27.2 is not alternating. On the other hand $g : R^{2\times 1} \times R^{2\times 1} \to R$ defined by

$$f\left(\left(\begin{array}{c} a \\ c \end{array}\right), \left(\begin{array}{c} b \\ d \end{array}\right)\right) = 2ad - 2bc = 2\det\left(\begin{array}{cc} a & b \\ c & d \end{array}\right)$$

is alternating.

LEMMA 27.5. *If $f : V^n \to W$ is multilinear alternating and $v_1, ..., v_n \in V$ then for every indices $i < j$ we have*

$$f(v_1, ..., v_i, ..., v_j, ..., v_n) = -f(v_1, ..., v_j, ..., v_i, ..., v_n).$$

Here $v_1, ..., v_j, ..., v_i, ..., v_n$ is obtained from $v_1, ..., v_i, ..., v_j, ..., v_n$ by replacing v_i with v_j and v_j with v_i while leaving all the other vs unchanged.

Proof. We have

$$\begin{aligned} f(v_1, ..., v_i + v_j, ..., v_i + v_j, ..., v_n) &= f(v_1, ..., v_i, ..., v_i, ..., v_n) \\ &\quad + f(v_1, ..., v_i, ..., v_j, ..., v_n) \\ &\quad + f(v_1, ..., v_j, ..., v_i, ..., v_n) \\ &\quad + f(v_1, ..., v_j, ..., v_j, ..., v_n). \end{aligned}$$

Hence

$$0 = f(v_1, ..., v_i, ..., v_j, ..., v_n) + f(v_1, ..., v_j, ..., v_i, ..., v_n).$$

□

EXERCISE 27.6. Let $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ be a bijection. Then there exists $\epsilon(\sigma) \in \{-1, 1\}$ with the following property. Let $f : V^n \to W$ be any multilinear alternating map and $v_1, ..., v_n \in V$. Then

$$f(v_{\sigma(1)}, ..., v_{\sigma(n)}) = \epsilon(\sigma) \cdot f(v_1, ..., v_n).$$

Hint: Induction on n. For the induction step distinguish two cases: $\sigma(n) = n$ and $\sigma(n) \neq n$. In the first case one concludes directly by the induction hypothesis. The second case can be reduced to the first case via Lemma 27.5.

We identify $(R^{n\times 1})^n$ with $R^{n\times n}$ by identifying a tuple of columns $(b^1, ..., b^n)$ with the $n \times n$ matrix whose columns are $b^1, ..., b^n$. We denote $I = I_n$ the identity $n \times n$ matrix.

LEMMA 27.7. *There exists a multilinear alternating map*

$$f : R^{n\times n} \to R$$

such that $f(I) = 1$.

Proof. We proceed by induction on n. For $n = 1$ we take $f(a) = a$. Assume we constructed a multilinear alternating map

$$f_{n-1} : R^{(n-1)\times(n-1)} \to R$$

such that $f_{n-1}(I_{n-1}) = 1$. Let $A = (a_{ij})$ be an $n \times n$ matrix and let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i-th row and the j-th column. Fix i and define

$$f_n(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} f_{n-1}(A_{ij}).$$

One easily checks that f_n is multilinear, alternating, and takes value 1 on the identity matrix I_n. □

EXERCISE 27.8. Check the last sentence in the proof above.

LEMMA 27.9. *If f and g are multilinear alternating maps $R^{n\times n} \to R$ and $f(I) = 1$ then there exists $c \in R$ such that $g(A) = cf(A)$ for all A.*

Proof. Let $A = (a_{ij})$. Let $e^1, ..., e^n$ be the standard basis of $R^{n\times 1}$. Then

$$g(A) = g\left(\sum_{i_1} a_{i_1 1}e^{i_1}, ..., \sum_{i_n} a_{i_n n}e^{i_n}\right) = \sum_{i_1} ... \sum_{i_n} a_{i_1 1}...a_{i_n n} g(e^{i_1}, ..., e^{i_n}).$$

The terms for which $i_1, ..., i_n$ are not distinct are zero. The terms for which $i_1, ..., i_n$ are distinct are indexed by permutations σ. By Exercise 27.6 we get

$$g(A) = \left(\sum_{\sigma} \epsilon(\sigma) a_{\sigma(1)1}...a_{\sigma(n)n}\right) g(I).$$

A similar formula holds for $f(A)$ and the Lemma follows. $\square$

By Lemmas 27.7 and 27.9 we get:

THEOREM 27.10. *There exists a unique multilinear alternating map (called determinant)*

$$\det : R^{n\times n} \to R$$

such that $\det(I) = 1$.

EXERCISE 27.11. Using the notation in the proof of Lemma 27.7 prove that:

1) For all i we have

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

2) For all j we have

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

Hint: Use Lemma 27.9.

We also have:

THEOREM 27.12. *For every two matrices $A, B \in R^{n\times n}$ we have*

$$\det(AB) = \det(A)\det(B).$$

Proof. Consider the multilinear alternating map $f : R^{n\times n} \to R$ defined by

$$f(u^1, ..., u^n) = \det(Au^1, ..., Au^n)$$

for $u^1, ..., u^n \in R^{n\times 1}$. By Lemma 27.9 there exists $c \in R$ such that

$$f(u^1, ..., u^n) = c \cdot \det(u^1, ..., u^n).$$

Hence

$$\det(Au^1, ..., Au^n) = c \cdot \det(u^1, ..., u^n).$$

Setting $u^i = e^i$ we get $\det(A) = c \cdot \det(I) = c$. Setting $u^i = b^i$, the columns of B, we get $\det(AB) = c \cdot \det(B)$ and the theorem is proved. $\square$

EXERCISE 27.13. Prove that $\epsilon(\sigma\tau) = \epsilon(\sigma)\epsilon(\tau)$ for every permutations $\sigma, \tau \in S_n$; in other words $\epsilon : S_n \to \{1, -1\}$ is a group homomorphism.

Hint: For $e^1, ..., e^n$ the standard basis and $\sigma \in S_n$ let $F_\sigma : R^n \to R^n$ be the unique linear map such that $F(e^i) = e^{\sigma(i)}$ for all i and let A_σ be the matrix of F_σ with respect to the standard basis. We have $F_\sigma(F_\tau(e^i)) = F_\sigma(e^{\tau(i)}) = e^{\sigma(\tau(i))}$ so $F_\sigma \circ F_\tau = F_{\sigma\circ\tau}$ so $A_\sigma A_\tau = A_{\sigma\circ\tau}$ so $\det(A_\sigma)\det(A_\tau) = \det(A_{\sigma\circ\tau})$. On the other hand $\det(A_\sigma) = \epsilon(\sigma)$.

Recall that $R^\times$ denotes the set of invertible elements in a ring R.

Exercise 27.14. Prove that if $A \in R^{n\times n}$ is a matrix such that $\det(A) \in R^\times$ then A is invertible i.e., there exists $B \in R^{n\times n}$ such that $AB = BA = I$.

Hint: Define $B = (b_{ij})$ where

$$b_{ij} = (-1)^{i+j}\det(A_{ji})$$

(notation as in Lemma 27.7). Prove that $AB = BA = I$ using Exercise 27.11.

Exercise 27.15. Prove that if $A \in R^{n\times n}$ is a matrix then $\det(A) = \det(A^t)$.

Exercise 27.16. Let $A \in K^{n\times n}$ where K is a field. Prove that $\lambda \in K$ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$. In particular A and A^t have the same eigenvalues.

Exercise 27.17. Let V be a vector space over a field K. Let $F, G : V \to V$ be linear maps with $F \circ G = G \circ F$, let $\lambda \in K$ be an eigenvalue of A, and let V_λ be the eigenspace of F corresponding to λ. Prove that $G(V_\lambda) \subset V_\lambda$.

Remark 27.18. It follows from Exercises 27.16 and 27.17 plus the Fundamental Theorem of Algebra 28.14 to be stated (without proof) later that if $K = \mathbb{C}$ and $F, G : V \to V$ are linear maps with $F \circ G = G \circ F$ then F and G possess a common eigenvector. (Check!)

Exercise 27.19. Let R be a ring.

1) Prove that the set

$$GL_n(R) = \{A \in R^{n\times n} \mid \det(A) \in R^\times\}$$

is a group with respect to multiplication; $GL_n(R)$ is called the *general linear group*.

2) Prove that the set

$$SL_n(R) = \{A \in R^{n\times n} \mid \det(A) = 1\}$$

is a subgroup of $GL_n(R)$; $SL_n(R)$ is called the *special linear group*.

3) Prove that the sets

$$\begin{aligned} O_n(R) &= \{A \in R^{n\times n} \mid AA^t = I\}, \\ SO_n(R) &= \{A \in R^{n\times n} \mid \det(A) = 1,\ AA^t = I\} \end{aligned}$$

are subgroups of $GL_n(R)$; they are referred to as the *orthogonal group* and the *special orthogonal group*, respectively.

Check that for $n = 2$ the above correspond to the previously defined groups $GL_2(R), SL_2(R), SO_2(R)$.

Exercise 27.20.

1) Prove that if a linear map $F : V \to W$ between two modules is bijective then its inverse $F^{-1} : W \to V$ is also linear. Such a map will be called an *isomorphism* (of R-modules, or R-linear).

2) Prove that the set $GL(V)$ of all isomorphisms $V \to V$ is a group under composition.

3) Assume V has a basis $v_1, ..., v_n$ and consider the map $GL(V) \to GL_n(R)$, $F \mapsto A_F$ where A_F is the matrix of F with respect to $v_1, ..., v_n$. Prove that $GL(V) \to GL_n(R)$ is an isomorphism of groups.

EXERCISE 27.21. Define the function $|| \;||^2 : \mathbb{R}^{n\times 1} \to \mathbb{R}_{\geq 0}$ by sending every column vector $x = (x_1, ..., x_n)^t \in \mathbb{R}^{n\times 1}$ into

$$||x||^2 = x^t x = x_1^2 + ... + x_n^2 \in \mathbb{R}_{\geq 0}.$$

Prove that

$$SO_n(\mathbb{R}) = \{A \in SL_n(\mathbb{R}) \mid \text{for all } x \in \mathbb{R}^{n\times n},\ ||Ax||^2 = ||x||^2\}.$$

EXERCISE 27.22.

1) Prove that the sets

$$\begin{aligned} U_n(\mathbb{C}) &= \{A \in GL_n(\mathbb{C}) \mid A^*A = I\}, \\ SU_n(\mathbb{C}) &= \{A \in GL_n(\mathbb{C}) \mid \det(A) = 1, \quad A^*A = I\} \end{aligned}$$

are subgroups of $GL_n(\mathbb{C})$; they are called the *unitary group* and *special unitary group*, respectively.

2) Prove that

$$SU_2(\mathbb{C}) = \{\begin{pmatrix} z & w \\ \overline{w} & \overline{z} \end{pmatrix} \mid |z|^2 + |w|^2 = 1\}.$$

REMARK 27.23. By 2), if one sets $z = x_1 + ix_2$ and $w = x_3 + ix_4$ with x_i real then $SU_2(\mathbb{C})$ identifies with the "unit 3-sphere"

$$S^3 := \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}.$$

The latter is the union

$$S^3 = S^3_+ \cup S^3_-$$

where

$$S^3_+ := \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1,\ x_4 \geq 0\},$$

$$S^3_- := \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1,\ x_4 \leq 0\}.$$

Furthermore

$$S^3_+ \cap S^3_- = S^2 := \{(x_1, x_2, x_3) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 = 1\}$$

and both S^3_+ and S^3_- are in natural bijections with the "3-ball"

$$B^3 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 \leq 1\}$$

via the projections

$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2, x_3).$$

So S^3 can be obtained from two copies of B^3 "glued along their boundaries S^2." The geometry of this gluing is described in Dante's Divine Comedy as follows. The first ball B^3 has the Earth in its center and contains the spheres of the Moon, Sun, and the planets. The center of the Earth is the bottom of Hell where Lucifer is situated. Dante and Virgil descend through Hell; when they cross the center of the Earth gravitation changes its direction. Then Dante and Virgil reach the surface of the Earth at the top of the Purgatory mountain where Beatrice takes Virgil's place. Dante and Beatrice ascend through the celestial spheres in the first copy of B^3. When they approach the boundary S^2 of the first copy of B^3 the vapors falling towards the Earth (likened to falling snow flakes) stop their movement for a moment and then, as Dante and Beatrice enter the second copy of B^3, the vapors start falling in the opposite direction towards the center of the second copy of B^3:

a second change in the direction of gravitation! This center of the second B^3 is a point of blinding light where God is situated.

EXERCISE 27.24. Define the vector space (called the *Lie algebra of the special unitary group*):

$$su_2(\mathbb{C}) = \{M \in \mathbb{C}^{2\times 2} \mid M + M^* = 0,\ Tr(M) = 0\}.$$

1) Prove that

$$su_2(\mathbb{C}) = \{\begin{pmatrix} ix_1 & x_2 + ix_3 \\ -x_2 + ix_3 & -ix_1 \end{pmatrix} \mid x_1, x_2, x_3 \in \mathbb{R}\}.$$

2) Let $su_2(\mathbb{C}) \to \mathbb{R}^{3\times 1}$ be the isomorphism of vector spaces sending every

$$X = \begin{pmatrix} ix_1 & x_2 + ix_3 \\ -x_2 + ix_3 & -ix_1 \end{pmatrix} \in su_2(\mathbb{C})$$

into the column vector $\widehat{X} = (x_1, x_2, x_3)^t \in \mathbb{R}^{3\times 1}$. Prove that

$$\det(X) = ||\widehat{X}||^2.$$

3) Show that there is a group homomorphism

$$\pi : SU_2(\mathbb{C}) \to SO_3(\mathbb{R})$$

such that $f^{-1}(I) = \{\pm I\}$.

Hint: For $A \in SU_2(\mathbb{C})$ we let $\pi(A)$ be the 3×3 matrix $\widehat{A}$ defined by

$$\widehat{A}\widehat{X} = \widehat{AXA^*}$$

for $X \in su_2(\mathbb{C})$. Then we have

$$||\widehat{A}\widehat{X}||^2 = ||\widehat{AXA^*}||^2 = \det(AXA^*) = \det(A)\det(X)\det(A^*)$$
$$= \det(AA^*)\det(X) = \det(X) = ||\widehat{X}||^2$$

which proves that $\widehat{A}\widehat{A}^* = I$. In particular $\widehat{A}$ has determinant ± 1; to check the determinant is 1 one needs a more explicit analysis of this construction (or one can use a topological argument).

REMARK 27.25. The map π above can be shown to also be surjective but this is harder. The group $SU_2(\mathbb{C})$ plays a role in the non-relativistic quantum mechanics (spin of the electron) while $SO_3(\mathbb{R})$ plays a role in the Newtonian mechanics; the map π relates them.

EXERCISE 27.26. Define

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Define the *Lorentz group*

$$O_{1,3}(\mathbb{C}) = \{A \in GL_4(\mathbb{R}) \mid A^tQA = Q\}$$

and consider the map $q : \mathbb{R}^{4\times 1} \to \mathbb{R}$ defined by

$$q(x) = x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

1) Prove that $O_{1,3}(\mathbb{R})$ is a subgroup of $GL_4(\mathbb{R})$.
2) Prove that $O_{1,3}(\mathbb{R}) = \{A \in GL_4(\mathbb{R}) \mid \text{for all } x \in \mathbb{R}^{4\times 1},\ \ q(Ax) = q(x)\}$.

EXERCISE 27.27. Define the space of *Hermitian matrices*,

$$h_4(\mathbb{C}) = \{M \in \mathbb{C}^{4\times 4} \mid M^* = M\}.$$

1) Prove that

$$h_4(\mathbb{C}) = \{\begin{pmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{pmatrix} \mid x_0, x_1, x_2, x_3 \in \mathbb{R}\}.$$

2) Let $h_4(\mathbb{C}) \to \mathbb{R}^{4\times 1}$ be the isomorphism of vector spaces sending every

$$X = \begin{pmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{pmatrix} \in h_4(\mathbb{C})$$

into the column vector $\widehat{X} = (x_0, x_1, x_2, x_3)^t \in \mathbb{R}^{4\times 1}$. Prove that

$$\det(X) = q(\widehat{X}).$$

3) Prove that there is a non-trivial group homomorphism

$$\pi : SL_2(\mathbb{C}) \to O_{1,3}(\mathbb{R}).$$

Hint: Let π attach to every $A \in SL_2(\mathbb{C})$ the 4×4 matrix $\widehat{A}$ defined by $\widehat{A}\widehat{X} = \widehat{AXA^*}$ for $X \in h_4(\mathbb{C})$.

REMARK 27.28. The Lorentz group $O_{1,3}(\mathbb{R})$ plays a key role in Relativity Theory and the group $SL_2(\mathbb{C})$ plays a role in relativistic quantum mechanics; the map π above relates them. The space $\mathbb{R}^{4\times 1}$ is called the *Minkovski space-time* and the map $(x, y) \mapsto q(x - y)$ is called the *interval* between two space-time points $x, y \in \mathbb{R}^{4\times 1}$.

CHAPTER 28

Polynomials

Determining the roots of polynomials was one of the most important motivating problems in the development of algebra, especially in the work of Cardano, Lagrange, Gauss, Abel, and Galois. Here we introduce polynomials and discuss some basic facts about their roots.

DEFINITION 28.1. Let R be a ring. We define the *ring of polynomials* $R[x]$ in one variable with coefficients in R as follows. An element of $R[x]$ is a map $f : \mathbb{N} \cup \{0\} \to R$, $i \mapsto a_i$ with the property that there exists $i_0 \in \mathbb{N}$ such that for all $i \geq i_0$ we have $a_i = 0$; we also write such a map as

$$f = (a_0, a_1, a_2, a_3, ...).$$

We define $0, 1 \in R[x]$ by

$$0 = (0, 0, 0, 0, ...),$$

$$1 = (1, 0, 0, 0, ...).$$

If f is as above and $g = (b_0, b_1, b_2, b_3, ...)$ then addition and multiplication are defined by

$$\begin{array}{rcl} f + g & = & (a_0 + b_0, a_1 + b_1, a_2 + b_2, a_3 + b_3, ...), \\ fg & = & (a_0 b_0, a_0 b_1 + a_1 b_0, a_0 b_2 + a_1 b_1 + a_2 b_0, a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0, ...). \end{array}$$

We define the degree of $f = (a_0, a_1, a_2, a_3, ...)$ as

$$deg(f) = \min\{i \mid a_i \neq 0\}$$

if $f \neq 0$ and $deg(0) = 0$. We also define

$$x = (0, 1, 0, 0, ...)$$

and we write

$$a = (a, 0, 0, 0, ...)$$

for every $a \in R$.

EXERCISE 28.2.
1) Prove that $R[x]$ with the operations above is a ring.
2) Prove that the map $R \to R[x]$, $a \mapsto a = (a, 0, 0, 0, ...)$ is a ring homomorphism.
3) Prove that $x^2 = (0, 0, 1, 0, ...)$, $x^3 = (0, 0, 0, 1, 0, ...)$, etc.
4) Prove that if $f = (a_0, a_1, a_2, a_3, ...)$ then

$$f = a_n x^n + ... + a_1 x + a_0$$

where $n = deg(f)$. (We also write $f = f(x)$ but we DO NOT SEE $f(x)$ as a function; this is just a notation.)

EXAMPLE 28.3. If $R = \mathbb{Z}$ then

$$(3x^2 + 5x + 1)(8x^3 + 7x^2 - 2x - 1) =$$

$$= (3 \times 8)x^5 + (3 \times 7 + 5 \times 8)x^4 + (3 \times (-2) + 5 \times 7 + 1 \times 8)x^3 + ...$$

DEFINITION 28.4. For every $b \in R$ we define an element $f(b) \in R$ by

$$f(b) = a_n b^n + ... + a_1 b + a_0.$$

So for every polynomial $f \in R[x]$ we can define a map (called the *polynomial map* defined by the polynomial f):

$$R \to R, b \mapsto f(b).$$

(The polynomial map defined by f should not be confused with the polynomial f itself; they are two different entities.) An element $b \in R$ is called a *root* of f (or a zero of f) if $f(b) = 0$. (Sometimes we say "a root in R" instead of "a root.")

EXAMPLE 28.5. If $R = \mathbb{F}_2$ and we consider the polynomial $f(x) = x^2 + x \in R[x]$ then $f = (\overline{0}, \overline{1}, \overline{1}, \overline{0}, ...) \neq (\overline{0}, \overline{0}, \overline{0}, ...) = 0$ as an element of $R[x]$; but the polynomial map $R \to R$ defined by f sends $\overline{1} \mapsto \overline{1}^2 + \overline{1} = \overline{0}$ and $\overline{0} \mapsto \overline{0}^2 + \overline{0} = \overline{0}$ so this map is the constant map with value $\overline{0}$. This shows that different polynomials (in our case f and 0) can define the same polynomial map.

EXERCISE 28.6. Let $R = \mathbb{R}$. Show that
1) $x^2 + 1$ has no root in $\mathbb{R}$.
2) $\sqrt{\sqrt{3} + 1}$ is a root of $x^4 - 2x^2 - 1 = 0$.

REMARK 28.7. One can ask if every root in $\mathbb{C}$ of a polynomial with coefficients in $\mathbb{Q}$ can be expressed, using (possibly iterated) radicals of rational numbers. The answer to this is negative as shown by Galois in the early 19th century.

EXERCISE 28.8. Let $R = \mathbb{C}$. Show that
1) i is a root of $x^2 + 1$ in $\mathbb{C}$.
2) $\frac{1+i}{\sqrt{2}}$ is a root of $x^4 + 1 = 0$ in $\mathbb{C}$.

REMARK 28.9. Leibniz mistakenly thought that the polynomial $x^4 + 1$ should have no root in $\mathbb{C}$.

EXERCISE 28.10. Let $R = \mathbb{F}_7$. Show that:
1) $x^2 + \overline{1}$ has no root in R.
2) $\overline{2}$ is a root of $x^3 - x + \overline{1}$ in R.

EXERCISE 28.11. Let $R = \mathbb{F}_5$. Show that:
1) $\overline{2}$ is a root of $x^2 + \overline{1}$ in R.
2) $x^5 - x + 1$ has no root in $\mathbb{F}_5$.

DEFINITION 28.12. A field C is called *algebraically closed* if every polynomial in $C[x] \setminus C$ has a root in C.

The study of roots of polynomial functions is one of the main concerns of algebra. Here are two of the main basic theorems about roots.

THEOREM 28.13. *(Lagrange) If K is a field then every polynomial of degree $d \geq 1$ has at most d roots in K.*

THEOREM 28.14. *(Fundamental Theorem of Algebra, d'Alembert) The field $\mathbb{C}$ of complex numbers is algebraically closed.*

In what follows we prove Theorem 28.13. (Theorem 28.14 will be proved in Exercise 34.10 and 36.26.) We need a preparation.

DEFINITION 28.15. A polynomial $f(x) = a_n x^n + ... + a_1 x + a_0$ of degree n is *monic* if $a_n = 1$.

PROPOSITION 28.16. *(Long division). Let R be a ring. Let $f(x), g(x) \in R[x]$ with $g(x)$ monic of degree ≥ 1. Then there exist unique $q(x), r(x) \in R[x]$ such that*

$$f(x) = g(x)q(x) + r(x)$$

and $deg(r) < deg(g)$.

Proof. Fix g (of degree m) and let us prove by induction on n that the statement above is true if $deg(f) \leq n$. The case $deg(f) = 0$ is clear because we can then take $q(x) = 0$ and $r(x) = f(x)$. For the induction step we may take f of degree n and let $f(x) = a_n x^n + ... + a_0$, $a_n \neq 0$. We may assume $n \geq m$. Then

$$deg(f - a_n x^{n-m} g) \leq n - 1$$

so by the induction hypothesis

$$f(x) - a_n x^{n-m} g(x) = g(x)q(x) + r(x)$$

with $deg(r) < m$. So

$$f(x) = g(x)(a_n x^{n-m} + q(x)) + r(x)$$

and we are done. $\square$

Proof of Theorem 28.13. Assume there exists a polynomial $f \in K[x]$ of degree $d \geq 1$ that has $d+1$ roots. Choose f such that d is minimal and seek a contradiction. Let $a_1, ..., a_{d+1} \in K$ be distinct roots of f. By Long Division we can write

$$f(x) = (x - a_{d+1})g(x) + r(x)$$

with $deg(r) < deg(x - a_{d+1}) = 1$. So $deg(r) = 0$ i.e., $r(x) = c \in K$. Since $f(a_{d+1}) = 0$ we get $r(x) = 0$ hence $c = 0$. Since $0 = f(a_k) = (a_k - a_{d+1})g(a_k) + c$ for $k = 1, ..., d$ it follows that $0 = (a_k - a_{d+1})g(a_k)$. Since K is a field and $a_k - a_{d+1} \neq 0$ for $k = 1, ..., d$ it follows that $g(a_k) = 0$ for $k = 1, ..., d$. But $deg(g) = d - 1$ which contradicts the minimality of d. $\square$

Long division for polynomials can be used in the same way Euclidean division in $\mathbb{Z}$ was used to prove an analogue of the Fundamental Theorem of Arithmetic as follows.

DEFINITION 28.17. Let K be a field. For polynomials f and g in $K[x]$ we say f *divides* g if there exists $h \in K[x]$ such that $g = fh$. We write $f|g$. We also say f is a *divisor* of g; or that g is a *multiple* of f. If f does not divide g we write $f \not| g$.

DEFINITION 28.18. For $f \in K[x]$ denote by $fK[x]$ (sometimes by (f)) the set $\{fh \mid h \in K[x]\}$ of polynomials divisible by f. (It is an ideal.) For $f, g \in K[x]$ denote by $fK[x] + gK[x]$ (sometimes by (f, g)) the set $\{hf + kg \mid h, k \in K[x]\}$ of all polynomials expressible as a multiple of f plus a multiple of g. (This is again an ideal.)

EXERCISE 28.19.

1) For every polynomials f and g there exists a polynomial h such that

$$fK[x] + gK[x] = hK[x].$$

More generally every ideal in $K[x]$ has the form $hK[x]$ for some $h \in K[x]$.

2) If f and g are polynomials and have no common divisor of degree ≥ 1 then there exist polynomials h and k such that $hf + kg = 1$.

3) A polynomial p of degree d is called *irreducible* if and only if $d \geq 1$ and p has no divisors of degree e with $1 \leq e \leq d-1$. Prove that if p is an irreducible polynomial and $p|fg$ for polynomials f and g then either $p|f$ or $p|g$. (Hence an ideal in $K[x]$ is prime if and only if it is zero or of the form $pK[x]$ with p irreducible.)

4) Any polynomial f of degree > 1 can be written uniquely as a product of irreducible polynomials, i.e., there exist irreducible polynomials $p_1, p_2, ..., p_s$, where $s \geq 1$, such that

$$f = p_1 p_2 ... p_s.$$

Moreover any such representation is unique in the following sense: if

$$p_1 p_2 ... p_s = q_1 q_2 ... q_r$$

with p_i and q_j irreducible then $s = r$ and there exists a permutation $\sigma \in S_r$ and elements $c_1, ..., c_r \in K$ such that $p_i = c_i q_{\sigma(i)}$ for all $i \in \{1, ..., r\}$.

Hint: Imitate the case of the integers.

REMARK 28.20. The main problems about roots are:

1) Find the number of roots; in case $K = \mathbb{F}_p$ this leads to some of the most subtle problems in number theory.

2) Understand when roots of polynomials with rational coefficients, say, can be expressed by radicals; this leads to Galois theory.

The following concept plays an important role in the study of polynomials.

DEFINITION 28.21. By a *field extension* we mean an inclusion $K \subset L$ such that L is a field and K is a subfield of L. Note that if this is the case L has a naturally induced structure of vector space over K. If L has finite dimension as a vector space over K we denote this dimension by $[L : K]$ and we call it the *degree* of the extension.

EXERCISE 28.22. Prove that if $K \subset L$ and $L \subset F$ are finite degree field extensions then so is $K \subset F$ and

$$[F : K] = [F : L] \cdot [L : K].$$

Hint: If $\alpha_1, ..., \alpha_n$ is a basis of L as a vector space over K and if $\beta_1, ..., \beta_m$ is a basis of F as a vector space over L prove that $\alpha_1\beta_1, \alpha_1\beta_2, ..., \alpha_n\beta_m$ is a basis of F as a vector space over K.

DEFINITION 28.23. Let $K \subset L$ be a field extension. An element $\alpha \in L$ is called *algebraic* over K if there exists a polynomial $f \in K[x]$ of degree ≥ 1 such that α is a root of f: $f(\alpha) = 0$. We say α is *transcendental* over K if it is not algebraic.

EXERCISE 28.24. Let $K \subset L$ be a field extension and $\alpha \in L$ algebraic over K. Prove that there is a unique monic irreducible polynomial $f \in K[x]$ such that $f(\alpha) = 0$. It is called the *irreducible polynomial* of α over K and is denoted by

$\varphi_{\alpha/K}$. It has the property that it divides all the polynomials in $K[x]$ that vanish at α.

Hint: Let f be a monic polynomial of smallest degree in the set $P_{\alpha/K}$ of all polynomials in $K[x]$ which vanish at α. Then f is irreducible because if $f = gh$ then either g or h vanish at α. Also f divides all the polynomials g in $P_{\alpha/K}$ (in other words $P_{\alpha/K} = fK[x]$) because the remainder when g is divided by f also vanishes at α so the remainder must be zero. Uniqueness of f follows from the latter property.

Exercise 28.25. Let $K \subset L$ be a field extension and $\alpha \in L$ algebraic over K. Let d be the degree of $\varphi_{\alpha/K}$ and let $K(\alpha)$ be the set of all elements of L of the form $g(\alpha)$ with $g \in K[x]$.

1) Prove that $K(\alpha)$ is a field.

2) Prove that $1, \alpha, \alpha^2, ..., \alpha^{d-1}$ is a basis for the K-linear space $K(\alpha)$; so

$$[K(\alpha) : K] = d.$$

3) Prove that if $\beta \in L$ is another algebraic element over K then $K(\alpha)(\beta) = K(\beta)(\alpha)$; we denote this field by $K(\alpha, \beta)$.

Hint for 1): Clearly $K(\alpha)$ is a ring. Now for $g \in K[x]$ with $g(\alpha) \neq 0$ we have that g is not divisible by $f = \varphi_{\alpha/K}$. Hence $fK[x] + gK[x] = hK[x]$ with $h \in K^{\times}$ hence $kf + mg = 1$ for some $k, m \in K[x]$ hence evaluating the latter at α we get $m(\alpha)g(\alpha) = 1$.

Hint for 2): Linear independence follows from the fact that any linear combination of the elements $1, \alpha, \alpha^2, ..., \alpha^{d-1}$ which vanishes gives a polynomial of degree $\leq d$ that vanishes at α. On the other hand for all g we have $g = fq + r$ with $\deg(r) \leq d - 1$ so evaluating at α we get $g(\alpha) = r(\alpha)$.

Definition 28.26. A complex number $\alpha \in \mathbb{C}$ is *constructible* is there exist field extensions

$$\mathbb{Q} \subset K_1 \subset K_2 \subset ... \subset K_n$$

such that $[K_{i+1} : K_i] = 2$ for all i and $\alpha \in K_n$. (The terminology is justified by the fact that the above holds if and only if a segment of length α can be constructed using the ruler and compass only, starting with a segment of length 1. Check this assuming analytic geometry for which we refer to the Chapters on Lines and Conics!)

Exercise 28.27. Prove that if $\alpha \in \mathbb{R}$, $\alpha^3 = 2$, then α is not constructible. (This is a classical problem recorded by Plato which was only solved in the 19th century.)

Hint: One needs to show that $x^3 - 2$ is irreducible. If it were reducible it would have one factor of degree 1 hence a root in $\mathbb{Q}$. Prove this is impossible by contradiction.

Definition 28.28. (Derivative). Let $f \in R[x]$ be a polynomial with coefficients in a ring, $f(x) = \sum_{n=0}^{d} a_n x^n$. Its *derivative* is the polynomial $f'(x) = \sum_{n=0}^{d} n a_n x^{n-1}$.

Exercise 28.29. Prove that for polynomials $f, g \in R[x]$ we have

1) $(f + g)' = f' + g'$

2) $(fg)' = f'g + fg'$. (Leibniz rule)

Exercise 28.30. Let $K \subset L$ be a field extension, $\alpha \in L$, and $f \in K[x]$.

1) (Descartes) Prove that $f(\alpha) = 0$ if and only if $(x-\alpha)|f$ in $K[x]$.

2) Prove that $f(\alpha) = 0$ and $f'(\alpha) = 0$ if and only if $(x-\alpha)^2|f$ in $K[x]$. (If this is the case we say α is a multiple root of f; if this is not the case we say that α is a simple root of f.)

3) If $char(K) = 0$, f is irreducible in $K[x]$, and $f(\alpha) = 0$. Then α is a simple root.

Hint: 1) follows directly from dividing f by $x - \alpha$ with remainder. For the if part of 2) apply the Leibniz rule to $f = (x-\alpha)^2 h$. For the only if part if $f = (x-\alpha)g$ with $(x-\alpha) \not| g$ then by part 1) $g(\alpha) \neq 0$ so we get $f' = g + (x-\alpha)g'$ hence $f'(\alpha) \neq 0$. For 3) assume f is irreducible and α is a multiple root. We may assume f is monic hence $f = \varphi_{\alpha/K}$. Since $f'(\alpha) = 0$ it follows that $f|f'$. Since $\mathbb{Q} \subset K$ we have that f' has degree one less that that of f so it cannot be divisible by f.

Definition 28.31. Let $K \subset L$ be a field extension. We say that a polynomial $f \in K[x]$ *splits completely* in L if it is a product of polynomials of degree 1 (*linear polynomials*) with coefficients in L.

Exercise 28.32. Prove that if C is an algebraically closed field then every polynomial in $C[x]$ splits completely in $C[x]$.

Theorem 28.33. *(Theorem of the Primitive Element). Let $K \subset L$ be a field extension and $\alpha, \beta \in L$ algebraic over K. Assume that α and β are roots of two polynomials f and g in $K[x]$, respectively, such that f and g split completely in $L[x]$ and have only simple roots in L. Then there exists $c \in K$ such that*

$$K(\alpha, \beta) = K(\alpha + c\beta).$$

Proof (in case K is infinite). Let $\alpha_1, ..., \alpha_n$ and $\beta_1, ..., \beta_m$ be the roots of f and g in L, respectively, with $\alpha = \alpha_1$, $\beta = \beta_1$ and with α_i distinct and β_j distinct. Then there exists $c \in K$ such that $\alpha_i + c\beta_j \neq \alpha + c\beta$ for all $i, j \geq 2$; this is because the quotients $\frac{\alpha_i - \alpha}{\beta_j - \beta}$ form a finite set and K is infinite. Then the polynomials $g(x)$ and $f(\alpha + c\beta - cx)$ have only β as a common root in L. Write

$$g(x)K(\alpha + c\beta)[x] + f(\alpha + c\beta - cx)K(\alpha + c\beta)[x] = h(x)K(\alpha + c\beta)[x],$$

with $h \in K(\alpha + c\beta)[x]$. Then β is the only root of h in L. Also since h divides g in $L[x]$ it follows that h splits completely in $L[x]$ and has only simple roots in L. Hence we may assume $h(x) = x - \beta$. This implies $\beta \in K(\alpha + c\beta)$ which in its turn implies $\alpha \in K(\alpha + c\beta)$. □

Exercise 28.34. (Construction of fields). Let $f \in K[x]$ be an irreducible polynomial. Define an equivalence relation $\sim$ in $K[x]$ by letting $g \sim h$ if and only if $g - h \in (f) := fK[x]$; we write

$$g \equiv h \mod (f)$$

Denote by $K[x]/(f)$ the quotient $K[x]/\sim$. Define on $K[x]/(f)$ an addition and multiplication by the rules

$$\widehat{g} + \widehat{h} := \widehat{g+h}, \quad \widehat{g} \cdot \widehat{h} := \widehat{gh}.$$

1) Prove that $L := K[x]/(f)$ is a field and $\widehat{x}$ is a root in L.

2) Prove that the field extension $K \subset L$ has degree equal to $\deg(f)$.

3) Prove that if $K \subset F$ is a field extension and $\alpha \in F$ is algebraic over K with irreducible polynomial $f \in K[x]$ then $L := K[x]/(f)$ is isomorphic to $K(\alpha)$ by an isomorphism that is the identity on K.

4) Prove that every polynomial in $K[x] \setminus K$ has a root in some field extension of K.

5) Prove that for every finite set S of polynomials in $K[x] \setminus K$ there is a field extension E of K such that every polynomial in S has a root in E.

6) Prove that there is a field extension E of K such that every polynomial in $K[x] \setminus K$ has a root in E.

7) Prove that there is a field extension C of K such that every polynomial in $C[x] \setminus C$ has a root in C (i.e., C is algebraically closed).

8) Prove that there is a field extension K^a of K such that every polynomial in $K^a[x] \setminus K^a$ has a root in K^a and such that every element of K^a is algebraic over K. Such a K^a is called an *algebraic closure* of K.

9) Prove that for every finite extension $K \subset K'$ there exists a field homomorphism $\sigma : K' \to K^a$ which is the identity on K.

Hint: For 4) take an irreducible factor of the polynomial and use 1). For 5) apply induction. For 6) attach to every polynomial $f \in K[x] \setminus K$ a variable x_f and consider the union A of all rings of polynomials in finitely many such variables. Let I be the ideal of all (finite) sums $\sum_f a_f f(x_f)$ with $a_f \in A$. Prove that $I \neq A$. Take a maximal ideal M of A containing I. Then take $E := A/M$. For 7) take $C = \bigcup_n E_n$ where $K = E_1 \subset E_2 \subset E_3 \subset ...$ and E_{n+1} constructed from E_n using 6). For 8) take K^a be the set of all elements in C which are algebraic over K.

We discuss in what follows polynomial congruences which lie at the heart of number theory. The main results below are due to Fermat, Lagrange, Euler, and Gauss.

Definition 28.35. Let $f(x) \in \mathbb{Z}[x]$ be a polynomial and p a prime. An integer $c \in \mathbb{Z}$ is called a *root of* $f(x)$ *mod* p (or a solution to the congruence $f(x) \equiv 0$ $(mod\ p)$) if and only if $f(c) \equiv 0$ $(mod\ p)$; in other words if $p|f(c)$. Let $\overline{f} \in \mathbb{F}_p[x]$ be the polynomial obtained from $f \in \mathbb{Z}[x]$ by replacing the coefficients of f with their images in $\mathbb{F}_p$. Then c is a root of f mod p if and only if the image $\overline{c}$ of c in $\mathbb{F}_p$ is a root of $\overline{f}$. We denote by $N_p(f)$ the number of roots of $f(x)$ mod p contained in $\{0, 1, ..., p-1\}$; equivalently $N_p(f)$ is the number of roots of $\overline{f}$ in $\mathbb{F}_p$. If f, g are polynomials in $\mathbb{Z}[x]$ we write $N_p(f = g)$ for $N_p(f - g)$. If $Z_p(f)$ is the set of roots of $\overline{f}$ in $\mathbb{F}_p$ then of course $N_p(f) = |Z_p(f)|$.

Exercise 28.36.

1) 3 is a root of $x^3 + x - 13$ mod 17.

2) Any integer a is a root of $x^p - x$ mod p; this is Fermat's Little Theorem. In particular $N_p(x^p - x) = p$, $N_p(x^{p-1} = 1) = p - 1$.

3) $N_p(ax - b) = 1$ if $p \not| a$.

4) $N_p(x^2 = 1) = 2$ if $p \neq 2$.

Proposition 28.37. *For any two polynomials* $f, g \in \mathbb{Z}[x]$ *we have*

$$N_p(fg) \leq N_p(f) + N_p(g).$$

Proof. Clearly $Z_p(fg) \subset Z_p(f) \cup Z_p(g)$. Hence

$$|Z_p(fg)| = |Z_p(f) \cup Z_p(g)| \leq |Z_p(f)| + |Z_p(g)|.$$

□

EXERCISE 28.38. Consider the polynomials
$$f(x) = x^{p-1} - 1 \text{ and } g(x) = (x-1)(x-2)...(x-p+1) \in \mathbb{Z}[x].$$
Prove that all the coefficients of the polynomial $f(x) - g(x)$ are divisible by p. Conclude that p divides the sums
$$\sum_{a=1}^{p-1} a = 1 + 2 + 3 + ... + (p-1)$$
and
$$\sum_{1 \leq a < b \leq p-1} ab = 1\times2+1\times3\times...1\times(p-1)+2\times3+...+2\times(p-1)+...+(p-2)\times(p-1).$$

EXERCISE 28.39. Assume $p \geq 5$ is a prime. Prove that the numerator of any fraction that is equal to
$$1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{p-1}$$
is divisible by p^2.

REMARK 28.40. Fix a polynomial $f(x) \in \mathbb{Z}[x]$. Some of the deepest problems and theorems in number theory can be formulated as special cases of the following two problems:

1) Understand the set of primes p such that the congruence $f(x) \equiv 0$ $(mod\ p)$ has a solution or, equivalently, such that $p|f(c)$ for some $c \in \mathbb{Z}$.

2) Understand the set of primes p such that $p = f(c)$ for some $c \in \mathbb{Z}$.

In regards to problem 1) one would like more generally to understand the function whose value at a prime p is the number $N_p(f)$. In particular one would like to understand the set of all primes p such that $N_p(f) = k$ for a given k (equivalently such that the congruence $f(x) \equiv 0$ $(mod\ p)$ has k solutions in $\{0, 1, ..., p-1\}$). We note that if $deg(f) = 1$ the problem is trivial. For $deg(f) = 2$ the problem is already highly non-trivial although a complete answer was given by Gauss in his Quadratic Reciprocity Law (to be proved later). For the quadratic polynomial $f(x) = x^2 + 1$, for instance, we will prove below (without using quadratic reciprocity) that $p|f(c)$ for some c if and only if p is of the form $4k+1$. For $deg(f)$ arbitrary the problem (and its generalizations for polynomials $f(x, y, z, ...)$ of several variables) is essentially open and part of an array of tantalizing conjectures (part of the Langlands program) that link the function $N_p(f)$ to Fourier analysis and the theory of complex analytic functions. This is beyond the scope of our course.

In regards to problem 2), by a theorem of Dirichlet, for every linear polynomial $f(x) = ax + b$ for which a and b are coprime there exist infinitely many integers k such that $f(k)$ is prime. But it is not known, for instance, if there are infinitely many integers k such that $f(k)$ is prime when $f(x)$ is a quadratic polynomial such as $f(x) = x^2 + 1$. Problem 2) has an obvious analogue for polynomials in several variables.

The following is a direct consequence of Lagrange's Theorem 28.13:

COROLLARY 28.41. *Assume* $p \equiv 1$ $(mod\ d)$. *Then* $N_p(x^d - 1) = d$.

Proof. By Lagrange's Theorem $N_p(x^d - 1) \leq d$. Assume $N_p(x^d - 1) < d$ and seek a contradiction. If $p - 1 = kd$ then $x^{p-1} - 1 = (x^d - 1)g(x)$ where

$$g(x) = x^{d(k-1)} + x^{d(k-2)} + ... + x^d + 1.$$

Since by Lagrange's Theorem $N_p(g) \leq d(k-1)$ we get

$$p-1 = N_p(x^{p-1}-1) = N_p((x^d-1)g) \leq N_p(x^d-1)+N_p(g) < d+d(k-1) = dk = p-1,$$

a contradiction. □

Corollary 28.42.

1) If $p \equiv 1$ $(mod\ 4)$ then $N_p(x^2 = -1) = 2$. Equivalently every prime p of the form $4k+1$ divides some number of the form c^2+1 where c is an integer.

2) If $p \equiv 3$ $(mod\ 4)$ then $N_p(x^2 = -1) = 0$. Equivalently no prime p of the form $4k+3$ can divide a number of the form c^2+1 where c is an integer.

Proof. 1) By Corollary 28.41 if $p \equiv 1$ $(mod\ 4)$ then $N_p(x^4 - 1) = 4$. But

$$4 = N_p(x^4-1) \leq N_p((x^2-1)(x^2+1)) \leq N_p(x^2-1) + N_p(x^2+1) \leq N_p(x^2+1) + 2$$

hence $N_p(x^2+1) \geq 2$ and we are done.

2) Assume $p \equiv 3$ $(mod\ 4)$ so $p = 4k+3$ and assume $N_p(x^2 = -1) > 0$ so there exists $c \in \mathbb{Z}$ such that $c^2 \equiv -1$ $(mod\ p)$; we want to derive a contradiction. We have (by Fermat's Little Theorem) that $c^p \equiv c$ $(mod\ p)$. Since $p \not| c$ we get $c^{p-1} \equiv 1$ $(mod\ p)$. But

$$c^{p-1} \equiv c^{4k+2} \equiv (c^2)^{2k+1} \equiv (-1)^{2k+1} \equiv -1 \ (mod\ p),$$

a contradiction. □

Exercise 28.43. Prove that:

1) If $p \equiv 1$ $(mod\ 3)$ then $N_p(x^2+x+1) = 2$. Equivalently every prime p of the form $3k+1$ divides some number of the form c^2+c+1.

2) If $p \equiv 2$ $(mod\ 3)$ then $N_p(x^2+x+1) = 0$. Equivalently no prime p of the form $3k+2$ can divide a number of the form c^2+c+1.

Definition 28.44. Let a be an integer not divisible by a prime p. The *order* of a mod p is the smallest positive integer k such that $a^k \equiv 1$ $(mod\ p)$. We write $k = o_p(a)$. Clearly $o_p(a)$ equals the order $o(\overline{a})$ of the image $\overline{a}$ of a in $\mathbb{F}_p$.

Definition 28.45. An integer g is a *primitive root* mod p if it is not divisible by p and $o_p(g) = p-1$, equivalently, if the image $\overline{g}$ of g in $\mathbb{F}_p^\times$ is a generator of the group $\mathbb{F}_p^\times$.

Exercise 28.46. Prove that g is a primitive root mod p if and only if it is not divisible by p and

$$g^{(p-1)/q} \not\equiv 1 \ (mod\ p)$$

for all primes $q|p-1$.

Exercise 28.47. Prove that 3 is a primitive root mod 7 but 2 is not a primitive root mod 7.

The following Theorem about the existence of primitive roots was conjectured by Euler and proved by Gauss:

Theorem 28.48. *(Gauss) If p is a prime there exists a primitive root mod p. Equivalently the group $\mathbb{F}_p^\times$ is cyclic.*

Proof. By the Fundamental Theorem of Arithmetic, $p-1 = p_1^{e_1}...p_s^{e_s}$ with $p_1,...,p_s$ distinct primes and $e_1,...,e_s \geq 1$. Let $i \in \{1,...,s\}$. By Corollary 28.41 $N_p(x^{p_i^{e_i}}-1) = p_i^{e_i}$ and $N_p(x^{p_i^{e_i-1}}-1) = p_i^{e_i-1}$. So $x^{p_i^{e_i}}-1$ has a root c_i mod p which is not a root mod p of $x^{p_i^{e_i-1}}-1$. So

$$c_i^{p_i^{e_i}} \equiv 1 \ (mod\ p),$$
$$c_i^{p_i^{e_i-1}} \not\equiv 1 \ (mod\ p).$$

It follows that the order of c_i is a divisor of $p_i^{e_i}$ but not a divisor of $p_i^{e_i-1}$. Hence

$$o_p(c_i) = p_i^{e_i}.$$

By Proposition 24.25

$$o_p(c_1...c_s) = p_1^{e_1}...p_s^{e_s} = p-1$$

so $c_1...c_s$ is a primitive root mod p. □

Note that one can consider polynomials in several variables as follows.

Definition 28.49. Let R be a ring. We define the *ring of polynomials in* 2 *variables* as the set of all maps $\mathbb{Z}_{\geq 0}^2 \to R$ equipped with the following addition and multiplication. If a map $(i,j) \mapsto a_{ij}$ is given we denote it by (a_{ij}) and we set

$$(a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij}),$$
$$(a_{ij}) \times (b_{ij}) = (c_{ij}), \quad c_{ij} = \sum_{i_1+i_2=i, j_1+j_2=j} a_{i_1 j_1} b_{i_2 j_2}.$$

We denote this ring by $R[x,y]$. We let $x \in R[x,y]$ be the map that sends $(1,0) \mapsto 1$ and all other pairs into 0; similarly we let $y \in R[x,y]$ be the map that sends $(0,1) \mapsto 1$ and all other pairs into 0. We define the degree of $(a_{ij}) \neq 0$ as the maximum d such that there exist i,j with $i+j=d$ and $a_{ij} \neq 0$. For $a \in R$ we continue to denote by $a \in R[x,y]$ the map that sends $(0,0) \mapsto a$ and sends all other pairs into 0.

Exercise 28.50. Prove that

$$(a_{ij}) = \sum_{ij} a_{ij} x^i y^j.$$

Exercise 28.51. We defined the ring of polynomials $R[x]$ in one variable x with coefficients in R. Now $R[x]$ is again a ring so we can consider the ring of polynomials $R[x][y]$ in one variable y with coefficients in $R[x]$. Prove that $R[x,y]$ is isomorphic to $R[x][y]$ and also to $R[y][x]$.

Exercise 28.52. Define the ring of polynomials in n variables $x_1,...,x_n$ (and denote it by $R[x_1,...,x_n]$).

Definition 28.53. For K a field the field of fractions of $K[x]$ is denoted by $K(x)$ and referred to as the *field of rational functions* in one variable x. Similarly the field of fractions of $K[x_1,...,x_n]$ is denoted by $K(x_1,...,x_n)$ and referred to as the *field of rational functions* in n variables.

CHAPTER 29

Invariants

Let R be a ring and $R[x_1, ..., x_n]$ the ring of polynomials in n variables. Consider the following polynomials in this ring (called the *elementary symmetric polyomials*):

$$\begin{array}{lll} s_1 & := & x_1 + ... + x_n, \\ s_2 & := & \sum_{1 \leq i < j \leq n} x_i x_j \\ ... & ... & ... \\ s_k & := & \sum_{1 \leq i_1 < ... < i_k \leq n} x_{i_1} ... x_{i_k} \\ ... & ... & ... \\ s_n & := & x_1 ... x_n. \end{array}$$

DEFINITION 29.1. A polynomial $f \in R[x] := R[x_1, ..., x_n]$ is called a *symmetric polynomial* if for every permutation $\sigma \in S_n$ we have

$$f(x_{\sigma(1)}, ..., x_{\sigma(n)}) = f(x_1, ..., x_n).$$

Let $R[x_1, ..., x_n]^{S_n}$ be the ring of symmetric polynomials.

EXAMPLE 29.2.
1) $s_1, ..., s_n$ are symmetric polynomials.
2) If $y_1, ..., y_n$ are variables then every polynomial in the image of the ring homomorphism

$$\phi : R[y_1, ..., y_n] \to R[x_1, ..., x_n], \quad g(y_1, ..., y_n) \mapsto g(s_1, ..., s_n)$$

is symmetric. We denote by $R[s_1, ..., s_n]$ the image of ϕ; so

$$R[s_1, ..., s_n] \subset R[x_1, ..., x_n]^{S_n}.$$

PROPOSITION 29.3. *The ring homomorphism ϕ above is injective.*

Proof. Induction on n. For the induction step assume ϕ is not injective for some n and let f be of minimal degree d such that $\phi(f) = 0$,

$$f(y_1, ..., y_n) = f_0(y_1, ..., y_{n-1}) + f_1(y_1, ..., y_{n-1})y_n + ... + f_d(y_1, ..., y_{n-1})y_n^d.$$

So $f_0 \neq 0$ and

$$0 = f(s_1, ..., s_n) = f_0(s_1, ..., s_{n-1}) + f_1(s_1, ..., s_{n-1})s_n + ... + f_d(s_1, ..., s_{n-1})s_n^d.$$

Make $x_n = 0$ in the the above. One gets

$$f_0(s_1', ..., s_{n-1}') = 0$$

where s'_k is the k-th elementary symmetric polynomial in $n-1$ variables $x_1, ..., x_{n-1}$. By the induction hypothesis $f_0 = 0$, a contradiction. $\square$

THEOREM 29.4. *(Fundamental Theorem of Symmetric Polynomials).*

$$R[x_1, ..., x_n]^{S_n} = R[s_1, ..., s_n].$$

Proof (in case R is a field K). Every symmetric polynomial can be uniquely written as

$$\sum_{0 \le k_1 \le ... \le k_n} \lambda_{k_1...k_n} t_{k_1...k_n}$$

where $\lambda_{k_1...k_n} \in K$ and

$$t_{k_1...k_n} = \sum_{\sigma \in S_n} x_{\sigma(1)}^{k_1} ... x_{\sigma(n)}^{k_n} = \sum_{\sigma \in S_n} x_1^{k_{\sigma(1)}} ... x_n^{k_{\sigma(n)}}.$$

Say that a polynomial in $K[y_1, ..., y_n]$ has *weight* d if all its monomials $y_1^{l_1} ... y_n^{l_n}$ satisfy

$$l_1 + 2l_2 + ... + nl_n = d$$

and let $K[y_1, ..., y_n]^d$ be the collection of all polynomials of weight d. Also say that a polynomial in $K[x_1, ..., x_n]$ is *homogeneous of degree* d if all monomials $x_1^{k_1} ... x_n^{k_n}$ in it satisfy $k_1 + ... + k_n = d$. Denote by $K[x_1, ..., x_n]_d$ be the collection of all homogeneous polynomials of degree d. Let

$$K[x_1, ..., x_n]_d^{S_n} := K[x_1, ..., x_n]^{S_n} \cap K[x_1, ..., x_n]_d$$

be the set of all symmetric homogeneous polynomials of degree d. Clearly every symmetric polynomial is a sum of symmetric homogeneous polynomials of various degrees. Also

$$\phi(K[y_1, ..., y_n]^d) \subset K[x_1, ..., x_n]_d^{S_n}.$$

Since ϕ is injective it is enough to show that the the following dimensions of vector spaces over K are equal:

$$\dim K[y_1, ..., y_n]^d = \dim K[x_1, ..., x_n]_d^{S_n}.$$

We already established that the right hand side has dimension equal to the cardinality of the set T_d of all tuples of integers $k_1, ..., k_n$ such that $0 \le k_1 \le k_2 \le ... \le k_n$ and $k_1 + k_2 + ... + k_n = d$. On the other hand a basis for $K[y_1, ..., y_n]^d$ consists of all monomials

$$y_1^{k_n - k_{n-1}} y_2^{k_{n-1} - k_{n-2}} ... y_n^{k_1} \quad \text{with} \quad (k_1, ..., k_n) \in T_d.$$

$\square$

EXAMPLE 29.5. There exists a polynomial $\Delta = \Delta_n \in K[y_1, ..., y_n]$ such that

$$\Delta(s_1, ..., s_n) = \prod_{1 \le i < j \le n} (x_i - x_j)^2$$

because the right hand side is a symmetric polynomial. Δ is called the *discriminant* polynomial. Note that if

$$f(t) = t^n - \gamma_1 t^{n-1} + ... + (-1)^n \gamma_n \in K[t]$$

splits in a field extension L of K as

$$f(t) = (t - \lambda_1)...(t - \lambda_n)$$

with $\lambda_i \in L$ then

$$\gamma_i = s_i(\lambda_1, ..., \lambda_n)$$

for all i. Hence two of the λ_is coincide (equivalently f has a multiple root in L) if and only if

$$\Delta(\gamma_1, ..., \gamma_n) = 0.$$

The above theorem fits into the following general setting.

Definition 29.6. Let G be group (with identity element e) and X a set. By an *action* of G on X one understands a map

$$G \times X \to X, \quad (g, x) \mapsto gx$$

such that for all $g, h \in G$ and all $x \in X$ one has

1) $g(hx) = (gh)x$;
2) $ex = x$.

One says that $x \in X$ is *fixed* by G if $gx = x$ for all $g \in G$ and one lets X^G be the set of all fixed elements.

Definition 29.7. Let A be a group (resp. a module over a ring, resp. a ring). By an action of a group G on A one understands an action of G on the set A such that, in addition, for all $g \in G$ the map $X \to X$, $x \mapsto gx$, is a group homomorphism (resp. a linear map, resp. a ring homomorphism).

Remark 29.8.

1) Note that in the notation above $x \mapsto gx$ is automatically bijective with inverse $x \mapsto g^{-1}x$ which is then, again, a group homomorphism (resp. a linear map, resp. a ring homomorphism). Also A^G is, again, a group (resp. a module, resp. a ring) called the group (resp. module, resp. ring) of *invariants* of G in A.

2) In our discussion above S_n acts on the ring $R[x_1, ..., x_n]$ by the formula

$$(\sigma, f) \mapsto \sigma f := f(x_{\sigma(1)}, ..., x_{\sigma(n)})$$

and the notation $R[x_1, ..., x_n]^{S_n}$ in Definition 29.1 agrees with the corresponding one in Definition 29.6.

3) The main problem of invariant theory is to "compute" the ring of invariants A^G for a given action of a group G on a ring A.

Exercise 29.9. Let K be a field and $n \geq 2$. Consider the action of $GL_n(K)$ on the ring of polynomials

$$K[X] := K[x_{11}, x_{12}, ..., x_{nn}]$$

in n^2 variables given for $U = (u_{ij}) \in GL_n(K)$ and $f = f(x_{11}, x_{12}, ..., x_{nn})$ by

$$(U, f) \mapsto f(\ell_{11}, \ell_{12}, ..., \ell_{nn})$$

where, for X the matrix $X = (x_{ij})$, ℓ_{ij} is the ij-entry of the matrix UXU^{-1}. Prove that the coefficients $c_i = c_i(X) := c_i(x_{11}, x_{12}, ..., x_{nn}) \in K[X]$ of the *characteristic polynomial*

$$\Phi_X(t) := \det(tI - X) = \sum_{i=0}^{n} c_i(X)t^{n-i} \in K[X][t]$$

belong to the ring of invariants $K[X]^{GL_n(K)}$. Note that

$$c_0(X) = 1, \quad c_1(X) = -Tr(X), \quad c_n(X) = (-1)^n \det(X).$$

If $A \in K^{n\times n}$ then one defines the characteristic polynomial of A as

$$\Phi_A(t) = \det(tI - A) = \Phi_X(t)_{|X=A}.$$

Note: In fact one can prove that every element of $K[X]^{GL_n(K)}$ is a polynomial combination of $c_1, ..., c_n$ i.e., it has the form $g(c_1, ..., c_n)$ for some polynomial g in n variables with coefficients in K. On the other hand it is easy to find examples of matrices $A = (a_{ij}), B = (b_{ij}) \in K^{n\times n}$ such that

$$c_i(A) = c_i(B)$$

for all $i = 1, ..., n$ but for which there is no $U \in GL_n(K)$ such that $B = UAU^{-1}$. E.g., one can take $A = I$, the identity matrix and B a matrix with $b_{ii} = 1$ for all i and $b_{ij} = 0$ for all $i > j$.

Theorem 29.10. *(Hamilton-Cayley Theorem). Let $A \in K^{n\times n}$ be a matrix and $\Phi_A(t) = t^n - c_1(A)t^{n-1} + ... + (-1)^n c_n(A) \in K[t]$ be its characteristic polynomial. Then*

$$A^n - c_1(A)A^{n-1} + ... + (-1)^n c_n(A)I = 0 \in K^{n\times n}.$$

Proof. We first prove the theorem in case the characteristic polynomial $\Phi_A(t)$ has only simple roots $\lambda_1, ..., \lambda_n$ in an algebraic closure K^a of K. Replacing K by K^a let $v_1, ..., v_n \in K^n = K^{n\times 1}$ be eigenvectors belonging to $\lambda_1, ..., \lambda_n$. Since the λ_is are distinct $v_1, ..., v_n$ form a basis on K^n. So it is enough to prove that

$$(A^n - c_1(A)A^{n-1} + ... + (-1)^n c_n(A)I)v_i = 0 \in K^n$$

for all i. But the left hand side equals

$$(\lambda_i^n - c_1(A)\lambda_i^{n-1} + ... + (-1)^n c_n(A))v_i = 0 \cdot v_i = 0 \in K^n$$

and we are done.

We now prove the theorem in the general case. It is enough to prove that

$$X^n - c_1(X)X^{n-1} + ... + (-1)^n c_n(X)I = 0 \in K[X]^{n\times n}$$

because then we can simply set X equal to A. So we shall be done if we check that $\Phi_X(t)$ has simple roots only in an algebraic closure of the fraction field of $K[X]$. So we need to show that the discriminant polynomial Δ evaluated at the coefficients of $\Phi_X(t)$ is non-zero. But if it were zero we would get a contradiction by setting X equal to a matrix which is, say, diagonal, with distinct entries on the diagonal. $\square$

We next present some elements of Galois Theory.

Definition 29.11. Let $K \subset L$ be a field extension. We denote by $G(L/K)$ the group of all field isomorphisms $\sigma : L \to L$ such that $\sigma(c) = c$ for all $c \in K$.

Definition 29.12. Let G be a group acting on a field L. We denote (as usual) by L^G the *fixed field* i.e., the field of all $\alpha \in L$ such that $\sigma(\alpha) = \alpha$ for all $\sigma \in G$. We say that G acts *effectively* on L if whenever $\sigma \in G$ satisfies $\sigma(\alpha) = \alpha$ for all $\alpha \in L$ we must have that σ is the identity element of G. (In this case we can, and will, identify G with a subgroup of the group of all isomorphisms $L \to L$.)

The following result is “half” of the Main Theorem of Galois Theory:

Theorem 29.13. *Let G be a finite group acting effectively on a field L. Then the following hold:*

1) Every $\alpha \in L$ is algebraic over L^G; moreover, the irreducible polynomial φ_{α/L^G} splits completely in $L[x]$, has only simple roots in L, and has degree at most the order $|G|$ of G.

2) There exists $\gamma \in L$ such that $L = L^G(\gamma)$.

3) $G(L/L^G) = G$.

Proof. For 1) let $\{\sigma_1, ..., \sigma_m\}$ be a subset of G with σ_i distinct and m maximal with the property that the elements $\sigma_1(\alpha), ..., \sigma_m(\alpha)$ of L are distinct. Then one checks that the polynomial

$$f(x) := \prod_{i=1}^{m}(x - \sigma_i(\alpha))$$

has coefficients in L^G, vanishes at α, has only simple roots in L, and has degree $\leq |G|$. Hence φ_{α/L^G} has the same properties.

For 2) pick an element $\gamma \in L$ such that $[L^G(\gamma) : L^G]$ is maximal. We claim that $L = L^G(\gamma)$; for if this is not the case take $\beta \in L \setminus L^G$ and write (by the Theorem of the Primitive element) $L^G(\gamma, \beta) = L^G(\eta)$ for some η and note that $[L^G(\eta) : L^G] = [L^G(\eta) : L^G(\gamma)] \cdot [L^G(\gamma) : L^G]$ contradicting the maximality above.

For 3) note that $G \subset G(L/L^G)$ trivially. On the other hand, with γ as in 2) every $\sigma \in G(L/L^G)$ must send γ into a root of φ_{γ/L^G}. There are exactly $[L : L^G] = \deg(\varphi_{\gamma/L^G})$ such roots so there are at most $\deg(\varphi_{\gamma/L^G})$ elements in $G(L/L^G)$. But by 1) we have $\deg(\varphi_{\gamma/L^G}) \leq |G|$. So we conclude that $|G(L/L^G)| \leq \deg(\varphi_{\gamma/L^G}) \leq |G|$ which forces the equality in 3). □

Definition 29.14. A finite field extension $K \subset L$ is called *normal* if there exists a polynomial $f \in K[x]$ such that f splits completely in $L[x]$ and $L = K(\alpha_1, ..., \alpha_n)$ where $\alpha_1, ..., \alpha_n$ are the roots of f in L. (Here $K(\alpha_1, ..., \alpha_n)$ is, by definition, $K(\alpha_1)(\alpha_2)...(\alpha_n)$.) One also says that L is the *splitting field* of f over K.

The following is the Main Theorem of Galois theory in characteristic zero. A similar theorem holds in characteristic p but for simplicity we skip that.

Theorem 29.15. *Let $K \subset L$ be a finite normal field extension with $char(K) = 0$. Let $\mathcal{F}$ be the set of subfields of L containing K and let $\mathcal{G}$ be the set of all subgroups H of $G(L/K)$. Then the maps*

$$\mathcal{G} \to \mathcal{F}, \quad H \mapsto L^H$$

and

$$\mathcal{F} \to \mathcal{G}, \quad E \mapsto G(L/E)$$

are inverse to each other.

Proof. The equality $G(L/L^H) = H$ follows from Theorem 29.13. For the equality $L^{G(L/E)} = E$ note that the inclusion $\supset$ is clear. Assume now there exists $\alpha \in L^{G(L/E)} \setminus E$ and seek a contradiction. Since $\alpha \notin E$ the degree of $\varphi_{\alpha/E}$ is ≥ 2. Since the characteristic is zero α is a simple root of f so if L^a is an algebraic closure of L then L is also an algebraic closure E^a of E and it contains a root β of f different from α. Consider the composition of homomorphisms

$$E(\alpha) \simeq E[x]/(\varphi_{\alpha/E}) \simeq E(\beta) \subset E^a.$$

By 9) in Exercise 28.34 one can extend the above to a homomorphism $\sigma : L \to E^a$. Since L is normal $\sigma(L) \subset L$ hence by degree considerations $\sigma(L) = L$. So σ defines an element of $G(L/E)$ such that $\sigma(\alpha) = \beta \neq \alpha$. This contradicts the fact that $\alpha \in L^{G(L/E)}$. $\square$

CHAPTER 30

Lines

Geometry is the study of shapes such as lines and planes, or, more generally, curves and surfaces, etc. There are two paths towards this study: the synthetic one and the analytic one. Synthetic geometry is geometry whose main concepts are introduced without the help of coordinates. Analytic geometry is geometry whose main concepts are introduced using coordinates. Analytic geometry roughly comes in two flavors: algebraic geometry and differential geometry. The first is based on algebra while the second is based on calculus. Synthetic geometry originates with the Greek Mathematics of antiquity (e.g., the treatise of Euclid). Analytic geometry was invented by Fermat and Descartes. We already encountered the synthetic approach in the discussion of the affine plane and the projective plane which were purely combinatorial objects. Here we introduce some of the most elementary structures of algebraic geometry and discuss the simplest plane curves: lines, conics, and cubics. Differential geometry will be discussed later (see Chapter 37).

Throughout this Chapter K denotes a field.

Definition 30.1. The *affine plane* $\mathbb{A}^2 = \mathbb{A}^2(K)$ over K is the set $K^2 = K \times K$. A *point* $P = (x, y)$ in the plane is an element of K^2; x, y will be called the *coordinates* of P.

Definition 30.2. A subset $L \subset K^2$ is called a *line* if there exist $a, b, c \in K$ such that $(a, b) \neq (0, 0)$ and

$$L = \{(x, y) \in K^2 \mid ax + by + c = 0\}.$$

We say a point P *lies* on the line L (or we say L *passes* through P) if $P \in L$. Two lines are said to be *parallel* if they either coincide or their intersection is empty (in the last case we say they don't meet). Three points are *collinear* if they lie on the same line. We sometimes write $L = L(K)$ if we want to stress that coordinates are in K.

Remark 30.3. Note the use here of the letters x, y to denote elements of K (rather than "variables" of a polynomial); this should not introduce any confusion as long as we keep in mind the two possible meanings of these letters.

Exercise 30.4. Prove that:

1) There exist 3 points which are not collinear.

2) For any two distinct points P_1 and P_2 there exists a unique line L (denoted by $L_{P_1P_2}$) passing through P_1 and P_2. In particular every two non-parallel distinct lines meet in exactly one point.

3) Given a line L and a point P there exists exactly one line L' passing through P and parallel to L. (This is called *Euclid's fifth postulate* but in our exposition here this is not a postulate.)

Hence $\mathbb{A}^2 = K^2$ together with the set $\mathcal{L}$ of all lines (in the sense above) is an affine plane in the sense of Definition 11.40.

Hint for 2): Hint: If $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and if

$$m = (y_2 - y_1)(x_2 - x_1)^{-1}$$

then the unique line through P_1 and P_2 is:

$$L = \{(x, y) \in K^2 \mid y - y_1 = m(x - x_1)\}.$$

REMARK 30.5. Not all affine planes in the sense of Definition 11.40 are affine planes over a field in the sense above. Hilbert proved that an affine plane is the affine plane over some field if and only if the theorems of Desargues and Pappus (stated below) hold. See below for the "only if direction."

EXERCISE 30.6. Prove that every line in $\mathbb{F}_p^2$ has exactly p points.

EXERCISE 30.7. How many lines are there in the plane $\mathbb{F}_p^2$?

EXERCISE 30.8. (Desargues' Theorem) Let $A_1, A_2, A_3, A_1', A_2', A_3'$ be distinct points in the plane. Also for all $i \neq j$ assume $L_{A_iA_j}$ and $L_{A_i'A_j'}$ are not parallel and let P_{ij} be their intersection. Assume the 3 lines $L_{A_1A_1'}, L_{A_2A_2'}, L_{A_3A_3'}$ have a point in common. Then prove that the points P_{12}, P_{13}, P_{23} are collinear.

Hint: Consider the "space" $K^3 := K \times K \times K$ and define planes and lines in this space. Prove that if two planes meet and don't coincide then they meet in a line. Then prove that through any two points in space there is a unique line and through any 3 non-collinear points there is a unique plane. Now consider the projection $K^3 \to K^2$, $(x, y, z) \mapsto (x, y)$ and show that lines project onto lines. Next show that configuration of points $A_i, A_i' \in K^2$ can be realized as the projection of a similar configuration of points $B_i, B_i' \in K^3$ not contained in a plane. (Identifying K^2 with the set of points in space with zero third coordinate we take $B_i = A_i$, $B_i' = A_i'$ for $i = 1, 2$, we let B_3 have a nonzero third coordinate, and then we choose B_3' such that the lines $L_{B_1B_1'}, L_{B_2B_2'}, L_{B_3B_3'}$ have a point in common.) Then prove "Desargues' Theorem in Space" (by noting that if Q_{ij} is the intersection of $L_{B_iB_j}$ with $L_{B_i'B_j'}$ then Q_{ij} is in the plane containing B_1, B_2, B_3 and also in the plane containing B_1', B_2', B_3'; hence Q_{ij} is in the intersection of these planes which is a line). Finally deduce the original plane Desargues by projection.

EXERCISE 30.9. (Pappus' Theorem) Let P_1, P_2, P_3 be points on a line L and let Q_1, Q_2, Q_3 be points on a line $M \neq L$. Assume the lines $L_{P_2Q_3}$ and $L_{P_3Q_2}$ are not parallel and let A_1 be their intersection; define A_2, A_3 similarly. Then prove that A_1, A_2, A_3 are collinear.

Hint (for the case L and M meet): One can assume $L = \{(x, 0) \mid x \in K\}$, $M = \{(0, y) \mid y \in K\}$ (explain why). Let the points $P_i = (x_i, 0)$ and $Q_i = (0, y_i)$ and compute the coordinates of A_i. Then check that the line through A_1 and A_2 passes through A_3.

REMARK 30.10. One can identify the projective plane $(\overline{\mathbb{A}^2}, \overline{\mathcal{L}})$ attached to the affine plane $(\mathbb{A}^2, \mathcal{L})$ (cf. Definition 11.46) with the pair $(\mathbb{P}^2, \check{\mathbb{P}}^2)$ defined as follows.

Let $\mathbb{P}^2 = K^3/\sim$ where $(x, y, z) \sim (x', y', z')$ if and only if there exists $0 \neq \lambda \in K$ such that $(x', y', z') = (\lambda x, \lambda y, \lambda z)$. Denote the equivalence class of (x, y, z) by $(x : y : z)$. Identify a point (x, y) in the affine plane $\mathbb{A}^2 = K^2 = K \times K$ with the point $(x : y : 1) \in \mathbb{P}^2$. Identify a point $(x_0 : y_0 : 0)$ in the complement $\mathbb{P}^2 \backslash \mathbb{A}^2$ with the class of lines in $\mathbb{A}^2$ parallel to the line $y_0 x - x_0 y = 0$. This allows one to identify the complement $\mathbb{P}^2 \backslash \mathbb{A}^2$ with the line at infinity L_∞ of $\mathbb{A}^2$. In particular we get an identification between L_∞ and the set of all lines in $\mathbb{A}^2$ passing through $(0, 0)$. Hence we get an identification of $\mathbb{P}^2$ with $\overline{\mathbb{A}^2}$ and we may write

$$L_\infty = \{(x : y : 0) \mid (x, y) \in K^2 \setminus \{(0, 0)\}\} = (K^2 \setminus \{(0, 0)\})/\sim =: \mathbb{P}^1,$$

where the latter $=$ is a definition and $\sim$ is defined by $(x, y) \sim (x', y')$ iff there exists $\lambda \in K^\times$ such that $x' = \lambda x$, $y' = \lambda y$. Finally define a *line* in $\mathbb{P}^2$ (or a *projective line*) as a set of the form

$$\overline{L} = \{(x : y : z) \mid ax + by + cz = 0\}.$$

So under the above identifications,

$$\overline{L} = \{(x : y : 1) \mid ax + by + c = 0\} \cup \{(x : y : 0) \mid ax + by = 0\} = L \cup \{\widehat{L}\}$$

where L is the line in K^2 defined by $ax + by + c = 0$. Then define $\check{\mathbb{P}}^2$ to be the set of all lines $\overline{L}$ in $\mathbb{P}^2$. We get an identification of $\check{\mathbb{P}}^2$ with $\overline{\mathcal{L}}$.

Some familiar concepts such as area and distance can be defined in the above context. Assume in what follows that K is a field such that $2 \neq 0$ and identify K^2 with $K^{2\times 1}$.

Definition 30.11. Let $P_1, P_2, P_3 \in K^{2\times 1}$ be 3 points in the plane. Define

$$\text{area}(P_1, P_2, P_3) = \frac{1}{2}\det(P_2 - P_1, P_3 - P_1) \in K.$$

Exercise 30.12. Prove that

1) $\text{area}(P_{\sigma(1)} P_{\sigma(2)} P_{\sigma(3)}) = \epsilon(\sigma) \cdot \text{area}(P_1, P_2, P_3)$ for all permutations $\sigma \in S_3$.

2) $\text{area}(P_1, P_2, P_3) = 0$ if and only if P_1, P_2, P_3 are collinear.

3) Let $F : K^2 \to K^2$ be an isomorphism of vector spaces and A its matrix with respect to the canonical basis. Then $A \in SL_2(K)$ if and only if "F preserves areas" in the sense that for all $P_1, P_2, P_3 \in K^2$ we have

$$\text{area}(F(P_1), F(P_2), F(P_3)) = \text{area}(P_1, P_2, P_3).$$

4) For every $P_0 \in K^{2\times 1}$ area is "invariant under translation by P_0" in the sense that

$$\text{area}(P_1 + P_0, P_2 + P_0, P_3 + P_0) = \text{area}(P_1, P_2, P_3).$$

5) Reformulate 3) and 4) in terms of the area being an invariant for a group action on a ring. Reformulate 1) in a similar way.

Hint for 5): Consider the actions of the groups $SL_2(K)$, K^2, and S_3 on the ring

$$A = Fun(K^2 \times K^2 \times K^2, K)$$

of all K-valued functions on the set of triples of points in K^2 and view area and area^2 as elements of A.

Definition 30.13. Let $P_1, P_2 \in K^{2\times 1}$ be 2 points in the plane. Define the *distance squared* between these points as

$$\text{dist}^2(P_1, P_2) = (P_2 - P_1)^t (P_2 - P_1) \in K.$$

EXERCISE 30.14. Prove that

1) $\mathrm{dist}^2(P_1, P_2) = \mathrm{dist}^2(P_2, P_1)$.

2) If $K = \mathbb{R}$ then $\mathrm{dist}^2(P_1, P_2) = 0$ if and only if $P_1 = P_2$. (Show that this may fail for other fields.)

3) Let $F : K^2 \to K^2$ be an isomorphism of vector spaces and A its matrix with respect to the canonical basis. Then $A \in SO_2(K)$ if and only if "F preserves areas" and also "preserves distances" in the sense that for all $P_1, P_2 \in K^{2\times 1}$ we have

$$\mathrm{dist}^2(F(P_1), F(P_2)) = \mathrm{dist}^2(P_1, P_2).$$

4) For every $P_0 \in K^{2\times 1}$, dist^2 is "invariant under translation by P_0" in the sense that

$$\mathrm{dist}^2(P_1 + P_0, P_2 + P_0) = \mathrm{dist}^2(P_1, P_2).$$

5) Reformulate 1), 3) and 4) in terms of the area^2 being an invariant for a group action on a ring.

For the next Exercises recall that for every matrix $A \in M_2(K)$ we may consider the linear map (still denoted by)

$$K^{2\times 1} = K^2 \to K^2, \quad v \mapsto Av.$$

EXERCISE 30.15. (Three lines through the origin). Prove that if L_1, L_2, L_3 are three distinct lines and L_1', L_2', L_3' are three distinct lines, all these lines passing through the point $O = (0, 0)$, then there exists $A \in GL_2(K)$ such that $AL_i = L_i'$ for $i = 1, 2, 3$.

Hint: We may assume L_1, L_2, L_3 are the lines $\ell_0, \ell_1, \ell_\infty$ given by $y = 0$, $y = x$, and $x = 0$, respectively. Then one computes directly.

EXERCISE 30.16. (Four lines through the origin: cross ratio.) This goes back to Desargues. Let L_1, L_2, L_3, L_4 be four distinct lines in the affine plane passing through the point $O = (0, 0)$,

$$L_i = \{(x, y) \in K^2 \mid a_i x + b_i y = 0\}, \quad (a_i, b_i) \in K^2 \setminus \{(0, 0)\}, \quad i \in \{1, 2, 3, 4\}.$$

Define

$$[i, j] := \det \begin{pmatrix} a_i & b_i \\ a_j & b_j \end{pmatrix}, \quad i, j \in \{1, 2, 3, 4\},$$

and define the *cross ratio* of the four lines by

$$\mathrm{cross}(L_1, L_2, L_3, L_4) := \frac{[12][34]}{[13][24]} \in K^\times.$$

1) Prove that if $A \in GL_2(K)$ and $L_i' := AL_i$ then we have

$$\mathrm{cross}(L_1', L_2', L_3', L_4') = \mathrm{cross}(L_1, L_2, L_3, L_4).$$

2) Prove that, conversely, if

$$\mathrm{cross}(L_1', L_2', L_3', L_4') = \mathrm{cross}(L_1, L_2, L_3, L_4)$$

for distinct L_1, L_2, L_3, L_4 and distinct L_1', L_2', L_3', L_4' then there exists $A \in GL_2(K)$ such that $L_i' = AL_i$ for all i.

3) Prove that

$$\begin{aligned}
\operatorname{cross}(L_4, L_3, L_2, L_1) &= \operatorname{cross}(L_1, L_2, L_3, L_4)^{-1}, \\
\operatorname{cross}(L_3, L_2, L_1, L_4) &= 1 - \operatorname{cross}(L_1, L_2, L_3, L_4) \\
\operatorname{cross}(L_2, L_1, L_3, L_4) &= \operatorname{cross}(L_1, L_2, L_3, L_4)/(\operatorname{cross}(L_1, L_2, L_3, L_4) - 1).
\end{aligned}$$

4) Find similar relations in the case of the other 3 transpositions in S_4, i.e., for

$$\operatorname{cross}(L_1, L_3, L_2, L_4), \quad \operatorname{cross}(L_1, L_4, L_3, L_2), \quad \operatorname{cross}(L_1, L_2, L_4, L_3).$$

5) Prove that if $z := \operatorname{cross}(L_1, L_2, L_3, L_4)$ then for all $\sigma \in S_4$ we have

$$\operatorname{cross}(L_{\sigma(1)}, L_{\sigma(2)}, L_{\sigma(3)}, L_{\sigma(4)}) \in \{z, \frac{1}{z}, 1 - z, \frac{1}{1-z}, \frac{z}{z-1}, \frac{z-1}{z}\}.$$

6) Reformulate 1) in terms of cross being an invariant for a certain action of a group on a ring. Note: there is a way to reformulate 5) in this way but we will not consider this here.

Hint for 1): Follows from the multiplicativity of det.

Hint for 2): Let $\ell_0, \ell_1, \ell_\infty, \ell_\lambda$ be the lines given by the equations $y = 0$, $y = x$, $x = 0$, $y = \lambda x = 0$, respectively. By Exercise 30.15 and Part 1) of the present Exercise we may assume $L_1 = L_1' = \ell_0$, $L_2 = L_2' = \ell_1$ and $L_3 = L_3' = \ell_\infty$, $L_4 = \ell_\lambda$, $L_4' = \ell_{\lambda'}$. But then a trivial computation gives $\lambda = \lambda'$.

Hint for 3) and 4): Direct computation. We may assume $a_i \in \{0, 1\}$ for all i which somewhat simplifies the computations.

Hint for 5): By 3) and 4) the statement is true for all 6 transpositions in S_4. Then use the fact that every permutation is a product of transpositions plus the fact that the 6 expressions in 5), viewed as functions of z, form a group with respect to composition. (This group is, by the way, isomorphic to S_3.)

REMARK 30.17. Since by Remark 30.10 we have an identification between the set of lines in $\mathbb{A}^2$ and the set $\mathbb{P}^1$ we have a well defined notion of *cross ratio* for 4-tuples of distinct points of $\mathbb{P}^1$.

CHAPTER 31

Conics

So far we were concerned with lines in the plane. Let us discuss now "higher degree curves." We start with conics. Assume K is a field with $2 := 1 + 1 \neq 0$; equivalently K does not contain the field $\mathbb{F}_2$.

Definition 31.1. The *circle* of center $(a, b) \in K \times K$ and radius r is the set

$$C(K) = \{(x, y) \in K^2 \mid (x - a)^2 + (y - b)^2 = r^2\}.$$

Exercise 31.2. Prove that:
1) A circle and a line meet in at most 2 points.
2) Two circles meet in at most 2 points.

Exercise 31.3. How many points does a circle of radius 1 have if $K = \mathbb{F}_{13}$? Same problem for $\mathbb{F}_{11}$.

Exercise 31.4. Prove that the circle $C(K)$ with center $(0, 0)$ and radius 1 is an Abelian group with $e = (1, 0)$, $(x, y)' = (x, -y)$, and group operation

$$(x_1, y_1) \star (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

Prove that the map

$$C(K) \to SO_2(K), \quad (a, b) \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

is a bijective group homomorphism. (Cf. Exercise 24.13 for $SO_2(K)$.)

Exercise 31.5. Consider the circle $C(\mathbb{F}_{17})$. Show that $(\bar{3}, \bar{3}), (\bar{1}, \bar{0}) \in C(\mathbb{F}_{17})$ and compute $(\bar{3}, \bar{3}) \star (\bar{1}, \bar{0})$ and $2(\bar{1}, \bar{0})$ (where the latter is, of course, $(\bar{1}, \bar{0}) \star (\bar{1}, \bar{0})$).

Circles are special cases of conics:

Definition 31.6. A *conic* is a subset $Q \subset K \times K$ of the form

$$Q = Q(K) = \{(x, y) \in K^2 \mid ax^2 + bxy + cy^2 + dx + ey + f = 0\}$$

for some $a, b, c, d, e, f \in K$ with $(a, b, c) \neq (0, 0, 0)$. We say a conic *passes* through a point if it contains it.

Note that, according to our definition, a conic may be, at the same time, a line. Indeed this is the case, for instance, when the polynomial $ax^2+bxy+cy^2+dx+ey+f$ is the square of a polynomial of degree 1. This is also the case, for instance, when the polynomial $ax^2+bxy+cy^2$ takes the value 0 for all the values of x, y in K (which can happen even if a, b, c are not all 0). All of this is not, of course, a contradiction but rather a potentially confusing feature of our terminology. In Algebraic Geometry a refinement of the terminology used here is being introduced that greatly clarifies the situation; however introducing this refined terminology here would take us too far.

Exercise 31.7. (Conic through 5 points). Prove that if 5 points are given in $\mathbb{P}^2$ such that no 4 of them are collinear then there exists a unique conic passing through these given 5 points. If no 3 of the 5 points are collinear then the unique conic is not a line or the union of two lines.

Hint: Consider the vector space of all polynomials of degree ≤ 2 that vanish on a set S of points. Next note that if one adds a point to S the dimension of this space either stays the same or drops by one. Since the space of all polynomials of degree ≤ 2 has dimension 6 it is enough to show that for $r \leq 5$ the space of polynomials that vanish at $r-1$ of the r points is strictly bigger than the space of polynomials that vanish at all r points. For $r = 5$, for instance, this is done as follows. Let $P_1, ..., P_5$ be our points and let L_{ij} be the line that passes through P_i and P_j. If neither L_{12} nor L_{34} passes through P_5 the quadric $L_{12} \cup L_{34}$ will not pass through P_5. Assume now that one of the lines L_{12} or L_{34}, for instance L_{12} passes through P_5. Then one checks that none of the lines L_{13} or L_{24} passes through P_5. (Indeed if L_{13} passes through P_5 then L_{12} and L_{13} have 2 points P_1, P_5 in common so they coincide, so P_1, P_2, P_3, P_5 lie on a line, a contradiction; on the other hand if L_{24} passes through P_5 then L_{12} and L_{24} have 2 points P_2, P_5 in common, hence they coincide, hence P_1, P_2, P_4, P_5 line on a line, a contradiction.) So the quadric $L_{13} \cup L_{24}$ does not pass through P_5.

Definition 31.8. For a conic Q as in Definition 31.6 we define its *projective closure*

$$\overline{Q} := \{(x : y : z) \in \mathbb{P}^2 \mid ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0\} \subset \mathbb{P}^2.$$

We call such a $\overline{Q}$ a *projective conic.*

Exercise 31.9. Assume K contains an element i such that $i^2 = -1$. Prove that the projective closure of every circle meets the line at infinity L_∞ in the two points $(1 : i : 0)$ and $(1 : -i : 0)$.

Exercise 31.10. Prove that the projective closure of every "parabola" given by $y = ax^2$ (with $a \in K$) meets the line at infinity L_∞ at exactly one point, $(0 : 1 : 0)$.

Definition 31.11. For every matrix $A = (a_{ij}) \in GL_3(K)$ we may consider the bijection (still denoted by) $A : \mathbb{P}^2 \to \mathbb{P}^2$ defined by

$$A(x_1 : x_2 : x_3) = (\sum_k a_{1k}x_k : \sum_k a_{2k}x_k : \sum_k a_{3k}x_k).$$

Call such a bijection a *projective transformation.* Two subsets of X and Y of $\mathbb{P}^2$ are called *projectively equivalent* if there exists $A \in GL_3(K)$ such that $AX = Y$.

Exercise 31.12. (Sylvester's Theorem). Prove that every projective conic is projectively equivalent to a conic given by an equation of the form

$$ax^2 + by^2 + cz^2 = 0, \quad a, b, c \in K.$$

If $K = \mathbb{C}$ one can take $a, b, c \in \{0, 1\}$; so every projective conic is, in this case, projectively equivalent to either the projective closure of a circle or to the union of two projective lines or to a projective line.

Hint: If our conic is given by a polynomial $F := ax^2 + ...$ write F in the form $F = au^2 + f(y, z)$ where $u = x + \ell(y, z)$, ℓ is a polynomial of degree 1, and f is a polynomial of degree 2. Then perform the same operation, applied to f. Finally consider the matrix A that "reflects" these substitutions. The only way this will not work is if F or f do not contain "square terms." If that is the case with F, for instance, then one takes a "mixed term" appearing in F, say xy, and writes $x = u + v$, $y = u - v$, so $xy = u^2 - v^2$. So F as a function of u, v, z contains a square and we are reduced to the preceding case. Same in case f does not contain "square terms."

Sylvester's Theorem can be generalized, by the way, to any number of variables.

CHAPTER 32

Cubics

Definition 32.1. Let K be a field in which $2 := 1+1 \neq 0$, $3 := 1+1+1 \neq 0$. Equivalently K does not contain $\mathbb{F}_2$ or $\mathbb{F}_3$. A subset $Z = Z(K) \subset K^2$ is called an *affine elliptic curve* if and only if there exist $a, b \in K$ with $4a^3 + 27b^2 \neq 0$ such that

$$Z(K) = \{(x,y) \in K^2 \mid y^2 = x^3 + ax + b\}.$$

We call $Z(K)$ the *affine elliptic curve* over K defined by the equation $y^2 = x^3 + ax + b$. Next we introduce the *elliptic curve* defined by the equation $y^2 = x^3 + ax + b$ as the set

$$E(K) = Z(K) \cup \{\infty\}$$

where ∞ is an element not belonging to $Z(K)$. We call ∞ the *point at infinity* on $E(K)$. If $(x,y) \in E(K)$ define $(x,y)' := (x,-y)$. Also define $\infty' := \infty$. Next we define a binary operation $\star$ on $E(K)$ called the *chord-tangent* operation; we will see that $E(K)$ becomes a group with respect to this operation. First define $(x,y) \star (x,-y) = \infty$, $\infty \star (x,y) = (x,y) \star \infty = (x,y)$, and $\infty \star \infty = \infty$. Also define $(x,0) \star (x,0) = \infty$. Next assume $(x_1,y_1), (x_2,y_2) \in E(K)$ with $(x_2,y_2) \neq (x_1,-y_1)$. If $(x_1,y_1) \neq (x_2,y_2)$ we let L_{12} be the unique line passing through (x_1,y_1) and (x_2,y_2). Recall that explicitly

$$L_{12} = \{(x,y) \in K^2 \mid y - y_1 = m(x - x_1)\}$$

where

$$m = (y_2 - y_1)(x_2 - x_1)^{-1}.$$

If $(x_1,y_1) = (x_2,y_2)$ we let L_{12} be the "line tangent to $Z(K)$ at (x_1,y_1)" which is by definition given by the same equation as before except now m is defined to be

$$m = (3x_1^2 + a)(2y_1)^{-1}.$$

(This definition is inspired by the definition of slope in analytic geometry.) Finally one defines

$$(x_1,y_1) \star (x_2,y_2) = (x_3,-y_3)$$

where (x_3,y_3) is the "third point of intersection of $E(K)$ with L_{12}"; more precisely (x_3,y_3) is defined by solving the system consisting of the equations defining $E(K)$ and L_{12} as follows: replacing y in $y^2 = x^3 + ax + b$ by $y_1 + m(x - x_1)$ we get a cubic equation in x:

$$(y_1 + m(x - x_1))^2 = x^3 + ax + b$$

which can be rewritten as

$$x^3 - m^2x^2 + \ldots = 0.$$

x_1, x_2 are known to be roots of this equation. We define x_3 to be the third root which is then (check!)

$$x_3 = m^2 - x_1 - x_2;$$

so we define

$$y_3 = y_1 + m(x_3 - x_1).$$

Summarizing, the definition of (x_3, y_3) is

$$(x_3, y_3) = \big((y_2 - y_1)^2(x_2 - x_1)^{-2} - x_1 - x_2, y_1 + (y_2 - y_1)(x_2 - x_1)^{-1}(x_3 - x_1)\big)$$

if $(x_1, y_1) \neq (x_2, y_2)$, $(x_1, y_1) \neq (x_2, -y_2)$ and

$$(x_3, y_3) = \big((3x_1^2 + a)^2(2y_1)^{-2} - x_1 - x_2, y_1 + (3x_1^2 + a)(2y_1)^{-1}(x_3 - x_1)\big)$$

if $(x_1, y_1) = (x_2, y_2)$, $y_1 \neq 0$.

Exercise 32.2. Prove that $E(K)$ with the above defined operations $\star$ and $'$ is an Abelian group. (N.B. Checking associativity is a very laborious exercise.)

Exercise 32.3. Consider the group $E(\mathbb{F}_{13})$ defined by the equation $y^2 = x^3 + \bar{8}$. Show that $(\bar{1}, \bar{3}), (\bar{2}, \bar{4}) \in E(\mathbb{F}_{13})$ and compute $(\bar{1}, \bar{3}) \star (\bar{2}, \bar{4})$ and $2(\bar{2}, \bar{4})$ (where the latter is, of course, $(\bar{2}, \bar{4}) \star (\bar{2}, \bar{4})$).

Affine elliptic curves are special cases of the following:

Definition 32.4. A *cubic* is a subset $X = X(K) \subset K^2$ of the form

$$X(K) = \{(x, y) \in K^2 \mid ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0\}$$

where $a, b, c, ..., j \in K$, $(a, b, c, d) \neq (0, 0, 0, 0)$. We say a cubic *passes* through a point if it contains it.

Note that, according to our definition above, a cubic may be, at the same time, a conic or a line. (For instance the cubic polynomial used to define it can be the cube of a polynomial of degree 1 in which case our cubic is, at the same time, a line; and, as in the case of conics, there other ways this can happen.)

Exercise 32.5. (Three Cubics Theorem) Prove that if two distinct cubics meet in 9 distinct points such that no 4 of the 9 points lie on a line and no 7 of the 9 points lie on a conic then every cubic that passes through 8 of the 9 points must pass through the 9th point as well.

Hint: First show that if $r \leq 8$ and r points are given then the vector space of polynomials of degree ≤ 3 vanishing at these points is strictly smaller than the vector space of polynomials of degree ≤ 3 vanishing at $r - 1$ of the r points. (In order to find, for instance, a cubic passing through $P_1, ..., P_7$ but not through P_8 one considers the cubics $C_i = Q_{1234i} \cup L_{jk}$, $\{i, j, k\} = \{5, 6, 7\}$, where Q_{1234i} is the unique conic passing through P_1, P_2, P_3, P_4, P_i and L_{jk} is the unique line through P_j and P_k. Assume C_5, C_6, C_7 all pass through P_8 and derive a contradiction as follows. Note that P_8 cannot lie on 2 of the 3 lines L_{jk} because this would force us to have 4 collinear points. So we may assume P_8 does not lie on either of the lines L_{57}, L_{67}. Hence P_8 lies on both Q_{12345} and Q_{12346}. So these conics have 5 points in common so they coincide. So this conic contains 7 points, a contradiction.) Once this is proved let $P_1, ..., P_9$ be the points of intersection of the cubics with equations F and G. We know that the space of polynomials of degree ≤ 3 vanishing at $P_1, ..., P_8$ has dimension 2 and contains F and G. So every polynomial in this space is a linear combination of F and G, hence will vanish at P_9.

EXERCISE 32.6. (Pascal's Theorem) Let $P_1, P_2, P_3, Q_1, Q_2, Q_3$ be points on a conic C. Let A_1 be the intersection of $L_{P_2Q_3}$ with $L_{P_3Q_2}$, and define A_2, A_3 similarly. (Assume the lines in question are not parallel.) Then prove that A_1, A_2, A_3 are collinear.

Hint: The cubics

$$L_{Q_1P_2} \cup L_{Q_2P_3} \cup L_{Q_3P_1} \quad \text{and} \quad L_{P_1Q_2} \cup L_{P_2Q_3} \cup L_{P_3Q_1}$$

pass through all of the following 9 points:

$$P_1, P_2, P_3, Q_1, Q_2, Q_3, A_1, A_2, A_3.$$

On the other hand the cubic $C \cup L_{A_2A_3}$ passes through all these points except possibly A_1. Then by the Three Cubics Theorem $C \cup L_{A_2A_3}$ passes through A_1. Hence $L_{A_2A_3}$ passes through A_1.

EXERCISE 32.7. Show how Pascal's Theorem implies Pappus' Theorem.

Although the above theory does not require the introduction of the projective plane $\mathbb{P}^2$ by doing so one gets a conceptually improved picture as follows:

DEFINITION 32.8. If X is a cubic as in Definition 32.4 one defines the *projective closure* $\overline{X} \subset \mathbb{P}^2$ of X by

$$\overline{X} = \{(x : y : z) \mid ax^3+bx^2y+cxy^2+dy^3+ex^2z+fxyz+gy^2z+hxz^2+iyz^2+jz^3 = 0\}.$$

Such an $\overline{X}$ is called a *projective cubic.*

EXERCISE 32.9. Let Z be an affine elliptic curve as in Definition 32.1. Prove that the projective closure $\overline{Z}$ of Z meets the line at infinity L_∞ in the point

$$(0 : 1 : 0).$$

So identifying ∞ from that Definition with the point $(0 : 1 : 0)$ we have an identification of the elliptic curve E in that Definition with $\overline{Z}$.

EXERCISE 32.10. State and prove a version of the Three Cubics Theorem for projective cubics.

REMARK 32.11. The projective version of the Three Cubics Theorem implies the associativity of the chord-tangent operation on a cubic. The idea is as follows. Let E be the elliptic curve and Q, P, R points on it different from ∞. Let

$$\begin{aligned}
\overline{L_{PQ}} \cap E &= \{P, Q, U\} \\
L_{\infty U} \cap E &= \{\infty, U, V\} \\
\overline{L_{VR}} \cap E &= \{V, R, W\} \\
\overline{L_{PR}} \cap E &= \{P, R, X\} \\
L_{\infty X} \cap E &= \{\infty, X, Y\}.
\end{aligned}$$

Here $L_{\infty A}$ is the "vertical" line passing through a point A to which we add the point ∞. Note that

$$Q \star P = V, \quad V \star R = W', \quad P \star R = Y.$$

We want to show that

$$(Q \star P) \star R = Q \star (P \star R).$$

This is equivalent to

$$V \star R = Q \star Y$$

i.e., that

$$W' = Q \star Y$$

i.e., that Q, Y, W are collinear. Now the two cubics

$$E \quad \text{and} \quad \overline{L_{PQ}} \cup \overline{L_{WR}} \cup \overline{L_{YX}}$$

both pass through the 9 points

$$P, Q, R, U, V, W, X, Y, \infty.$$

On the other hand the cubic

$$\Gamma = \overline{L_{UV}} \cup \overline{L_{PR}} \cup \overline{L_{QY}}$$

passes through all 9 points except W. By the Three Cubics Theorem (with the hypotheses of that theorem satisfied) we get that Γ passes through W hence $\overline{L_{QY}}$ passes through W. The above argument only applies when one avoids the corresponding "degeneracies," e.g., tangencies or "4 points on a line" or "7 points on a conic." To conclude that associativity holds in the "degenerate" cases as well one needs to use arguments from algebraic geometry (the fact that "two morphisms of algebraic varieties that coincide generically must coincide").

CHAPTER 33

Limits

We start discussing now some simple topics in *analysis* (which is just another name for *calculus* the way it is understood in modern Mathematics). Analysis is the study of functions with special emphasis on their "infinitesimal behavior." Euler referred to this subject as the "analysis of the infinity" (which enters the title of one of his books). The key words here are sequences, convergence, limits, differential calculus, and integral calculus. Here we will discuss limits. Analysis emerged through work of Abel, Cauchy, Riemann, and Weierstrass, as a clarification of the early calculus of Newton, Leibniz, Euler, and Lagrange.

Recall the following:

DEFINITION 33.1. A *sequence* in $\mathbb{R}$ is a map $F : \mathbb{N} \to \mathbb{R}$; if $F(n) = a_n$ we denote the sequence by $a_1, a_2, a_3, \ldots$ or by (a_n). We let $F(\mathbb{N})$ be denoted by $\{a_n \mid n \geq 1\}$; the latter is a subset of $\mathbb{R}$.

DEFINITION 33.2. A *subsequence* of a sequence $F : \mathbb{N} \to \mathbb{R}$ is a sequence of the form $F \circ G$ where $G : \mathbb{N} \to \mathbb{N}$ is a strictly increasing map. If $a_1, a_2, a_3, \ldots$ is F then $F \circ G$ is $a_{k_1}, a_{k_2}, a_{k_3}, \ldots$ (or (a_{k_n})) where $G(n) = k_n$.

DEFINITION 33.3. A sequence (a_n) is *convergent* to $a_0 \in \mathbb{R}$ if and only if for every real number $\epsilon > 0$ there exists an integer N such that for all $n \geq N$ we have $|a_n - a_0| < \epsilon$. We write

$$a_n \to a_0 \quad \text{or} \quad \lim a_n = a_0 \quad \text{or} \quad \lim_{n \to \infty} a_n = a_0$$

and we say a_0 is the limit of (a_n). A sequence is called *convergent* if and only if there exists $a \in \mathbb{R}$ such that the sequence converges to a. A sequence is called *divergent* if and only if it is not convergent.

EXERCISE 33.4. Prove that $a_n = \frac{1}{n}$ converges to 0.

Hint: Let $\epsilon > 0$; we need to find N such that for all $n \geq N$ we have

$$|\frac{1}{n} - 0| < \epsilon;$$

it is enough to take N to be any integer such that $N > \frac{1}{\epsilon}$.

EXERCISE 33.5. Prove that $a_n = \frac{1}{\sqrt{n}}$ converges to 0.

EXERCISE 33.6. Prove that $a_n = \frac{1}{n^2}$ converges to 0.

EXERCISE 33.7. Prove that $a_n = n$ is divergent.

EXERCISE 33.8. Prove that $a_n = (-1)^n$ is divergent.

EXERCISE 33.9. Prove that if $a_n \to a_0$ and $b_n \to b_0$ then

1) $a_n + b_n \to a_0 + b_0$

2) $a_n b_n \to a_0 b_0$.

If in addition $b_0 \neq 0$ then there exists N such that for all $n \geq N$ we have $b_n \neq 0$; moreover if $b_n \neq 0$ for all n then

3) $\frac{a_n}{b_n} \to \frac{a_0}{b_0}$.

Hint for 1): Consider any $\epsilon > 0$. Since $a_n \to a_0$ there exists N_a such that for all $n \geq N_a$ we have $|a_n - a_0| < \frac{\epsilon}{2}$. Since $b_n \to b_0$ there exists N_b such that for all $n \geq N_b$ we have $|b_n - b_0| < \frac{\epsilon}{2}$. Let $N = \max\{N_a, N_b\}$ be the maximum between N_a and N_b. Then for all $n \geq N$ we have

$$|(a_n + b_n) - (a_0 + b_0)| \leq |a_n - a_0| + |b_n - b_0| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

EXERCISE 33.10. Prove that if $a_n \to a$, $b_n \to b$, and $a_n \leq b_n$ for all $n \geq 1$ then $a \leq b$.

EXERCISE 33.11. Prove that a subset A of $\mathbb{R}$ is closed if and only if for all convergent sequences a_n in $\mathbb{R}$ if $a_n \in A$ for all n and if $a_n \to a$ then $a \in A$.

DEFINITION 33.12. A sequence F is *bounded* if and only if the set $F(\mathbb{N}) \subset \mathbb{R}$ is bounded. In other words a sequence (a_n) is bounded if and only if there exists $m, M \in \mathbb{R}$ such that for all n we have $m \leq a_n \leq M$.

DEFINITION 33.13. A sequence F is (strictly) *increasing* if and only if F is (strictly) increasing. A sequence F is (strictly) *decreasing* if and only if F is (strictly) decreasing. In other words a sequence (a_n) is increasing if for all $n \in \mathbb{N}$ we have $a_n \leq a_{n+1}$; and a sequence (a_n) is decreasing if for all $n \in \mathbb{N}$ we have $a_n \geq a_{n+1}$. Similarly the strict version of these.

DEFINITION 33.14. A sequence (a_n) is *Cauchy* if and only if for every real $\epsilon > 0$ there exists an integer N such that for all integers $m, n \geq N$ we have $|a_n - a_m| < \epsilon$.

EXERCISE 33.15. Prove that every convergent sequence is Cauchy.

EXERCISE 33.16. Prove the following statements in the prescribed order:

1) Every Cauchy sequence is bounded.

2) Every bounded sequence contains a sequence which is either increasing or decreasing.

3) Every bounded sequence which is either increasing or decreasing is convergent.

4) Every Cauchy sequence which contains a convergent subsequence is itself convergent.

5) Every Cauchy sequence is convergent.

Hints: For 1) let $\epsilon = 1$, let N correspond to this ϵ, and get that $|a_n - a_N| < 1$ for all $n \geq N$; conclude from here. For 2) consider the sets $A_n = \{a_m \mid m \geq n\}$. If at least one of these sets has no maximal element we get a strictly increasing subsequence by Proposition 19.7. If each A_n has a maximal element b_n then $b_n = a_{k_n}$ for some k_n and the subsequence a_{k_n} is decreasing. For 3) we view each $a_n \in \mathbb{R}$ as a Dedekind cut i.e., as a subset $a_n \subset \mathbb{Q}$; the limit will be either the union of the sets a_n or the intersection (possibly with its minimum removed). Statement 4) is easy. Statement 5) follows by combining the previous statements.

Definition 33.17. A function $F : I \to \mathbb{R}$ is *continuous at a point* a_0 of a (closed or open) interval I if and only if for every sequence (a_n) in I converging to a_0 we have that the sequence $(F(a_n))$ converges to $F(a_0)$.

Exercise 33.18. (ϵ and δ criterion for continuity). Prove that a function $F : I \to \mathbb{R}$ is continuous at $a_0 \in I$ if and only if for every real $\epsilon > 0$ there exists a real $\delta > 0$ such that for every $a \in I$ with $|a - a_0| < \delta$ we have $|F(a) - F(a_0)| < \epsilon$.

Exercise 33.19. Prove that a function $F : I \to \mathbb{R}$ is continuous (for the Euclidean topology on both the source and the target) if and only if it is continuous at every point of I.

Exercise 33.20. Prove that every polynomial function $f : \mathbb{R} \to \mathbb{R}$ (i.e., every function of the form $a \mapsto f(a)$ where f is a polynomial) is continuous.

Exercise 33.21. Prove that $\mathbb{R}$ with the Euclidean topology is connected.

Hint: Assume $\mathbb{R} = A \cup B$ with A, B open, non-empty, and disjoint, and seek a contradiction. Let $a \in A$ and $b \in B$. Assume $a \leq b$; the case $b \leq a$ is similar. Show that there exists sequences (a_n) and (b_n), the first increasing, the second decreasing, with $a_n \leq b_n$ and $b_n - a_n \to 0$. (To check this use recursion to define a_{n+1}, b_{n+1} in terms of a_n, b_n by the following rule: if $c_n = \frac{a_n+b_n}{2}$ then set $a_{n+1} = c_n$ and $b_{n+1} = b_n$ in case $c_n \in A$; and set $a_{n+1} = a_n$ and $b_{n+1} = c_n$ in case $c_n \in B$.) Note that $a_n \to a_0$ and $b_n \to b_0$ and $a_0 = b_0$. Since A, B are open and disjoint they are closed. So $a_0 \in A$ and $b_0 \in B$. But this contradicts the fact that A and B are disjoint.

Exercise 33.22. Prove that if $A \subset \mathbb{R}$ is a connected subset if and only if one of the following holds:
i) A is a closed interval;
ii) A is obtained from a closed interval by removing one or both ends;
iii) A is a set of the form

$$(-\infty, a] := \{x \in \mathbb{R} \mid x \leq a\} \quad \text{or} \quad [a, \infty) := \{x \in \mathbb{R} \mid x \geq a\};$$

iv) A is obtained from a set as in iii) by removing a.

Exercise 33.23. (Heine-Borel Theorem) Prove that every closed interval in $\mathbb{R}$ is compact.

Hint: Assume $[a, b]$ is not compact and derive a contradiction as follows. We know $[a, b]$ has an open covering $(U_i)_{i \in I}$ that does not have a finite open subcovering. Show that there exists sequences (a_n) and (b_n), the first increasing, the second decreasing, with $a_n \leq b_n$ and $b_n - a_n \to 0$, such that $[a_n, b_n]$ cannot be covered by finitely many U_is. (To check this use recursion to define a_{n+1}, b_{n+1} in terms of a_n, b_n by the following rule: let $c_n = \frac{a_n+b_n}{2}$; then at least one of the two intervals $[a_n, c_n]$ or $[c_n, b_n]$ cannot be covered by finitely many U_is; if this is the case with the first interval then set $a_{n+1} = a_n$ and $b_{n+1} = c_n$; in the other case set $a_{n+1} = c_n$ and $b_{n+1} = b_n$.) Note that $a_n \to a_0$ and $b_n \to b_0$ and $a_0 = b_0$. But $a_0 = b_0$ is in one of the U_is; this U_i will completely contain one of the intervals $[a_n, b_n]$ which is a contradiction.

Exercise 33.24. Prove that a subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

Hint: Use the Borel-Heine theorem.

EXERCISE 33.25. Prove that if $F : [a, b] \to \mathbb{R}$ is continuous then there exists $c \in [a, b]$ such that $F(c) \geq F(x)$ for all $x \in [a, b]$. (We say that F *attains its maximum* at c.)

Hint: By the Heine-Borel theorem $[a, b]$ is compact. Since F is continuous, $F([a, b])$ is compact. Since $\mathbb{R}$ is Hausdorff, $F([a, b])$ is closed and bounded. So its supremum is contained in it.

EXERCISE 33.26. (Darboux property). Prove that if $F : [a, b] \to \mathbb{R}$ is continuous and $F(a) \leq \gamma \leq F(b)$ then there exists $c \in [a, b]$ such that $F(c) = \gamma$.

Hint: Since $[a, b]$ is connected and F is continuous, $F([a, b])$ is connected. If there is no $c \in [a, b]$ such that $F(c) = \gamma$ one can cover $F([a, b])$ by the intervals defined by $x < \gamma$ and $x > \gamma$, a contradiction.

EXERCISE 33.27. Prove that every polynomial with real coefficients and odd degree has a root in $\mathbb{R}$.

Hint: Use the Darboux property.

EXERCISE 33.28. Prove that if $F : [a, b] \to \mathbb{R}$ is continuous then for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in [a, b]$ if $|x - y| < \delta$ then $|F(x) - F(y)| < \epsilon$. (The latter property is called *uniform continuity*.)

Hint: Assume the above is false for some ϵ. The there exist sequences a_n, b_n such that $|a_n - b_n| < \frac{1}{n}$ and $|F(a_n) - F(b_n)| \geq \epsilon$. Replace a_n by a convergent subsequence and further replace b_n by a convergent subsequence. Then $a_n \to \alpha$, $b_n \to \beta$, $\alpha, \beta \in [a, b]$. One proves that $\alpha = \beta$. One can find n such that $|F(a_n) - F(\alpha)| < \epsilon/2$, $|F(b_n) - F(\alpha)| < \epsilon/2$. One gets $|F(a_n) - F(b_n)| < \epsilon$, a contradiction.

DEFINITION 33.29. A sequence (f_n) of functions $f_n : I \to \mathbb{R}$ on an (open or closed) interval is called *uniformly convergent* to a function $f : I \to \mathbb{R}$ if and only if for all $\epsilon > 0$ there exists $N \geq 1$ such that for all $n \geq N$ and all $x \in I$ we have

$$|f_n(x) - f(x)| < \epsilon.$$

DEFINITION 33.30. A sequence (f_n) of functions $f_n : I \to \mathbb{R}$ on an (open or closed) interval is called *uniformly Cauchy* if and only if for all $\epsilon > 0$ there exists $N \geq 1$ such that for all $n, m \geq N$ and all $x \in I$ we have

$$|f_n(x) - f_m(x)| < \epsilon.$$

EXERCISE 33.31.

1) Prove that if (f_n) is uniformly convergent to some f then (f_n) is uniformly Cauchy.

2) Prove that if (f_n) is uniformly Cauchy then (f_n) is uniformly convergent to some f.

3) Prove that if Prove that if (f_n) is uniformly convergent to f and if each f_n is continuous then f is continuous.

Hint for 2): For each $x \in I$ the sequence $(f_n(x))$ is Cauchy so convergent and let $f(x)$ be its limit. Now let $\epsilon > 0$. There exists N such that $|f_n(x) - f_m(x)| < \epsilon/2$ for all $n, m \geq N$ and all $x \in I$. Taking $m \to \infty$ in the latter we get $|f_n(x) - f(x)| \leq \epsilon/2 < \epsilon$ for all $n \geq N$ and all $x \in I$.

Hint for 3): Fix $x_0 \in I$ and $x_n \to x_0$. Let $\epsilon > 0$. There exists N such that for all $n \geq N$ and all $x \in I$ we have $|f_n(x) - f(x)| < \epsilon/3$; in particular $|f_N(x) - f(x)| < \epsilon/3$

for all $x \in I$. Since f_N is continuous there exists M such that for all $m \geq M$ we have $|f_N(x_m) - f_N(x_0)| < \epsilon/3$. Then for all $m \geq M$ we have

$$\begin{aligned} |f(x_m) - f(x_0)| &\leq |f(x_m) - f_N(x_m)| + |f_N(x_m) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

EXERCISE 33.32. Prove that (f_n) on $[0,1]$ defined by $f_n(x) = x^n$ is not uniformly Cauchy.

Most of the content of this Chapter can be generalized to the case of functions of several variables. For example we have:

DEFINITION 33.33. A sequence $(a_1, b_1), (a_2, b_2), \ldots$ in $\mathbb{R}^2$ is said to *converge* to (a_0, b_0) (and we write $(a_n, b_n) \to (a_0, b_0)$) if and only if $a_n \to a_0$ and $b_n \to b_0$.

DEFINITION 33.34. A function $F : I \times J \to \mathbb{R}$ is *continuous at a point* $(a_0, b_0) \in I \times J$ (where I, J are closed or open intervals) if and only if for every sequence (a_n, b_n) in $I \times J$ converging to (a_0, b_0) we have that the sequence $(F(a_n, b_n))$ converges to $F(a_0, b_0)$.

EXERCISE 33.35. (ϵ and δ criterion for continuity). Prove that a function $F : I \times J \to \mathbb{R}$ is continuous at $(a_0, b_0) \in I \times J$ if and only if for every real $\epsilon > 0$ there exists a real $\delta > 0$ such that for every $a \in I$ and $b \in J$ with $|a - a_0| < \delta$ and $|b - b_0| < \delta$ we have $|F(a, b) - F(a_0, b_0)| < \epsilon$.

EXERCISE 33.36. Prove that a function $F : I \times J \to \mathbb{R}$ is continuous (for the Euclidean topology on both the source and the target) if and only if it is continuous at every point of $I \times J$.

EXERCISE 33.37. Prove that a function $U \to \mathbb{R}^m$, where $U \subset \mathbb{R}^m$ is open, is continuous if and only if its components are continuous.

In what follows we discuss series.

DEFINITION 33.38. Let (a_n) be a sequence and $s_n = \sum_{k=1}^n a_k$. The sequence (s_n) is called the sequence of *partial sums*. If (s_n) is convergent to some s we say $\sum_{n=1}^\infty a_n$ is a *convergent series* and that this series *converges* to s; we write

$$\sum_{k=1}^\infty a_n = s$$

or simply

$$\sum a_n = s.$$

If the sequence (s_n) is divergent we say that $\sum_{n=1}^\infty a_n$ is a *divergent series*.

EXERCISE 33.39. Prove that

$$\sum_{n=1}^\infty \frac{1}{n(n+1)} = 1.$$

Hint: Start with the equality

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$$

and compute

$$\sum_{n=1}^{N} \frac{1}{n(n+1)} = 1 - \frac{1}{N}.$$

EXERCISE 33.40. Prove that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

is convergent.

Hint: Prove the sequence of partial sums is bounded using the inequality

$$\frac{1}{n^2} \leq \frac{1}{n(n+1)}$$

plus Exercise 33.39.

EXERCISE 33.41. Prove that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^k}$$

is convergent for $k \geq 3$.

EXERCISE 33.42. Prove that the series

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

is divergent. This series is called the *harmonic series*.

Hint: Assume the series is convergent. Then the sequence of partial sums is convergent hence Cauchy. Get a contradiction from the inequality:

$$\frac{1}{2^n+1} + \frac{1}{2^n+2} + \frac{1}{2^n+3} + \ldots + \frac{1}{2^n+2^n} > 2^n \times \frac{1}{2^n+2^n} = \frac{1}{2}.$$

EXERCISE 33.43. Prove that $a^n \to 0$ if $|a| < 1$.

Hint: We may assume $0 < a < 1$. Note that (a^n) is strictly decreasing. Since it is bounded it is convergent. Let α be its limit. Assume $\alpha \neq 0$ and get a contradiction by noting that

$$\frac{1}{a} = \frac{a^n}{a^{n+1}} \to \frac{\alpha}{\alpha} = 1.$$

EXERCISE 33.44. Prove that

$$\sum_{n=1}^{\infty} a^n = \frac{1}{1-a}$$

if $|a| < 1$.

EXERCISE 33.45. Let (a_n) be a sequence and assume there exist real numbers $M > 0$ and $0 < c < 1$ such that for all $n \geq 1$,

$$|a_n| \leq M \cdot c^n.$$

Prove that the series $\sum a_n$ is convergent.

Hint: It is enough to prove the sequence of partial sums is Cauchy.

EXERCISE 33.46. (Quotient criterion). Let (a_n) be a sequence and assume

$$\lim \frac{|a_{n+1}|}{|a_n|} = b < 1.$$

Prove that the series $\sum a_n$ is convergent.

Hint: There exist $N \geq 1$ and $c \in (b, 1)$ such that $|a_{n+1}| \leq c \cdot |a_n|$ for $n \geq N$. Hence $|a_{n+k}| \leq c^k \cdot |a_n|$ for $n \geq N$ and $k \geq 1$. The result follows the previous Exercise.

EXERCISE 33.47. Prove that the series

$$\sum_{n=0}^{\infty} \frac{a^n}{n!}$$

is convergent for all $a \in \mathbb{R}$; its limit is denoted by $e^a = \exp(a)$; $e = e^1$ is called the *Euler number*; the map

$$\mathbb{R} \to \mathbb{R}, \quad a \mapsto \exp(a)$$

is called the *exponential map*. Prove that

$$\exp(a+b) = \exp(a)\exp(b).$$

Hint: Use Exercise 33.46.

EXERCISE 33.48. Prove that the function $\exp : \mathbb{R} \to \mathbb{R}$ is continuous.

Hint: By Exercise 33.31 it is enough to show that the sequence (f_n) defined by

$$f_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$$

is uniformly Cauchy. This can be shown using Exercise 33.54.

EXERCISE 33.49. Let $S \subset Fun(\mathbb{N}, \{0,1\})$ be the set of all sequences (a_n) such that there exist N with $a_n = 1$ for all $n \geq N$. Prove that the map

$$\{0,1\}^{\mathbb{N}} \backslash S \to \mathbb{R}, \quad (a_n) \mapsto \sum_{n=1}^{\infty} \frac{a_n}{2^n}$$

is (well defined and) injective. Conclude that $\mathbb{R}$ is uncountable.

EXERCISE 33.50. Prove that there exist transcendental numbers in $\mathbb{R}$.

Hint: $\mathbb{R}$ is uncountable whereas the set of algebraic numbers is countable; cf. Exercise 38.5. This is Cantor's proof of existence of transcendental numbers.

Real analysis (analysis of sequences, continuity, and other concepts of calculus like differentiation and integration of functions on $\mathbb{R}$) can be extended to complex analysis. Indeed we have:

EXERCISE 33.51. Identifying $\mathbb{C}$ with $\mathbb{R}^2$ recall that if (z_n) is a sequence in $\mathbb{C}$ and $z_n = a_n + b_n i$, $a_n, b_n \in \mathbb{R}$, $z_0 = a_0 + b_0 i$, then, by definition, $z_n \to z_0$ if and only if $a_n \to a_0$ and $b_n \to b_0$. Prove that a sequence (z_n) in $\mathbb{C}$ is convergent to $z_0 \in \mathbb{C}$ if and only if for every real number $\epsilon > 0$ there exists an integer N such that for all $n \geq N$ we have $|z_n - z_0| < \epsilon$. We write $z_n \to z_0$ and we say z_0 is the limit of (z_n).

Definition 33.52. A sequence (z_n) in $\mathbb{C}$ is *Cauchy* if and only if for every real $\epsilon > 0$ there exists an integer N such that for all integers $m, n \geq N$ we have $|z_n - z_m| < \epsilon$.

Exercise 33.53. Prove that a sequence in $\mathbb{C}$ is convergent if and only if it is Cauchy.

Exercise 33.54. Prove that:

1) For all $z \in \mathbb{C}$ the series

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

is convergent (by which we mean that the sequence of partial sums is convergent); its limit is denoted by $e^z = \exp(z)$.

2) $\exp(z+w) = \exp(z)\exp(w)$ for all $z, w \in \mathbb{C}$.

3) $\overline{\exp(z)} = \exp(\overline{z})$ for all $z \in \mathbb{C}$.

4) $|\exp(it)| = 1$ for all $t \in \mathbb{R}$.

5) The map

$$\mathbb{C} \to \mathbb{C}, \quad z \mapsto \exp(z),$$

is continuous. This map is called the (complex) *exponential map*.

There is a version of the above theory in what is referred to as *p-adic analysis* (which is crucial to number theory). Recall from Definition 20.31 the ring of p-adic numbers $\mathbb{Z}_p$ whose elements are denoted by $\alpha = [a_n]$ and whose norm is denoted by $|\ |$.

Definition 33.55. Consider a sequence $\alpha_1, \alpha_2, \alpha_3, \ldots$ in $\mathbb{Z}_p$.

1) $\alpha_1, \alpha_2, \alpha_3, \ldots$ is called a *Cauchy sequence* if and only if for every real (or, equivalently, rational) $\epsilon > 0$ there exists an integer N such that for all $m, m' \geq N$ we have $|\alpha_m - \alpha_{m'}| \leq \epsilon$.

2) We say that $\alpha_1, \alpha_2, \alpha_3, \ldots$ *converges* to some $\alpha_0 \in \mathbb{Z}_p$ if and only if for every real (or, equivalently, rational) $\epsilon > 0$ there exists an integer N such that for all $m \geq N$ we have $|\alpha_m - \alpha_0| \leq \epsilon$. We say α_0 is the *limit* of (α_n) and we write $\alpha_n \to \alpha_0$.

Exercise 33.56. Prove that a sequence in $\mathbb{Z}_p$ is convergent if and only if it is Cauchy.

Exercise 33.57. Prove that $\mathbb{Z}_p$ is compact in the p-adic topology.

The following is in deep contrast with the case of $\mathbb{R}$ or $\mathbb{C}$:

Exercise 33.58. Prove that if (α_n) is a sequence in $\mathbb{Z}_p$ with $\alpha_n \to 0$ then the series $\sum_{n=1}^{\infty} \alpha_n$ is convergent in $\mathbb{Z}_p$ (by which we mean that the sequence of partial sums is convergent).

Exercise 33.59.

1) Prove that $\sum_{n=1}^{\infty} p^{n-1}$ is the inverse of $1-p$ in $\mathbb{Z}_p$.

2) Prove that if $\alpha \in \mathbb{Z}_p$ has $|\alpha| = 1$ then α is invertible in $\mathbb{Z}_p$.

3) Prove that for all $n \geq 1$ and all $a \in \mathbb{Z}_p$ with $|a| < 1$ there exists an element of $\mathbb{Z}_p$ denoted by $\frac{a^n}{n!}$ such that $(n!) \cdot \frac{a^n}{n!} = a^n$.

4) Prove that

$$\sum_{n=1}^{\infty} \frac{a^n}{n!}$$

is convergent in $\mathbb{Z}_p$ for all $a \in \mathbb{Z}_p$ with $|a| < 1$. One denotes the limit by $\exp_p(a)$.

Hint for 3): Use the fact that if p does not divide an integer $a \in \mathbb{Z}$ then there exist integers $m, n \in \mathbb{A}$ such that $ma + np = 1$; then use 1) above.

For the next Exercise recall the derivative a polynomial introduced in Definition 28.28.

EXERCISE 33.60. (Hensel's Lemma). Let $a_1 \in \mathbb{Z}_p$ and let $f \in \mathbb{Z}_p[x]$ be a polynomial with coefficients in $\mathbb{Z}_p$. Assume that $|f(a_1)| < 1$ and $|f'(a_1)| = 1$. Prove that there exists a unique $a \in \mathbb{Z}_p$ such that $|a - a_1| < 1$ and $f(a) = 0$.

Hint: Prove by recursion that there exists a sequence (a_n) such that

$$|a_{n+1} - a_n| \leq p^{-n} \quad \text{and} \quad |f(a_n)| \leq p^{-n}$$

for all n. To prove this set $f(a_n) = p^n c_n$ with $c_n \in \mathbb{Z}_p$ and define $a_{n+1} = a_n + p^n b_n$ where $b_n \in \mathbb{Z}_p$ is chosen such that $c_n + b_n f'(a_n) = p d_n$ with $d_n \in \mathbb{Z}_p$. Then set $a = \lim a_n$. This shows the existence part. To prove uniqueness prove that for any two roots as in Hensel's Lemma the norm of their difference is $\leq p^{-n}$ for all n.

EXERCISE 33.61.
1) Prove that the polynomial $x^{p-1} - 1$ has $p - 1$ roots in $\mathbb{Z}_p$.
2) Prove that $x^p - 1$ has only one root, 1, in $\mathbb{Z}_p$.

Hint for 1): Use Hensel's Lemma and Fermat's Little Theorem.
Hint for 2): Prove that if $a^p = 1$ then $|a - 1| \leq p^{-n}$ for all n.

CHAPTER 34

Trigonometry

Trigonometry arose long before calculus mainly motivated by geometry and astronomy. A rigorous approach to trigonometry requires some elements of analysis that we already covered and hence can be used in what follows. We will define the functions sin and cos and also the number π.

Definition 34.1. For all $t \in \mathbb{R}$ define $\cos t, \sin t \in \mathbb{R}$ as being the unique real numbers such that

$$\exp(it) = \cos t + i \sin t.$$

(This is called *Euler's formula* but here this is a definition and not a theorem.)

Exercise 34.2. Prove the following equalities:
1) $\cos(t_1 + t_2) = \cos t_1 \cos t_2 - \sin t_1 \sin t_2$;
2) $\sin(t_1 + t_2) = \sin t_1 \cos t_2 + \cos t_1 \sin t_2$.

Exercise 34.3. Prove that the map $f : \mathbb{R} \to SO_2(\mathbb{R})$ defined by

$$f(t) = \left(\begin{array}{cc} \cot t & \sin t \\ -\sin t & \cos t \end{array} \right)$$

is a group homomorphism.

Exercise 34.4. Prove that if H is a closed subgroup of $\mathbb{R}$ and $H \neq \mathbb{R}$, $H \neq \{0\}$ then there exists a unique $T \in \mathbb{R}$, $T > 0$, such that

$$H = \{nT \mid n \in \mathbb{Z}\}.$$

Hint: One first shows that if T is the infimum of the set

$$\{a \in H \mid a > 0\}$$

then $T \neq 0$. In order to check this assume $T = 0$ and seek a contradiction. Indeed from $T = 0$ we get that there exists a sequence (a_n) with $a_n \in H$, and $a_n \to 0$. Deduce from this and the fact that H is closed that $G = \mathbb{R}$, a contradiction. Finally one shows that $H = \{nT \mid n \in \mathbb{Z}\}$ using an argument similar to the one used to prove Proposition 20.6.

Exercise 34.5.
1) Prove that the "circle"

$$S^1 := \{z \in \mathbb{C}^\times \mid |z| = 1\}$$

is a subgroup of the group $\mathbb{C}^\times := \mathbb{C} \setminus \{0\}$ under multiplication.
2) Prove that if $K \neq \{1\}$ is a connected subgroup of S^1 then $K = S^1$.

Hint: For 2) use the topological fact that a connected subset of S^1 is either a "segment" of S^1 or S^1 minus a point or S^1 itself (this can be shown by using "stereographic projections" of S^1 minus a point into lines) plus the algebraic fact that for all $z \in S^1$ we have $|z^{n+1} - z^n| = |z - 1|$.

Exercise 34.6. Prove that the map

$$F : \mathbb{R} \to S^1, F(t) = \exp(it)$$

is surjective and there exists a unique real number $\pi \in \mathbb{R}$, $\pi > 0$, such that

$$Ker\ F = \{2n\pi \mid n \in \mathbb{Z}\}.$$

Hint: Since F is continuous Im F is a connected non-trivial subgroup of S^1 so it is S^1 by Exercise 34.5. Also Ker F is a closed subgroup of $\mathbb{R}$ different from $\mathbb{R}$ so we may conclude by Exercise 34.4.

Exercise 34.7. Prove that

$$\exp(\pi i) + 1 = 0.$$

Hint: Note that $(\exp(\pi i))^2 = \exp(2\pi i) = 1$.

Remark 34.8. This is a celebrated formula of Euler. For us, however, this formula is a trivial consequence of our definition of π.

Exercise 34.9. (de Moivre). For $n \in \mathbb{N}$, $n \geq 3$, set

$$\zeta_n := \exp\left(\frac{2\pi i}{n}\right).$$

Prove that the roots of the polynomial $x^n - 1$ in $\mathbb{C}$ are the following *roots of unity*:

$$1, \zeta, \zeta^2, ..., \zeta^{n-1}.$$

More generally prove that for all $a \in \mathbb{C}$ the equation $z^n = a$ has n roots in $\mathbb{C}$.

Hint: For the second assertion use the surjectivity of the map $\mathbb{R} \to S^1$, $t \mapsto \exp(it)$ to show that every complex number has the form $r \cdot \exp(it)$ for some real r and t.

Exercise 34.10. Recall the Fundamental Theorem of Algebra (Theorem 28.14) saying that every non-constant polynomial $P(x) \in \mathbb{C}[x]$ has a root in $\mathbb{C}$. Give a proof of this theorem following the following steps. (This is essentially d'Alembert's proof. Gauss criticized it for lack of rigor and gave other proofs. However the argument of d'Alembert can be made rigorous as indicated below.)

1) Show that for every sequence (z_n) of complex numbers such that $|z_n| \to \infty$ we have $|P(z_n)| \to \infty$. Conclude that the function $z \mapsto |P(z)|$ must attain its minimum at some point in $\mathbb{C}$.

2) (d'Alembert's Lemma) Prove that if $P(z_0) \neq 0$ for some $z_0 \in \mathbb{C}$ then every disk with center z_0 contains a point z_1 such that $|P(z_1)| < |P(z_0)|$.

Clearly 1) and 2) prove the theorem. To prove 2) we may assume $z_0 = 0$. Write

$$P(z) = a_0 + a_k z^k + a_{k+1} z^{k+1} + ... + a_d z^d$$

with $a_0 \neq 0$ and $a_k \neq 0$. Let $\epsilon \in \mathbb{R}_{>0}$ and denote by $\epsilon^{\frac{a}{b}}$ the positive root of $x^b = \epsilon^a$. Let $w \in \mathbb{C}$ be such that $w^k = -\frac{a_0}{a_k}$ (which exists by de Moivre, see the previous exercise). Setting $z = z_1 := \epsilon^{\frac{1}{k}} w$ we get

$$P(z_1) = (1-\epsilon)a_0 + \epsilon^{\frac{k}{k+1}} M(\epsilon)$$

with $M(\epsilon)$ bounded for ϵ close to 0. Then for sufficiently small ϵ we have

$$|P(z_1)| \leq (1-\epsilon)|a_0| + \epsilon^{\frac{k}{k+1}} \cdot M(\epsilon) \leq (1 - \frac{\epsilon}{2})|a_0| < |P(0)|.$$

CHAPTER 35

Differentiation

Calculus was invented by Newton and Leibniz, motivated by problems in mechanics and analytic geometry. The main concepts of calculus are differentiation and integration.

Definition 35.1. Let $F : I \to \mathbb{R}$ be a map, where $I \subset \mathbb{R}$ is an open interval, and $a_0 \in I$. We say F is *differentiable* at a_0 if and only if there exists a real number (denoted by) $F'(a_0) \in \mathbb{R}$ such that for every sequence $a_n \to a_0$ with $a_n \neq a_0$ we have

$$\frac{F(a_n) - F(a_0)}{a_n - a_0} \to F'(a_0).$$

Exercise 35.2. Prove that if F is differentiable at $a_0 \in I$ then it is continuous at a_0.

Exercise 35.3. Prove that if F is a constant function (i.e., $F(x) = F(y)$ for all $x, y \in I$) then F is differentiable at every a and $F'(a) = 0$.

Definition 35.4. We say $F : I \to \mathbb{R}$ is *differentiable* on an open interval I if and only if F is differentiable at every $a \in I$. If this is the case the map $a \mapsto F'(a)$ is called the *derivative* of F and is denoted by $DF = F' = \frac{dF}{dx} : I \to \mathbb{R}$. We say F is *k times differentiable* ($k \geq 2$) if and only if it is $k-1$ times differentiable and $D^{k-1}F$ is differentiable; one writes $D^kF = \frac{d^kF}{dx^k} = F^{(k)}$. We also write $F'' = D^2F$. We say that F is of *class C^k* if and only if it is k times differentiable and $F^{(k)}$ is continuous. We say that F is *infinitely differentiable* (or *smooth*) if and only if it is k times differentiable for every $k \in \mathbb{N}$. One denotes by $C^k(I)$ and $C^\infty(I)$ the set of n times differentiable, respectively smooth functions.

Sometimes it is useful to extend the concept of class C^1 to closed intervals:

Definition 35.5. We say that $F : [a, b] \to \mathbb{R}$ is of class C^1 if it of class C^1 on (a, b) and F' can be extended to a continuous function on $[a, b]$. (The extension is then unique still denoted by F'.)

Exercise 35.6. Prove that for every F, G differentiable on an open interval I we have $F + G, F \cdot G$ are differentiable on I and

1) $D(F + G) = D(F) + D(G)$ (additivity);

2) $D(F \cdot G) = F \cdot D(G) + G \cdot D(F)$ (Leibniz rule);

here $F + G, F \cdot G$ are the pointwise addition and multiplication of F and G. In particular $C^\infty(I)$ is a ring with respect to $+$ and $\cdot$; 0 and 1 are the functions $0(x) = 0$ and $1(x) = 1$.

Exercise 35.7. Prove that every polynomial function $F : \mathbb{R} \to \mathbb{R}$ is smooth and

$$F(x) = \sum_{k=0}^{n} a_n x^n \Rightarrow F'(x) = \sum_{k=0}^{n} n a_n x^{n-1}.$$

Hint: It is enough to look at $F(x) = x^k$. In this case

$$\frac{a_n^k - a_0^k}{a_n - a_0} = a_n^{k-1} + a_n^{k-2}a_0 + ... + a_0^{k-1} \to k a_0^{k-1}.$$

Exercise 35.8. Prove that the function $F : \mathbb{R}_{>0} \to \mathbb{R}$ defined by $F(x) = \sqrt{x}$ is smooth and $F'(x) = \frac{1}{2\sqrt{x}}$.

Exercise 35.9. Prove that $F(x) = \exp(x)$ is differentiable and $F'(x) = \exp(x)$. Hence F is smooth.

Exercise 35.10. Prove that $F(x) = \sin(x)$ is differentiable and $F'(x) = \cos(x)$. Prove that $G(x) = \cos(x)$ is differentiable and $G'(x) = -\sin(x)$. Hence F and G are smooth.

Exercise 35.11. Let $F : [a, b] \to \mathbb{R}$ be continuous and assume F is differentiable on (a, b). Assume F attains its maximum at some $c \in (a, b)$. Prove that $F'(c) = 0$.

Hint: Use the fact that $\frac{F(x)-F(c)}{x-c}$ is ≥ 0 for $x < c$ and it is ≤ 0 for $x > c$ to deduce that $F'(c) \geq 0$ and $F'(c) \leq 0$.

Exercise 35.12. (Rolle's Theorem). Let $F : [a, b] \to \mathbb{R}$ be continuous and assume F is differentiable on (a, b). Assume $F(a) = F(b) = 0$. Prove that there exists $c \in (a, b)$ such that $F'(c) = 0$.

Hint. Since F is continuous it attains its maximum or minimum at some $c \in (a, b)$ and apply the previous exercise to F (or $-F$).

Exercise 35.13. (Mean Value Theorem). Let $F : [a, b] \to \mathbb{R}$ be continuous and assume F is differentiable on (a, b). Then there exists $c \in (a, b)$ such that

$$F(b) - F(a) = F'(c)(b - a).$$

Hint: Apply Rolle's theorem to the function

$$G(x) = F(x) - F(a) - \frac{F(b) - F(a)}{b - a} \cdot (x - a).$$

Exercise 35.14. Let $F : (a, b) \to \mathbb{R}$ be differentiable.
1) Prove that if $F'(x) = 0$ for all $x \in (a, b)$ then F is constant.
2) Prove that if $F'(x) > 0$ for all $x \in (a, b)$ then F is strictly increasing.
3) Prove that if $F'(x) < 0$ for all $x \in (a, b)$ then F is strictly decreasing.

Exercise 35.15. (Chain rule) Prove that if $F \in C^1(J)$, $G \in C^1(I)$, I, J open intervals, $G(I) \subset J$, then $F \circ G \in C^1(I)$ and

$$D(F \circ G) = (D(F) \circ G) \cdot D(G).$$

Hint: For $a_n \mapsto a_0$, $a_n \neq a_0$, by the Mean Value Theorem, there exists c_n between $G(a_n)$ and $G(a_0)$ such that

$$F(G(a_n)) - F(G(a_0)) = F'(c_n)(G(a_n) - G(a_0)).$$

Divide by $a_n - a_0$, take the limit, and use the fact that $c_n \to G(a_0)$ and F' is continuous.

EXERCISE 35.16. Prove that the map $\exp : \mathbb{R} \to \mathbb{R}_{>0}$ is bijective.

Hint: Since exp is smooth with positive derivative it is strictly increasing so injective. To prove surjectivity use the Darboux property.

DEFINITION 35.17. We define the *logarithm* function

$$\log : \mathbb{R}_{>0} \to \mathbb{R}$$

to be the inverse of $\exp : \mathbb{R} \to \mathbb{R}_{>0}$.

EXERCISE 35.18. Prove that for all $a, b \in \mathbb{R}_{>0}$ we have

$$\log(ab) = \log(a) + \log(b).$$

EXERCISE 35.19. Prove that $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is differentiable and

$$\frac{d}{dx}(\log(x)) = \frac{1}{x}.$$

Hint: Use the chain rule for $\exp(\log(x)) = x$.

More generally one can define derivatives of functions of several variables as follows:

DEFINITION 35.20. Let $F : U \to \mathbb{R}$ be a function, where $U = I_1 \times ... \times I_n$ and $I_1, ..., I_n$ are open intervals in $\mathbb{R}$. Let $a = (a_1, ..., a_n) \in U$ and define $F_i : I_i \to \mathbb{R}$ by

$$F_i(x) = F(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_n)$$

(with the obvious adjustment if $i = 1$ or $i = n$). We say that F is *differentiable with respect to* x_i at a if and only if F_i is differentiable at a_i; in this case we define

$$\frac{\partial F}{\partial x_i}(a) = F_i'(a_i).$$

We say that F is *differentiable with respect to* x_i if and only if it is differentiable with respect to x_i at every $a \in U$. For such a function we have a well defined function $\frac{\partial F}{\partial x_i} : U \to \mathbb{R}$ which is also denoted by $D_i F$. Moreover we introduce, by recursion, the following terminology:

0) We say that F is of *class* C^0 if and only if F is continuous.

1) We say that F is of *class* C^1 if and only if F is of class C^0, F is differentiable with respect to each variable, and each $D_i F$ is continuous.

2) For $k \geq 1$ we say that F is of *class* C^{k+1} if and only if F is of class C^k, each $D_{i_1} ... D_{i_k} F$ is differentiable with respect to each variable, and all $D_{i_1} ... D_{i_k} D_{i_{k+1}} F$ are continuous.

3) We say that F is of *class* C^∞ (or *smooth*) if and only if F is of class C^k for every $k \geq 0$.

If U is an arbitrary open set in $\mathbb{R}^n$ (rather than a product of intervals) we say that F is of class C^k or C^∞ is the restriction of F to every product of intervals in U has this property. We also write

$$\frac{\partial^k F}{\partial x_{i_1} ... \partial x_{i_k}} := D_{i_1} ... D_{i_k} F.$$

We let $C^k(U)$, $C^\infty(U)$ be the sets of functions of class C^k, resp. C^∞ on U. $C^\infty(U)$ is a ring with respect to the addition and multiplication of functions.

Definition 35.21. If K is a product of closed intervals we say that F defined on K is of class C^k if it is of class C^k on the product of the corresponding open intervals and all $D_{i_1}...D_{i_l}F$, $l \leq k$, can be extended to continuous functions on K (which are necessarily unique and still denoted by the same symbol).

Exercise 35.22. (Chain rule). Prove that if $F : I \times J \to \mathbb{R}$, $(x, y) \mapsto F(x, y)$ is of class C^1 and $u : V \to I$, $v : V \to J$ are of class C^1, with I, J, V open intervals, then the map $t \mapsto (F(u(t), v(t)))$ is of class C^1 on V and for $t \in V$ we have

$$\frac{d}{dt}(F(u(t), v(t))) = \frac{\partial F}{\partial x}(u(t), v(t)) \cdot \frac{du}{dt}(t) + \frac{\partial F}{\partial y}(u(t), v(t)) \cdot \frac{dv}{dt}(t).$$

Hint: Let $t_n \to t_0$. Use the Mean Value Theorem to find c_n between $u(t_0)$ and $u(t_n)$ such that

$$F(u(t_n), v(t_n)) - F(u(t_0), v(t_n)) = \frac{\partial F}{\partial x}(c_n, v(t_n))(u(t_n) - u(t_0))$$

and a real number d_n between $v(t_n)$ and $v(t_0)$ such that

$$F(u(t_0), v(t_n)) - F(u(t_0), v(t_0)) = \frac{\partial F}{\partial x}(u(t_0), d_n)(v(t_n) - v(t_0));$$

then use the continuity of $\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}$.

Definition 35.23. A function $F : I \to \mathbb{R}^n$, $F(x) = (F_1(x), ..., F_n(x))$, I an open interval, is *k times differentiable, resp. smooth*, if and only if each of F_i is so. We write

$$F'(x) = (F_1'(x), ..., F_n'(x)).$$

A function $F : I \to \mathbb{C}^n$ is of *class* C^k, resp. *smooth*, if and only if it is so after we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

More generally, a function $F : U \to \mathbb{R}^n$ resp. to $\mathbb{C}^n$, with $U \subset \mathbb{R}^m$ open, is of class C^k (resp. smooth) if and only if its components are so. For

$$F(x_1, ..., x_m) = (F_1(x_1, ..., x_m), ..., F_n(x_1, ..., x_m))$$

we write

$$\frac{\partial F}{\partial x_i} = (\frac{\partial F_1}{\partial x_i}(x_1, ..., x_m), ..., \frac{\partial F_n}{\partial x_i}(x_1, ..., x_m)).$$

Exercise 35.24. Prove that compositions of C^1 functions of several variables are C^1 and a "chain rule" holds. For instance if in Exercise 35.22 we replace V by a subset of $\mathbb{R}^2$ then we have

$$\frac{\partial}{\partial t}(F(u(t,s), v(t,s))) = \frac{\partial F}{\partial x}(u(t,s), v(t,s)) \cdot \frac{\partial u}{\partial t}(t,s) + \frac{\partial F}{\partial y}(u(t,s), v(t,s)) \cdot \frac{\partial v}{\partial t}(t,s),$$

$$\frac{\partial}{\partial s}(F(u(t,s), v(t,s))) = \frac{\partial F}{\partial x}(u(t,s), v(t,s)) \cdot \frac{\partial u}{\partial s}(t,s) + \frac{\partial F}{\partial y}(u(t,s), v(t,s)) \cdot \frac{\partial v}{\partial s}(t,s).$$

Exercise 35.25. Let $F : I \times J \to \mathbb{R}$ be of class C^1, I and J open intervals, and assume

$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial y} = 0.$$

Prove that F is constant.

Hint: For all x the function $y \mapsto K(x, y)$ is constant; and for all y the function $x \mapsto K(x, y)$ is constant. So for all x, y, z, u we have $K(x, y) = K(x, z) = K(u, z)$.

EXERCISE 35.26. (Symmetry of second derivatives). Assume that $F : I \times J \to \mathbb{R}$, $(x, y) \mapsto F(x, y)$, is of class C^2 on a product of open intervals. Then

$$\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x}.$$

Hint: Fix $(a, b) \in I \times J$. Define

$$\Delta(x, y) := \frac{F(x, y) - F(x, b) - F(a, y) + F(a, b)}{(x - a)(y - b)}.$$

Let

$$\varphi(x, y) := F(x, y) - F(x, b)$$

Hence (using the Mean Value Theorem),

$$\Delta(x, y) = \frac{\varphi(x, y) - \varphi(a, y)}{(x - a)(y - b)} = \frac{\frac{\partial \varphi}{\partial x}(\xi, y)}{y - b} = \frac{\frac{\partial F}{\partial x}(\xi, y) - \frac{\partial F}{\partial x}(\xi, b)}{y - b}$$

for some ξ between x and a. Applying again the Mean Value Theorem we get

$$\frac{\frac{\partial F}{\partial x}(\xi, y) - \frac{\partial F}{\partial x}(\xi, b)}{y - b} = \frac{\partial^2 F}{\partial y \partial x}(\xi, \eta)$$

for some η between y and b. Let $x_n \to a$ and $y_n \to b$, $x_n \neq a$, $y_n \neq b$. Then, by the continuity of the second derivative we get

$$\frac{\partial^2 F}{\partial y \partial x}(a, b) = \lim \Delta(x_n, y_n).$$

On the other hand if we define

$$\psi(x, y) := F(x, y) - F(a, y)$$

then

$$\Delta(x, y) = \frac{\varphi(x, y) - \varphi(x, b)}{(x - a)(y - b)}.$$

Repeating the above argument we get

$$\frac{\partial^2 F}{\partial x \partial y}(a, b) = \lim \Delta(x_n, y_n).$$

DEFINITION 35.27. Let $P \in C^\infty(U)$, $U \subset \mathbb{R}^{r+2}$ open. An equation of the form

$$P\left(x, y(x), \frac{dy}{dx}(x), \frac{d^2 y}{dx^2}(x), ..., \frac{d^r y}{dx^r}(x)\right) = 0$$

is called a *differential equation* satisfied by $y \in C^\infty(I)$, $x \mapsto y(x)$, I an open interval (where we assume the composition makes sense).

The study of differential equations has numerous applications within Mathematics (e.g., geometry) as well as natural sciences (e.g., physics); below are some examples. In these examples, when x stands for "time" and y stands for "position" one uses a different notation: x and y are replaced by t and x respectively.

EXERCISE 35.28. (Classical particle in a field). Define Newton's equation as the differential equation

$$m \frac{d^2 x}{dt^2} = F(x, t),$$

where $x : I \to \mathbb{R}$, $t \mapsto x(t)$, is a smooth function, $m \in \mathbb{R}$, and $F : \mathbb{R} \times I \to \mathbb{R}$ is a smooth function. (In physical Argot $t \mapsto x(t)$ is interpreted as the trajectory of a particle submitted to a force field $F(x,t)$, m is interpreted as the rest mass of the particle, $v := \frac{dx}{dt}$ is interpreted as the velocity, and $\frac{d^2x}{dt^2}$ is interpreted as acceleration.) Assume

$$F(x,t) = -\frac{\partial}{\partial x}U(x,t)$$

(in physical Argot U is interpreted as the potential energy) and set $p = mv$ (interpreted as the momentum).

1) Prove that Newton's equation is equivalent to

$$\frac{dp}{dt} = -\frac{\partial U}{\partial x}.$$

2) Let $T = T(v) := \frac{mv^2}{2}$ (interpreted as the kinetic energy), assume $U = U(x)$ does not depend on t, and let $L = L(x,v) := T(v) - U(x)$ (called the Lagrangian). Prove that Newton's equation is equivalent to (the so called Euler-Lagrange equation):

$$\frac{d}{dt}\frac{\partial L}{\partial v} - \frac{\partial L}{\partial x} = 0.$$

3) Prove that

$$\frac{dT}{dt} = v\frac{dp}{dt}.$$

4) Prove that if one defines

$$H(x,p) = E(x,p) = T(v) + U(x) = \frac{p^2}{2m} + U(x)$$

(called the Hamiltonian, or energy) then $U(x) = E(x,0)$, $T = E - U$ and the Newton equation is equivalent to the *Hamilton system*:

$$\begin{aligned} \tfrac{dx}{dt} &= \tfrac{\partial H}{\partial p} \\ \tfrac{dp}{dt} &= -\tfrac{\partial H}{\partial x} \end{aligned}$$

5) Let $A(x,p)$ be a C^1 function of x and p. If $x(t)$, $p(t)$ satisfies a system of the form

$$\begin{aligned} \tfrac{dx}{dt} &= \tfrac{\partial H}{\partial p} \\ \tfrac{dp}{dt} &= -\tfrac{\partial H}{\partial x} \end{aligned}$$

(with H arbitrary) and one writes $\frac{dA}{dt} := \frac{d}{dt}A(x(t),p(t))$ then one has the "evolution equation:"

$$\frac{dA}{dt} = \{A,H\} := \frac{\partial A}{\partial x}\frac{\partial H}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial H}{\partial x}.$$

In particular we have the following "conservation of energy":

$$\frac{dH}{dt} = 0.$$

More generally, if $\{A,H\} = 0$ we have the "conservation law":

$$\frac{dA}{dt} = 0.$$

6) For functions A and B of x and p define the *Poisson bracket*

$$\{A,B\} := \frac{\partial A}{\partial x}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial x}.$$

Prove that for any 3 functions A, B, C as above one has the *Jacobi identity*:

$$\{\{A,B\},C\}+\{\{B,C\},A\}+\{\{C,A\},B\}=0.$$

7) Prove that if U is a constant, equivalently if $F=0$ (which is interpreted by saying that the particle is free, which is the context of Exercise 35.28) then $x=vt$ with v constant.

For a relativistic particle in a force field a similar theory is produced by relativity theory. We content ourselves with the following:

EXERCISE 35.29. (Two classical particles acting on each other). The Newton equations for "two particles on a line acting on each other by a force that depends only on the distance between them" are

$$\begin{aligned} \tfrac{d^2x_1}{dt^2} &= F(x_1-x_2) \\ \tfrac{d^2x_2}{dt^2} &= -F(x_1-x_2) \end{aligned}$$

Here $x_i(t)$ are interpreted as the trajectories of the two particles and the fact that the sum of the right hand sides is 0 is interpreted as the "equality of action and reaction." ($F(x)$ could be kx as in elastic force or kx^{-2} as in gravitational force, for instance.) Prove that if $(x_1(t), x_2(t))$ is a solution of the above system and if $u \in \mathbb{R}$ and

$$x_i^*(t) = x_i(t) + ut, \quad i \in \{1,2\},$$

then $(x_1^*(t), x_2^*(t))$ is again a solution of the above system. One interprets this as saying that the system above is invariant under the Galilean group; cf. Exercise 26.33.

EXERCISE 35.30. (Relativistic particle in a field). Let $v : I \to (-1,1) \subset \mathbb{R}$, $t \mapsto v(t)$ be a smooth map and let $p, E : I \to \mathbb{R}$ be the maps defined by $v \mapsto p = \frac{mv}{\sqrt{1-v^2}}$, $v \mapsto T = \frac{m}{\sqrt{1-v^2}} - m$, where $m \in \mathbb{R}$ is a constant. Prove that

$$\frac{dT}{dt} = v\frac{dp}{dt}.$$

(In physical Argot v, m, p, T are interpreted as velocity, rest mass, momentum, and kinetic energy for a not necessarily free particle in relativistic mechanics. There is an analogue of the Newton equations involving relativistic force but we will omit that discussion here.)

EXERCISE 35.31. (Wave Equation). A smooth function $\psi : \mathbb{R}^2 \to \mathbb{R}$, $(x,t) \mapsto \psi(x,t)$, is said to satisfy the *wave equation* (with velocity $c \in \mathbb{R}$) if

$$\frac{\partial^2\psi}{\partial t^2} = c^2\frac{\partial^2\psi}{\partial x^2}.$$

1) Prove that for every smooth function $f : \mathbb{R} \to \mathbb{R}$ the functions $\psi^-(x,t) := f(x-ct)$ and $\psi^+(x,t) := f(x+ct)$ satisfy the wave equation.

2) Prove that it is not the case that if $\psi = \psi(x,t)$ is a solution of the wave equation (with $c=1$) and $u \in \mathbb{R}$ then the function

$$\psi^*(x,t) := \psi(x+ut, t)$$

is also a solution of the wave equation. (One interprets this by saying that the wave equation is not "invariant under the Galilean group"; cf. Exercise 26.33).

3) Prove that if $\psi = \psi(x,t)$ is a solution of the wave equation (with $c = 1$) then for all $\alpha, \beta \in \mathbb{R}$, the function

$$\psi^{**}(x,t) := \psi(\alpha x + \beta t, \beta x + \alpha t)$$

is also a solution of the wave equation. (One interprets this by saying that the wave equation is "invariant under the Lorenz group"; cf. Exercise 26.34.)

The above is a "baby example" of the invariance of the so-called Maxwell equations under the Lorenz group in 1+3 dimensions and their non-invariance under the Galilean group. The attempt to create an alternative to Newtonian mechanics that is invariant under the Lorenz group (rather than the Galilean group) was what led Einstein to his restricted theory of relativity.

EXERCISE 35.32. (Laplace equation). A smooth function $\psi : \mathbb{R}^n \to \mathbb{R}$,

$$(x_1, ..., x_n) \mapsto \psi(x_1, ..., x_n)$$

is said to satisfy the *Laplace equation* if

$$\sum_{i=1}^{n} \frac{\partial^2 \psi}{\partial x_i^2} = 0.$$

Prove that if ψ is a solution of the Laplace equation then for all $(a_{ij}) \in O_n(\mathbb{R})$, the function

$$\psi^{\dagger}(x_1, ..., x_n) := \psi(\sum_j a_{1j}x_j, ..., \sum_j a_{nj}x_j)$$

is also a solution of the Laplace equation. (One interprets this by saying that the Laplace equation is "invariant under the group $O_n(\mathbb{R})$".)

EXERCISE 35.33. (Heat Equation). A smooth function $\psi : \mathbb{R}^2 \to \mathbb{R}$, $(x,t) \mapsto \psi(x,t)$, is said to satisfy the *heat equation* if

$$\frac{\partial \psi}{\partial t} = \frac{\partial^2 \psi}{\partial x^2}.$$

Prove that the function

$$\psi(t,x) := \frac{\exp(-x^2/4t)}{\sqrt{t}}$$

satisfies the heat equation.

EXERCISE 35.34. (Schrödinger Equation). Let h be a positive real number (the Planck constant). A smooth function $\psi : \mathbb{R}^4 \to \mathbb{C}$, $(t,x,y,z) \mapsto \psi(t,x,y,z)$, is said to satisfy the *Schrödinger equation* if

$$ih\frac{\partial \psi}{\partial t} = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2}.$$

Prove that if $f(x,y,z)$ satisfies the equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = E \cdot f$$

for some real E then the function

$$\psi(t,x,y,z) := \exp(-itE/h) \cdot f(x,y,z)$$

satisfies the Schrödinger equation.

EXERCISE 35.35. For $A \in \mathbb{C}^{n\times n}$ define

$$\exp(A) = \sum_{n=0} \frac{1}{n!} A^n \in \mathbb{C}^{n\times n}.$$

1) Prove that the entries of the partial sums are convergent sequences so the definition above is correct.

1) Prove that if $AB = BA$ then

$$\exp(A+B) = \exp(A)\exp(B).$$

2) Prove that

$$\frac{d}{dt}(\exp(tA)) = A \cdot \exp(tA).$$

3) Prove that if H is a Hermitian matrix then $\exp(-iH)$ is unitary.

4) Prove that if $\psi_0 \in \mathbb{C}^n$ and H is a Hermitian matrix (interpreted as the "Hamiltonian") then $\psi(t) := \exp(-iHt/h)\psi_0$ satisfies the "finite dimensional Schrödinger equation for time independent Hamiltonian"

$$ih\frac{d\psi}{dt} = H\psi.$$

(If we assume, in addition, that ψ_0 is an eigenvector of H with eigenvalue E, $H\psi_0 = E\psi_0$, then we have $\psi = \exp(-itE/h)\psi_0$. Since $x \mapsto \exp(ix)$ is periodic with period 2π one can interpret E/h as a "frequency" divided by 2π; since E corresponds to the energy, what we found is consistent with Planck's formula $E = h\nu$ where ν is the frequency and can actually be viewed as a justification for Schrödinger's equation.)

5) Assume $t \mapsto H(t)$ is a smooth function with $H(t)$ Hermitian. Let $t \mapsto \psi(t)$ be a smooth solution to the "finite dimensional Schrödinger equation for time dependent Hamiltonian"

$$ih\frac{d\psi}{dt} = H(t)\psi.$$

and let A be a Hermitian matrix. Then we have the following

$$ih\frac{d}{dt}(\mathbb{E}_{\psi(t)}(A)) = \mathbb{E}_{\psi(t)}([H(t), A]).$$

(This is an analogue of the evolution equation in classical mechanics: the commutator $[\ ,\ ]$ is the analogue of the Poisson bracket $\{\ ,\ \}$.)

Hint for 5): One has

$$\begin{aligned}
\tfrac{d}{dt}(\mathbb{E}_{\psi(t)}(A)) &= \tfrac{d}{dt}(\langle\psi(t), A\psi(t)\rangle \\
&= \langle\tfrac{d\psi}{dt}, A\psi\rangle + \langle\psi, A\tfrac{d\psi}{dt}\rangle \\
&= \tfrac{1}{ih}\langle H(t)\psi, A\psi\rangle - \tfrac{1}{ih}\langle\psi, AH(t)\psi\rangle \\
&= \tfrac{1}{ih}\langle\psi, H(t)A\psi\rangle - \tfrac{1}{ih}\langle\psi, AH(t)\psi\rangle \\
&= \tfrac{1}{ih}\langle\psi, [H(t), A]\psi\rangle \\
&= \tfrac{1}{ih}\mathbb{E}_{\psi(t)}([H(t), A]).
\end{aligned}$$

The additivity and the Leibniz rule have an algebraic character. This suggests the following:

DEFINITION 35.36. Let A be a commutative unital ring. A map $D : A \to A$ is called a *derivation* if and only if for all $a, b \in A$:
1) $D(a + b) = D(a) + D(b)$ (additivity);
2) $D(a \cdot b) = D(a) \cdot b + a \cdot D(b)$ (Leibniz rule).

EXAMPLE 35.37.
1) $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$, $D(F) := F'$, is a derivation.
2) The map $D : R[x] \to R[x]$, $D(f) := f'$ (cf. Definition 28.28) is a derivation.

EXERCISE 35.38. Prove that every derivation $D : \mathbb{Z} \to \mathbb{Z}$ is identically 0 i.e., $D(x) = 0$ for all $x \in \mathbb{Z}$.

Hint: By additivity $D(0) = 0$ and $D(-n) = -D(n)$. So it is enough to show $D(n) = 0$ for $n \in \mathbb{N}$. Proceed by induction on n. For the case $n = 1$, by the Leibniz rule,

$$D(1) = D(1 \cdot 1) = 1 \cdot D(1) + 1 \cdot D(1) = 2 \cdot D(1)$$

hence $D(1) = 0$. The induction step follows by additivity.

REMARK 35.39. Exercise 35.38 shows that there is no naive analogue of calculus in which rings of functions such as $C^\infty(\mathbb{R})$ are replaced by rings of numbers such as $\mathbb{Z}$. An analogue of calculus for $\mathbb{Z}$ is, however, considered desirable for the purposes of number theory. Such a theory has been developed. (Cf. A. Buium, *Arithmetic Differential Equations*, Math. Surv. and Monographs 118, American Mathematical Society, 2005.) In that theory the analogue of x is a fixed prime p and the analogue of the derivation $D = \frac{d}{dx} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is the operator

$$\frac{d}{dp} : \mathbb{Z} \to \mathbb{Z}, \quad \frac{dx}{dp} = \frac{x - x^p}{p}$$

which is well defined by Fermat's Little Theorem. For example,

$$\frac{d4}{d5} = \frac{4 - 4^5}{5} = -204.$$

CHAPTER 36

Integration

There are a number of concepts of integrability (due to Newton, Leibniz, Riemann, Lebesgue, etc.). The first rigorous definition is due to Riemann as follows:

Definition 36.1. A function $f : [a, b] \to \mathbb{R}$ is called (Riemann) *integrable* if and only if there exists a number $A \in \mathbb{R}$ with the following property. For all $\epsilon > 0$ there exists $\delta > 0$ such that for all

$$a = a_0 < \xi_1 < a_1 < \xi_2 < a_2 < \dots < a_{n-1} < \xi_n < a_n = b$$

if $\max_i(a_{i+1} - a_i) < \delta$ then

$$|(\sum_{i=1}^{n} f(\xi_i)(a_i - a_{i-1})) - A| < \epsilon.$$

One checks that A is unique and one calls A the (Riemann) *integral* of f. One writes

$$A = \int_a^b f(x)dx.$$

One calls the sums above *Riemann sums*.

Exercise 36.2. Assume the notation in the above definition.

1) Prove that if f is continuous then f is integrable.

2) Assume f is integrable. Prove that the restriction of f to any segment $[c, d] \subset [a, b]$ is integrable.

3) Assume f is integrable. Prove that for $a \leq c \leq b$ we have

$$\int_a^c f(x)dx + \int_c^b f(x)dx = \int_a^b f(x)dx.$$

4) Prove that if f and g are integrable and $\lambda \in \mathbb{R}$ then $f + g$ and λf are integrable and

$$\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx,$$

$$\int_a^b \lambda f(x)dx = \lambda \int_a^b f(x)dx.$$

5) Prove that if f is continuous then

$$|\int_a^b f(x)dx| \leq \int_a^b |f(x)|dx \leq (b - a) \cdot \max\{|f(x)| \mid a \leq x \leq b\}.$$

6) (The Fundamental Theorem of Calculus). Prove that if f is continuous and

$$F(x) := \int_a^x f(t)dt$$

then F is differentiable on (a,b) and $\frac{dF}{dx} = f$.

7) (Variant of the Fundamental Theorem of Calculus). Let $F : [a,b] \to \mathbb{R}$ be of class C^1. Then we have

$$\int_a^b F'(x)dx = F(b) - F(a).$$

8) (Interchanging limits and integral sign). If (f_n) is a sequence of continuous functions on $[a,b]$ converging uniformly to f (hence f is also continuous) then

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx.$$

9) (Interchanging differentiation and integral sign). Assume $f : [a,b] \times [c,d] \to \mathbb{R}$, $(x,y) \mapsto f(x,y)$, is of class C^1. Then the following holds for $x \in (a,b)$:

$$\frac{d}{dx}\left(\int_c^d f(x,y)dy\right) = \int_c^d \frac{\partial f}{\partial x}(x,y)dy.$$

10) (Interchangind integrals: Fubini's Theorem). Assume $f : A := [a,b] \times [c,d] \to \mathbb{R}$, $(x,y) \mapsto f(x,y)$, is continuous. Then the following holds:

$$\int_a^b \left(\int_c^d f(t,s)ds\right) dt = \int_c^d \left(\int_a^b f(t,s)dt\right) ds.$$

We denote the integrals above by

$$\int_A f.$$

11) (Commutation of limits and differentiation). Let (F_n) be a sequence of C^1 functions on an open interval (a,b). Suppose that the sequence (F'_n) converges uniformly to a function f and suppose that for some $c \in (a,b)$ the sequence $F_n(c)$ converges. Then the sequence (F_n) converges uniformly to a function F which is C^1 and for which $F' = f$.

Hint for 1): For $\Delta = \{\xi_1, ..., \xi_n\}$ the Riemann sum is between the sums

$$s(\Delta) = \sum_{i=1}^n m_i \cdot (a_i - a_{i-1}) \quad \text{and} \quad S(\Delta) = \sum_{i=1}^n M_i \cdot (a_i - a_{i-1})$$

where m_i and M_i are the minimum and the maximum of $f([\xi_{i-1}, \xi_i])$, respectively. One says that $\Delta \leq \Delta'$ if every interval determined by Δ is a union of intervals determined by Δ'. One checks that in the notation above,

$$s(\Delta) \leq s(\Delta') \leq S(\Delta') \leq S(\Delta).$$

One also notes that for every Δ and Δ' there exists Δ'' such that $\Delta \leq \Delta''$ and $\Delta' \leq \Delta''$. Using the uniform continuity of f one checks that for all ϵ there exists δ such that if Δ has maximum length of its segments $< \delta$ then $S(\Delta) - s(\Delta) < \epsilon$. In particular the supremum of the sums $s(\Delta)$ equals the infimum of the sums $S(\Delta)$. One checks that this common value, A, satisfies the condition in the definition of integrability.

Hint for 9): For $\lambda_n \to \lambda_0 \in (a,b)$, $a_n \neq a_0$, consider the functions

$$g_n(y) := \frac{f(\lambda_n, y) - f(\lambda_0, y)}{\lambda_n - \lambda_0}.$$

Then

$$\frac{1}{\lambda_n-\lambda_0}\cdot\left(\int_c^d f(\lambda_n,y)dy-\int_c^d f(\lambda_0,y)dy\right)=\int_c^d g_n(y)dy$$

so it is enough to show that $g_n(y)$ converges uniformly to $\frac{\partial f}{\partial x}(\lambda_0,y)$. However, by the Mean Value Theorem, for all y there exists $\xi_n(y)$ between λ_n and λ_0 such that we have

$$g_n(y)=\frac{\partial f}{\partial x}(\xi_n(y),y)$$

We conclude by the fact that continuous functions in 2 variables defined on a product of two closed intervals are *uniformly continuous* in a sense similar to the one in Exercise 33.28 (a fact which is proved similarly: check this!).

Hint for 10): Define

$$G(x,y):=\int_a^x\left(\int_c^y f(t,s)ds\right)dt,\quad H(x,y):=\int_c^y\left(\int_a^x f(t,s)dt\right)ds$$

and using 9) we get that G and H have equal partial derivatives. So $G-H$ must be constant, hence 0.

Hint for 11): By the Fundamental Theorem of Calculus and the commutation of limits with integration we have for each x:

$$F_n(x)-F_n(c)=\int_c^x F_n'(t)dt\to\int_c^x f(t)dt$$

so $F_n(x)$ converges to some $F(x)$ and

$$F(x)-F(c)=\int_c^x f(t)dt.$$

So, applying again the Fundamental Theorem of Calculus, F is C^1 and $F'=f$.

EXERCISE 36.3. (Heat equation, again). Prove that if f is continuous on $[a,b]$ then the function

$$\psi(t,x):=\int_a^b f(s)\frac{\exp(-(x-s)^2/4t)}{\sqrt{t}}ds$$

satisfies the heat equation:

$$\frac{\partial\psi}{\partial t}=\frac{\partial^2\psi}{\partial x^2}.$$

EXERCISE 36.4. (Fundamental theorem of calculus: complex valued case). For $f:[a,b]\to\mathbb{C}$, $f(t)=f_1(t)+if_2(t)$ with $f_1,f_2:[a,b]\to\mathbb{R}$ continuous we define $\int_a^b f(t)dt=\int_a^b f_1(t)dt+i\int_a^b f_2(t)dt$. Prove that if $F:[a,b]\to\mathbb{C}$, $F'=f$, then $\int_a^b f(x)=F(b)-F(a)$.

EXERCISE 36.5. (Existence and uniqueness of solutions for differential equations).

1) Assume $f:I\times J\to\mathbb{R}^n$, $(t,x)\mapsto f(t,x)$, $f=(f_1,...,f_n)^t$, is C^1 where $I\subset\mathbb{R}$ is an open interval in $\mathbb{R}$ and $J\subset\mathbb{R}^n$ is a product of open intervals. (We say f is a "time dependent" vector field on J.) Let $t_0\in I$ and let $K\subset J$ be a compact subset. Prove that there exists $\epsilon>0$ with the following property: for all $x\in K$ there exists

a unique differentiable function $\varphi = \varphi^x : (t_0 - \epsilon, t_0 + \epsilon) \to J$, $\varphi = (\varphi_1, ..., \varphi_n)^t$, such that $\varphi(t_0) = x$ and

$$\frac{d\varphi}{dt} = f(t, \varphi(t)).$$

2) Assume the situation in 1) and write $\Phi^t(x) = \varphi^x(t)$ for

$$(t, x) \in (t_0 - \epsilon, t_0 + \epsilon) \times K;$$

we call the family $(\Phi^t)_t$ with $\Phi^t : K \to \mathbb{R}$ the *flow* associated to the vector field f. Prove that if f does not depend on t, i.e., if $f(t, x) = f(x)$ (in which case we say f is a "time independent" vector field on J) then

$$\Phi^0(x) = \left(\Phi^t(x)\right)_{t=t_0} = x, \quad \left(\frac{\partial \Phi^t}{\partial t}\right)_{t=t_0} = f(x)$$

and

$$\Phi^{t_1+t_2} = \Phi^{t_1} \circ \Phi^{t_2}$$

whenever these are defined.

3) Use 1) to solve differential equations of the form

$$\frac{d^n\varphi}{dt^n} = f(t, \varphi(t), \frac{d\varphi}{dt}(t), ..., \frac{d^{n-1}\varphi}{dt^{n-1}}(t)),$$

with f and φ functions with values in $\mathbb{R}$.

Hint for 1) in case $n = 1$; the case n arbitrary is similar. Our differential equation is equivalent to the "integral equation"

$$\varphi(t) = x + \int_{t_0}^{t} f(s, \varphi(s))ds$$

satisfied by a continuous function φ. (We view this problem as a "fixed point" problem for the "integral operator" sending φ into the right hand side of the above integral equation. The solution below is a standard method for solving fixed point problems.)

We may assume K is a closed interval and we may replace I and J by smaller open intervals containing t_0 and K respectively so that the absolute values of f and $\frac{\partial f}{\partial x}$ are bounded on $I \times J$ by constants N and M, respectively. Then, by the Mean Value Theorem, we get that for all $t \in I$ and $y, z \in J$,

$$|f(t, y) - f(t, z)| \leq M \cdot |y - z|.$$

Let $\varphi_1(t) = x$ for all $t \in I$ and define the sequence of functions (φ_n) on I by recursion:

$$\varphi_{n+1}(t) = x + \int_{t_0}^{t} f(s, \varphi_n(s))ds.$$

For $t \in [t_0 - \epsilon, t_0 + \epsilon] \subset I$ and any continuous function h on an interval $[t_0 - \epsilon, t_0 + \epsilon] \subset I$ write

$$||h||_\epsilon = \max\{|h(t)| \mid t \in [t_0 - \epsilon, t_0 + \epsilon]\}.$$

Say $K = [a, b]$ and $J = (\alpha, \beta)$. For ϵ satisfying $\epsilon < N^{-1} \cdot \min(\beta - b, a - \alpha)$ we have $[x - N\epsilon, x + N\epsilon] \subset J$ and one checks (by induction on n) that

$$\varphi_n([t_0 - \epsilon, t_0 + \epsilon]) \subset J.$$

Also we have

$$\begin{aligned} |\varphi_{n+1}(t) - \varphi_n(t)| &= |F\{\varphi_n\} - F\{\varphi_{n-1}\}| \\ &\leq \int_{t_0}^{t} |f(s, \varphi_n(s)) - f(s, \varphi_{n-1}(s))| ds \\ &\leq 2\epsilon \cdot M \cdot ||\varphi_n - \varphi_{n-1}||_\epsilon \end{aligned}$$

from which we get

$$||\varphi_{n+1} - \varphi_n||_\epsilon \leq 2\epsilon \cdot M \cdot ||\varphi_n - \varphi_{n-1}||_\epsilon.$$

Choose ϵ such that, in addition, $\epsilon < 1/2M$. We get that the sequence (φ_n) is uniformly Cauchy and hence uniformly convergent on $[t_0 - \epsilon, t_0 + \epsilon]$ and its limit will be a continuous function satisfying our integral equation. Uniqueness of the solution to the integral equation is checked by considering the difference of the two solutions φ and ψ and repeating the calculation above to get

$$||\varphi - \psi||_\epsilon \leq 2\epsilon \cdot M \cdot ||\varphi - \psi||_\epsilon$$

which, for $\epsilon < 1/2M$, implies $||\varphi - \psi||_\epsilon = 0$ hence $\varphi = \psi$.

Remark: An analysis of the above argument shows that if one fixes t_0 then ϵ can be taken to be independent of x as long as x is taken to be in an open set whose closure is contained in J.

Hint for 2): We may assume $t_0 = 0$. To prove $\Phi^{t_1+t_2} = \Phi^{t_1} \circ \Phi^{t_2}$ one checks that for all x the functions

$$t \mapsto \Phi^{t_1+t}(x) = \varphi^x(t_1 + t)$$

and

$$t \mapsto \Phi^t(\Phi^{t_1}(x)) = \varphi^{\varphi^x(t_1)}(t)$$

satisfy the same differential equation and have the same value at 0. So they must be equal by 1). (To check this one uses the chain rule and the fact that f does not depend on t.)

Hint for 3): The equation is equivalent to the system of equations

$$\begin{aligned} \tfrac{d\varphi_0}{dt} &= \varphi_1, \\ \tfrac{d\varphi_1}{dt} &= \varphi_2 \\ \ldots & \ldots \ldots \\ \tfrac{d\varphi_{n-1}}{dt} &= f(t, \varphi_0(t), \varphi_1(t), ..., \varphi_{n-1}(t)). \end{aligned}$$

Remark 36.6. Assume the situation in 2) of Exercise 36.5 with $t_0 = 0$ and f smooth. Then one can show $(t, x) \mapsto \Phi^t(x)$ is smooth. If Φ^t_i are the components of Φ^t consider the "Jacobian matrix"

$$\frac{\partial \Phi^t}{\partial x} := \left(\frac{\partial \Phi^t_i}{\partial x_j} \right).$$

Then one has

$$\left(\frac{\partial \Phi^t}{\partial x} \right)_{t=0} = I$$

and

$$\left(\frac{\partial}{\partial t}\left(\frac{\partial \Phi^t}{\partial x}\right)\right)_{t=0} = \left(\frac{\partial}{\partial x}\left(\frac{\partial \Phi^t}{\partial t}\right)\right)_{t=0} = \left(\frac{\partial}{\partial x}\left(\frac{\partial \Phi^t}{\partial t}\right)_{t=0}\right) = \frac{\partial f}{\partial x}.$$

Hence

$$\det\left(\frac{\partial \Phi^t}{\partial x}\right) = \det\left(I + t\frac{\partial f}{\partial x} + t\epsilon(t,x)\right)$$

with $\epsilon(t,x) \to 0$ for $t \to 0$. We get

$$\left(\frac{\partial}{\partial t}\det\left(\frac{\partial \Phi^t}{\partial x}\right)\right)_{|t=0} = Tr\left(\frac{\partial f}{\partial x}\right) = \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i} =: div(f).$$

One calls $div(f)$ the *divergence* of f. Assume next that

$$div(f) = 0.$$

Hence

$$\left(\frac{\partial}{\partial t}\det\left(\frac{\partial \Phi^t}{\partial x}\right)\right)_{|t=0} = 0.$$

Using $\Phi^{t+\tau} = \Phi^\tau \circ \Phi^t$ for $\tau \to 0$ and the chain rule we get that

$$\frac{\partial}{\partial t}\det\left(\frac{\partial \Phi^t}{\partial x}\right) = 0$$

for all t. To summarize we have:

THEOREM 36.7. *(Liouville's Theorem). Assume we are given a smooth time independent vector field $f = f(x)$ on a product of open intervals in $\mathbb{R}^n$, with $div(f) = 0$. Let $(\Phi^t)_t$ be the flow corresponding to f. Then for all x the function*

$$t \mapsto \det\left(\frac{\partial \Phi^t}{\partial x}\right)$$

is constant.

REMARK 36.8. This applies, for instance to the Hamilton system (Exercise 35.28) in which case $n = 2$, x_1, x_2 are x, p and $f_1 = \frac{\partial H}{\partial p}$ and $f_2 = -\frac{\partial H}{\partial x}$ for then we have:

$$div(f) = \frac{\partial^2 H}{\partial x \partial p} - \frac{\partial^2 H}{\partial p \partial x} = 0.$$

From this, using the change of variables for double integrals, one can deduce the Poincaré recurrence theorem which says that if the energy H of a Hamiltonian system has the property that $H^{-1}([-N,N])$ is compact for all N and if the functions Φ^t are defined for all $t > 0$ then for every disk D and every T there exists $n \geq 1$ such that $\Phi^{nT}(D) \cap D \neq \emptyset$. Intuitively, for every initial state s_0 of a system there are arbitrarily close states s_1 such that if the system runs with initial state s_1 for sufficiently long time it returns to a state s_2 close to s_1 (and hence to s_0). We will not provide here the details but the idea is that if one chooses $X := H^{-1}([-N,N])$ to contain D then, using Liouville's Theorem, one can show that the sets in the family $(\Phi^{nT}(D))_{n\geq 1}$ have all the same area, hence they cannot be all disjoint because then X would have an infinite area. This is true for all D. It is enough to prove the theorem for D replaced by a smaller D. One then uses that Φ^t are "local diffeomorphisms" to get from $\Phi^{nT}(D) \cap \Phi^{(n+m)T}(D) \neq \emptyset$ that $D \cap \Phi^{mT}(D) \neq \emptyset$.

EXERCISE 36.9. (Maupertuis' principle of least action). Let $(x, v) \mapsto L(x, v)$ be C^2 on $\mathbb{R}^2$ and $t \mapsto \varphi(t)$ be a C^1 real function on $[a, b]$. Define the *action*

$$S[\varphi] := \int_a^b L(\varphi(t), \varphi'(t))dt.$$

Prove that the following are equivalent:

1) $\varphi(t)$ satisfies the *Euler-Lagrange* equation

$$\frac{d}{dt}\frac{\partial L}{\partial v} - \frac{\partial L}{\partial x} = 0.$$

2) For all C^1 function η on $[a, b]$ with $\eta(a) = \eta(b) = 0$ the function $\lambda \mapsto S[\varphi+\lambda\eta]$ for $\lambda \in \mathbb{R}$ has vanishing derivative at $\lambda = 0$.

Hint: Use the following computation where $\varphi(t) = \varphi(t) + \lambda\eta(t)$:

$$\begin{aligned} \tfrac{d}{d\lambda}\textstyle\int_a^b L(\varphi(t), \varphi'(t))dt &= \textstyle\int_a^b (\tfrac{\partial L}{\partial x}(\varphi(t), \varphi'(t))\eta(t) + \tfrac{\partial L}{\partial x}(\varphi(t), \varphi'(t))\eta'(t))dt \\ &= \textstyle\int_a^b (\tfrac{\partial L}{\partial x}(\varphi(t), \varphi'(t))\eta(t) - \tfrac{d}{dt}\tfrac{\partial L}{\partial x}(\varphi(t), \varphi'(t))\eta(t))dt, \end{aligned}$$

where, in the last equality, we used the Leibniz formula, the Fundamental Theorem of Calculus and the fact that $\eta(a) = \eta(b) = 0$.

DEFINITION 36.10. (Curvilinear integrals). A (complex valued) *differential form* of class C^k on an open set $U \subset \mathbb{R}^2$ is a pair of functions $\omega = (P, Q)$ of class C^k, $P, Q : U \to \mathbb{C}$. We write

$$\omega = P(x, y)dx + Q(x, y)dy.$$

A C^1 *path* in U is a map $\gamma : [a, b] \to \mathbb{R}$, $\gamma(t) = (x(t), y(t))$, such that $x(t), y(t)$ are of class C^1. For ω and γ as above we define the *integral of* ω *along* γ by

$$\int_\gamma \omega := \int_a^b (P(x(t), y(t))x'(t) + Q(x(t), y(t))y'(t))dt \in \mathbb{C}.$$

We abusively write

$$\gamma^*\omega = (P(x(t), y(t))x'(t) + Q(x(t), y(t)y'(t)))dt$$

and we refer to it as a differential form on $[a, b]$. A *pathwise* C^1 *path* in U is a sequence $\gamma = (\gamma_1, ..., \gamma_m)$ of C^1 paths in U, $\gamma_i : [a_i, b_i] \to \mathbb{R}$, such that $\gamma_i(b_i) = \gamma_{i+1}(a_i)$ for all $i = 1, ..., m-1$. In this situation we define

$$\int_\gamma \omega = \sum_{i=1}^m \int_{\gamma_i} \omega.$$

If we identify $\mathbb{C}$ with $\mathbb{R}^2$ via $(x, y) = z = x + iy$ then define

$$F(z)dz := F(x, y)dx + iF(x, y)dy.$$

DEFINITION 36.11. A differential form $\omega = Pdx + Qdy$ on U is called *closed* if and only if

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

A differential form $\omega = Pdx + Qdy$ on U is called *exact* if and only if there exists a C^1 function H on U such that

$$P = \frac{\partial H}{\partial x}, \quad Q = \frac{\partial H}{\partial y}.$$

Exercise 36.12. Prove that every exact form is closed.

Exercise 36.13. Consider the path $\gamma : [0,1] \to U := \mathbb{C}^\times \setminus \{z_0\}$, $z_0 = a+ib \in \mathbb{C}$ defined by

$$\gamma(t) := (a + r\cos(2\pi t), b + r\sin(2\pi t)) = z_0 + \exp(2\pi i t)$$

(which we call the "conterclockwise" *boundary* of the disk of raduius r with center z_0.) Prove that $\frac{dz}{z-z_0}$ is closed and

$$\int_\gamma \frac{dz}{z-z_0} = 2\pi i.$$

Hint: Reduce to the case $z_0 = 0$ and compute using the definition.

Exercise 36.14. (Green formula). Consider the paths $\gamma_1, \gamma_2, \gamma_3, \gamma_4 : [0,1] \to \mathbb{C} = \mathbb{R}^2$ defined by

$$\gamma_1(t) = (t, 0), \quad \gamma_2(t) = (1, t), \quad \gamma_3(t) = (1-t, 1), \quad \gamma_4(t) = (0, 1-t),$$
$$\gamma := (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$$

(which we call the *boundary of the unit square*) and let $\omega = Pdx + Qdy$. Then

$$\int_\gamma \omega = \int_{[0,1]\times[0,1]} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right).$$

In particular if ω is closed then

$$\int_\gamma \omega = 0.$$

Hint: We prove

$$\int_\gamma Qdy = \int_A \frac{\partial Q}{\partial x}, \quad \int_\gamma Pdx = -\int_A \frac{\partial P}{\partial y}.$$

For the first equality,

$$\int_A \frac{\partial Q}{\partial x} = \int_0^1 \left(\int_0^1 \frac{\partial Q}{\partial x} dx \right) dy = \int_0^1 (Q(1,y) - Q(0,y)) dy = \int_\gamma Qdy.$$

The second equality is similar. (One needs to use the commutation property of integrals.)

Exercise 36.15. Let $\omega = Pdx + Qdy$ be a C^1 differential form on $U \subset \mathbb{R}^2$ and $\varphi : V \to U$, $\varphi(u,v) = (x(u,v), y(u,v))$, a C^2 map, $V \subset \mathbb{R}^2$ open. Define the C^1 form

$$\varphi^*\omega := P(x(u,v), y(u,v))(\frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv) + Q(x(u,v), y(u,v))(\frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv).$$

Prove that if ω is closed then $\varphi^*\omega$ is closed.

Hint: A direct computation using the chain rule and the symmetry of the second derivatives.

Exercise 36.16. Let $\varphi : [0,1] \times [0,1] \to U \subset \mathbb{R}^2$ be of class C^2 and let γ be the path in U obtained by composing φ (in the obvious sense) with the boundary of the unit square (cf. Exercise 36.14). Then for every closed C^1 differential form ω on U we have

$$\int_\gamma \omega = 0.$$

Hint: Use Exercises 36.15 and 36.14.

EXERCISE 36.17. Prove that every closed differential form on a product of open intervals is exact.

Hint: We may assume the intervals contain 0. If $\omega := Pdx + Qdy$ is a closed form then define

$$H(x,y) = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega = \int_{\gamma_3} \omega + \int_{\gamma_4} \omega,$$

where

$$\begin{array}{ll} \gamma_1 : [0,x] \to \mathbb{R}, & \gamma_1(t) = (t,0), \\ \gamma_2 : [0,y] \to \mathbb{R}, & \gamma_2(y) = (x,t), \\ \gamma_3 : [0,y] \to \mathbb{R}, & \gamma_1(t) = (0,t), \\ \gamma_4 : [0,x] \to \mathbb{R}, & \gamma_2(y) = (t,y), \end{array}$$

and the definition is correct by Exercise 36.16. So using $H = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega$ we have

$$H = \int_0^x P(t,0)dt + \int_0^y Q(x,t)dt$$

hence

$$\begin{aligned} \frac{\partial H}{\partial x} &= P(x,0) + \int_0^y \frac{\partial Q}{\partial x}(x,t)dt \\ &= P(x,0) + \int_0^y \frac{\partial P}{\partial y}(x,t)dt \\ &= P(x,0) + P(x,y) - P(x,0) \\ &= P(x,y) \end{aligned}$$

Similarly, using $H = \int_{\gamma_3} \omega + \int_{\gamma_4} \omega$ we get $\frac{\partial H}{\partial y} = Q$.

DEFINITION 36.18. Let $f : U \to \mathbb{C}$, $U \subset \mathbb{C}$ open, be a map and $z_0 \in U$. We say that F is *holomorphic* at z_0 if and only if there exists a complex number $f'(z_0) \in \mathbb{C}$ such that for every sequence (z_n) in U with $z_n \to z_0$, $z_n \neq z_0$, we have

$$\frac{f(z_n) - f(z_0)}{z_n - z_0} \to f'(z_0).$$

We say f is holomorphic on U if it is C^1 and holomorphic at all points of U. We call f' the *(complex) derivative* of f.

The C^1 condition is actually redundant but we will not use this.

EXERCISE 36.19.

1) The sum and product of two holomorphic functions is holomorphic; the reciprocal of a non-zero holomorphic function is holomorphic. Moreover $f \mapsto f'$ is a derivation.

2) Every polynomial with complex coefficients is holomorphic on $\mathbb{C}$.

3) Every quotient of such polynomials is holomorphic on the complement in $\mathbb{C}$ of the set of roots of the denominator.

4) (Abel's criterion). Assume $(a_n)_{n\geq 0}$ is a sequence of complex numbers such that there exist real constants $r > 0, M > 0$ with $|a_n| \leq \frac{M}{r^n}$ for all $n \geq 0$. Then the series $\sum_{n=0}^{\infty} a_n z^n$ is convergent to some $f(z)$ for every z in the open *disk*

$$D := \{z \in \mathbb{C} \mid |z| < r\}$$

and $z \mapsto f(z)$ is holomorphic in D with $f'(z) = \sum_{n=0}^{\infty} na_n z^{n-1}$. In particular $\exp(z)$ is holomorphic on $\mathbb{C}$ and $(\exp(z))' = \exp(z)$.

Hint for 4): One cannot use Exercise 36.2, Part 11) because we are looking at complex (rather than real) differentiation. Instead one has to directly show that the expression

$$\frac{1}{w}(\sum_{n=0}^{\infty} a_n(z+w)^n - \sum_{n=0}^{\infty} a_n z^n) - \sum_{n=0}^{\infty} na_n z^{n-1}$$

approaches 0 as w approaches 0.

Exercise 36.20. (Cauchy-Riemann equations). Assume f is holomorphic on U.

1) Prove that if f is holomorphic on U then the form $f(z)dz = fdx + ifdy$ is closed.

2) Prove that if $f = u + iv$ is holomorphic on U with u and v real functions then u and v satisfy the *Cauchy-Riemann equations*:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

In particular u and v satisfy the *Laplace equation*:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

3) Prove that if $f = u + iv$ is C^1 on U with u and v real functions then u and v satisfy the Cauchy-Riemann equations then f is holomorphic on U.

4) Prove that if U is a product of two open intervals, $z_0 \in U$, f is C^1 on U and $F(z) = \int_\gamma f(z)dz$ where $f(z)dz := fdx + ifdy$ and $\gamma(t) = (1-t)z_0 + tz$, $t \in [0,1]$, then F is holomorphic with $F'(z) = f(z)$.

Hint: 2) follows from 1). To check 1) take first z_n real and then z_n purely imaginary (i.e. i times real). For 3) one writes

$$f(z + (\epsilon + i\eta)) - f(z) = u(x+\epsilon, y+\eta) + iv(x+\epsilon, y+\eta) - u(x,y) - iv(x,y),$$

one adds and subtracts $u(x, y+\eta)$ and $iv(x, y+\eta)$ and one applies the intermediate value theorem for both u and v in each variable. For 4) one uses Exercise 36.16 to show that for $w_n \to w_0$, $F(w_n) - F(w_0) = \int_{\eta_n} f(z)dz$ where $\eta_n(t) = (1-t)w_n + tw_0$ and one directly estimates the limit of the above integral divided by $w_n - w_0$.

Exercise 36.21. (Cauchy integral formula for the circle). Let γ be the the counterclockwise boundary of the disk D of radius r and center $z_0 \in \mathbb{C}$ (cf. Exercise 36.13) and let f be holomorphic on an open set of $\mathbb{C}$ containing this boudary. Then for all $z \in D$ we have

$$\frac{1}{2\pi i}\int_{\gamma_r} \frac{f(w)dw}{w-z} = f(z).$$

Hint: Let γ_ϵ be the boundary of a disk contained in our disk, with center z and radius ϵ. Use Exercises 36.20 and 36.16 to show that

$$\int_\gamma \frac{f(w)dw}{w-z} = \int_{\gamma_\epsilon} \frac{f(w)dw}{w-z}.$$

(This is done by constructing a map Φ from the unit square to the contour obtained from the boundaries of the two disks together with segments that go from one boundary to the other and back; see below.) Then note that $\int_{\gamma_\epsilon} \frac{f(w)dw}{w-z}$ can be made to approach $2\pi i f(z)$ as $\epsilon \to 0$ due to Exercise 36.13. Here is a hint for constructing Φ. In the simplest case when the two disks have the same center 0 one can take Φ to be

$$\Phi_0(t,s) = (s\epsilon + (1-s)r)\exp(2\pi i t).$$

If the center of the big disk is 0 but the center of the little disk is $a+bi = c\exp(2\pi i\theta)$ then one can take $\Phi = \Phi_2 \circ \Phi_1 \circ \Phi_0$ where

$$\Phi_1 : \rho\exp(2\pi i t) \mapsto \rho\exp(2\pi i t) + \frac{a(1-\rho)}{1-\epsilon},$$

$$\Phi_2 : \rho\exp(2\pi i t) \mapsto \rho\exp(2\pi i(t+\theta)).$$

Note that Φ_1 leaves the boundary of the big disk fixed and translates the boundary of the little disk by a to the right. On the other hand Φ_2 rotates the whole picture by and "angle" of $2\pi\theta$.

EXERCISE 36.22. (Analyticity and Cauchy inequalities). Let f be holomorphic on an open set containing the boundary of an open disk D with center 0 and radius r. Then for all $z \in D$ one has an equality of the form

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

with $a_n \in \mathbb{C}$ and the series in the right hand side convergent inside the disk. The coefficients a_n are given by

$$a_n = \frac{1}{2\pi i}\int_\gamma \frac{f(w)dw}{w^{n+1}},$$

where γ is the boundary of the circle of radius r and center 0. In particular if $M(r)$ is the maximum of $|f(w)|$ on γ then for all n,

$$|a_n| \leq \frac{M(r)}{r^n}.$$

Hint: Write the quantity inside the Cauchy integral formula as a series in z using the equality

$$\frac{1}{w-z} = \sum_{n=0}^{\infty} \frac{z^n}{w^{n+1}}.$$

Then integrate and use the criterion on commutation of limits with integration.

The above Exercise has an obvious analogue for disks not centered at 0. So by Exercises 36.22 and 36.19 we have the following:

THEOREM 36.23. *If $f : U \to \mathbb{C}$ is a map on an open set $U \subset \mathbb{C}$ then the following are equivalent:*

1) f is holomorphic on U.

2) For every point $z_0 \in U$ there is an open disk $D \subset U$ with center z_0 and radius r and a sequence of complex numbers $(a_n)_{n\geq 0}$ satisfying $|a_n| \leq \frac{M}{r^n}$ for some constant $M > 0$ such that for all $z \in D$ we have

$$f(z) = \sum a_n(z-z_0)^n.$$

Note that by the condition on a_n in 2) the sum is convergent on D so the equality in 2) makes sense.

REMARK 36.24. A function satisfying condition 2) above is called *analytic*. So f is holomorphic if and only if it is analytic.

THEOREM 36.25. *(Liouville's Theorem). Let f be holomorphic on $\mathbb{C}$. If f is bounded on $\mathbb{C}$ then f is a constant.*

Proof. In Cauchy's inequalities $M(r)$ is bounded by a constant independent of r. Taking $r \to \infty$ we get $a_n = 0$ for all $n \geq 1$. $\square$

EXERCISE 36.26. Use Liouville's Theorem to prove the Fundamental Theorem of Algebra (Theorem 28.14) saying very non-constant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.

Hint: If f is a polynomial without roots in $\mathbb{C}$ then $\frac{1}{f}$ is holomorphic and bounded on $\mathbb{C}$(check!), hence constant by Liouville's Theorem.

CHAPTER 37

Curvature

In this Chapter we explain some of the main concepts of differential geometry and prove the Fundamental Theorem of Riemannian geometry. This subject was started by Gauss and Riemann in the mid 19th century and played a key role in Einstein's general theory of relativity as well as in the mathematics of the 20th century.

DEFINITION 37.1. (Vector fields) Let $D \subset \mathbb{R}^n$ be an open set with coordinates $x = (x_1, ..., x_n)$ and denote by $\mathfrak{X}(D)$ the $\mathbb{R}$-vector space of all $\mathbb{R}$-linear maps X from $C^\infty(D)$ to itself of the form

$$f \mapsto X(f) = \sum_{i=1}^{n} u_i \frac{\partial f}{\partial x_i},$$

where $u_i \in C^\infty(D)$. The elements of $\mathfrak{X}(D)$ are called *vector fields* on D. Note that $\mathfrak{X}(D)$ is a free $C^\infty(D)$-module of finite rank with basis

$$X_i := \frac{\partial}{\partial x_i}, \quad i \in \{1, ..., n\}.$$

REMARK 37.2. Vector fields are derivations on the ring $C^\infty(D)$. In fact one can prove that every derivation on $C^\infty(D)$ which vanishes on $\mathbb{R}$ is a vector field; we will not need this result. Instead we make the following observation. Let us say that a map $\Phi : D \to \tilde{D}$ between two open sets of $\mathbb{R}^n$ is a *diffeomorphism* if it is bijective and both Φ and Φ^{-1} are smooth. We have the following:

EXERCISE 37.3. Every diffeomorphism $\Phi : D \to \tilde{D}$ induces a natural linear isomorphism (still denoted by) $\Phi : \mathfrak{X}(D) \to \mathfrak{X}(\tilde{D})$. (We say that $\mathfrak{X}$ is "intrinsically defined" or "compatible with diffeomorphisms.")

Hint: Use the corresponding "chain rule."

For a vector space V over $\mathbb{R}$ let us denote by $End_{\mathbb{R}}(V)$ the space of all $\mathbb{R}$-linear maps $V \to V$. Also, for linear maps $A, B \in End_{\mathbb{R}}(V)$ we write

$$[A, B] := A \circ B - B \circ A \in L(V).$$

If V is a $C^\infty(D)$-module we denote by $End_{C^\infty(D)}(V)$ the space of all $C^\infty(D)$-linear maps $V \to V$; clearly

$$End_{C^\infty(D)}(V) \subset End_{\mathbb{R}}(V).$$

DEFINITION 37.4. (Tangent space). A *tangent vector* at a point $P \in D$ is an $\mathbb{R}$-linear map $C^\infty(D) \to \mathbb{R}$ of the form

$$f \mapsto \sum_{i=1}^{n} \lambda_i \cdot \frac{\partial f}{\partial x_i}(P),$$

with $\lambda_i \in \mathbb{R}$. We denote by $T_P D$ the set of all tangent vectors at P and we call this vector space the *tangent space* of D at P. For every vector field $X = \sum u_i X_i \in \mathfrak{X}(D)$ and $P \in D$ we define the *value* of X at P as being the tangent vector $X_P \in T_P D$,

$$X_P f := \sum_{i=1}^{n} u_i(P) \frac{\partial f}{\partial x_i}(P).$$

DEFINITION 37.5. (Connections). A (linear) *connection* on D i.e., a linear map of vector spaces over $\mathbb{R}$,

$$\nabla : \mathfrak{X}(D) \to End_{\mathbb{R}}(\mathfrak{X}(D)), \quad X \mapsto \nabla_X,$$

satisfying

$$\begin{aligned} \nabla_{fX} Y &= f \nabla_X Y \\ \nabla_X (fY) &= X(f)Y + f \nabla_X Y \end{aligned}$$

for $f \in C^\infty(D)$ and $X, Y \in \mathfrak{X}(D)$. We set

$$\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k$$

where

$$\Gamma_{ij}^k \in C^\infty(D)$$

are called the *Christoffel symbols of the second kind.*

DEFINITION 37.6. (Curves). A (smooth) *curve* $c : I \to D$ is a map where I is an open interval in $\mathbb{R}$ containing 0,

$$c(t) = (c_1(t), ..., c_n(t))$$

with $x_i(t)$ smooth, such that the vector

$$c'(t) := \left(\frac{dc_1}{dt}, ..., \frac{dc_n}{dt} \right)$$

is non-zero for all t. A *vector field along* c is a map that attaches to every $t \in I$ a tangent vector at $c(t)$,

$$t \mapsto \sum_{i=1}^{n} a_i(t) \left(\frac{\partial}{\partial x_i} \right)_{c(t)},$$

such that $t \mapsto a_i(t)$ are smooth. Every vector field Y on D induces a vector field which we denote by Y_c along c defined by $t \mapsto Y_{c(t)}$. Another example is the velocity of c, defined by

$$t \mapsto c'(t) \left(\frac{\partial}{\partial x} \right)_{c(t)} := \sum_{i=1}^{n} \frac{dc_i}{dt}(t) \left(\frac{\partial}{\partial x_i} \right)_{c(t)}.$$

Consider a curve c and assume one can choose (locally) a vector field $X(c)$ on D with the property that

$$X(c)_c = c'(t) \left(\frac{\partial}{\partial x} \right)_{c(t)}$$

for all $t \in I$. Assume we are given a connection ∇, a vector field Y on D and a curve c as above.

DEFINITION 37.7. (Derivative of vector fields). The *derivative* of Y_c is defined to be the vector field along c, denoted by Y_c', given by

$$t \mapsto (Y_c')_{c(t)} := (\nabla_{X(c)}Y)_{c(t)}.$$

REMARK 37.8. The map Y_c' only depends on c and Y_c but not on $X(c)$ or Y; in fact if X_i are coordinate vector fields, c is given in local coordinates around P_0 by an n-tuple of functions $(c_1(t), \dots, c_n(t))$ and $Y = \sum_j w_j X_j$ then we have

$$(Y_c')_{c(t)} = \sum_k \left(\frac{d}{dt}(w_k \circ c)(t) + \sum_{ij} \Gamma_{ij}^k(c(t)) w_j(c(t)) \frac{dc_i}{dt} \right) X_k.$$

DEFINITION 37.9. (Parallel transport). We say that Y_c is *parallel* (with respect to ∇) if

$$Y_c' = 0;$$

equivalently if

$$\frac{d}{dt}(w_k \circ c)(t) + \sum_{ij} \Gamma_{ij}^k(c(t))(\delta_t c_i) w_j(c(t)) = 0.$$

REMARK 37.10. By the existence and uniqueness theorem for differential equations for every tangent vector v_0 at P_0 there is a unique vector field along c (defined on some smaller interval) that is parallel and whose value at P_0 is v_0; the map sending v_0 into the above parallel vector field is referred to as *parallel transport*.

DEFINITION 37.11. (Geodesics). A curve c is called a *geodesic* through a point $P_0 \in D$ if $c(0) = P_0$ and we have

$$(\nabla_{X(c)} X(c))_{c(t)} = 0$$

for $t \in I$; in other words if $X(c)_c$ is parallel.

REMARK 37.12. This definition is independent on the choice of $X(c)$ and in fact if X_i are coordinate vector fields and c is given in coordinates around P_0 by an n-tuple of functions $(c_1(t), \dots, c_n(t))$ then c is a geodesic through P_0 if and only if $c_i(0) = 0$ and

$$\frac{d^2c}{dt^2} + \sum_{ij} \Gamma_{ij}^k(c(t)) \frac{dc_i}{dt} \frac{dc_j}{dt} = 0.$$

Again, for every tangent vector v_0 at P_0 there is a unique geodesic on M whose tangent vector at P_0 is v_0.

DEFINITION 37.13. (Symmetry). A connection ∇ is *symmetric* if

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

for all vector fields $X, Y \in \mathfrak{X}(D)$.

REMARK 37.14. Every diffeomorphism $\Phi : D \to \tilde{D}$ induces a natural bijection (which we still denote by) Φ between the set of connections on D and the set of connections on $\tilde{D}$. (We say that the set of connections is "intrinsically defined" or "compatible with diffeomorphisms.") Doreover ∇ is symmetric on D if and only if $\Phi\nabla$ is symmetric on $\tilde{D}$.

EXERCISE 37.15. Prove that the symmetry condition is equivalent to the equalities

$$\Gamma^k_{ij} = \Gamma^k_{ji}.$$

DEFINITION 37.16. (Curvature). The *curvature* of a connection ∇ is the map

$$\mathcal{R} : \mathfrak{X}(D) \times \mathfrak{X}(D) \to End_{\mathbb{R}}(\mathfrak{X}(D))$$

that attaches to every $(X, Y) \in \mathfrak{X}(D) \times \mathfrak{X}(D)$ the $\mathbb{R}$-linear map

$$\mathcal{R}(X, Y) := \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]} : \mathfrak{X}(D) \to \mathfrak{X}(D).$$

EXERCISE 37.17.

1) Prove that the image of $\mathcal{R}(X, Y)$ is actually contained in $End_{C^\infty(D)}(\mathfrak{X}(D))$.

2) Prove that

$$\mathcal{R}(fX, Y) = \mathcal{R}(X, fY) = f \cdot \mathcal{R}(X, Y)$$

for all $f \in C^\infty(D)$ and $X, Y \in \mathfrak{X}(D)$.

REMARK 37.18. Clearly for every diffeomorphism $\Phi : D \to \tilde{D}$ the curvature of $\Phi\nabla$ corresponds, under the natural map induced by $\Phi : \mathfrak{X}(D) \to \mathfrak{X}(\tilde{D})$ to the curvature of ∇. We say that the curvature is "invariantly defined" or "compatible with diffeomorphisms."

DEFINITION 37.19. (Metrics). A (semi-Riemannian) *metric* on D is a symmetric $n \times n$ matrix $g = (g_{ij})$ with coefficients $g_{ij} \in C^\infty(D)$ such that

$$\det(g) \in C^\infty(D)^\times$$

Given a metric as above there is a unique $C^\infty(D)$-bilinear map (still denoted by)

$$g : \mathfrak{X}(D) \times \mathfrak{X}(D) \to C^\infty(D), \quad (X, Y) \mapsto g(X, Y)$$

such that $g(X_i, X_j) = g_{ij}$. Note that the map g is symmetric in the sense that $g(X, Y) = g(Y, X)$. We abusively identify the metric g with the map g and also with formal expression

$$\sum_{ij} g_{ij} dx_i dx_j.$$

By the determinant condition it follows that for every $C^\infty(D)$-module homomorphism $\omega : \mathfrak{X}(D) \to C^\infty(D)$ there exists a unique $Z \in \mathfrak{X}(D)$ such that for all $Y \in \mathfrak{X}(D)$ we have

$$g(Z, Y) = \omega(Y).$$

EXERCISE 37.20. Clearly, every diffeomorphism $\Phi : D \to \tilde{D}$ induces a natural bijection Φ between the set of metrics on D and the set of metrics on $\tilde{D}$. (So the set of metrics is "intrinsically defined" or "compatible with diffeomorphisms.")

Hint: Use the corresponding "chain rule."

DEFINITION 37.21. (Metric condition). We say that a connection ∇ on D is *metric* with respect to a metric g if

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

for all vector fields $X, Y, Z \in \mathfrak{X}(D)$. Given a connection ∇ and a metric the *Christoffel symbols of the first kind* of ∇ are defined as

$$\Gamma_{ijk} := \sum_m \Gamma^m_{ij} g_{mk}.$$

THEOREM 37.22. *(Fundamental Theorem of Riemannian geometry). Given a metric g there is a unique connection that is symmetric and metric with respect to g. It is given by*

$$\Gamma_{ijk} = \frac{1}{2}(X_i g_{jk} + X_j g_{ik} - X_k g_{ij}).$$

This unique connection is called the *Levi-Civita connection.* Clearly if $\Phi : D \to \tilde{D}$ is a diffeomorphism and we have metrics on D and $\tilde{D}$ that correspond to each other via Φ then the Levi-Civita connections attached to these metrics correspond under Φ.

Sketch of proof. Existence is proved by a direct computation. To prove uniqueness note that the metric condition for $X = X_i$, $Y = X_j$, $Z = X_k$ yields

$$X_i g_{jk} = \sum_m \Gamma^m_{ij} g_{mk} + \sum_m \Gamma^m_{ik} g_{mj}$$

Permuting indices one derives similar expressions for $X_j g_{ik}, X_k g_{ij}$. One then checks directly that the formula in the Theorem holds. □

EXERCISE 37.23. Fill in the details in the proof of the above Theorem.

DEFINITION 37.24. If $\mathcal{R}$ is the curvature of the Levi-Civita connection ∇ one sets

$$\mathcal{R}(X_i, X_j)X_k = \sum_m R^m_{kij} X_m$$

with smooth functions R^m_{kij} and one defines the *Riemann curvature tensor* by

$$R_{lkij} := \sum_m R^m_{kij} g_{lm}.$$

One defines the *Ricci curvature tensor*

$$R_{ij} := \sum_k R^k_{ikj} = \sum_{k,l} g^{lk} R_{likj}$$

and the *scalar curvature*

$$Scal := R := \sum_{i,j} g^{ij} R_{ij} = \sum_{i,j,k,l} g^{ij} g^{lk} R_{likj},$$

where (g^{ij}) is the inverse of $g = (g_{ij})$.

EXERCISE 37.25. Prove that the map

$$\mathfrak{X}(D) \times \mathfrak{X}(D) \times \mathfrak{X}(D) \times \mathfrak{X}(D) \to C^\infty(D),$$
$$(X, Y, Z, W) \mapsto Riem(X, Y, Z, W) := g(\mathcal{R}(X, Y)Z, W)$$

is $C^\infty(D)$-multilinear and compatible with diffeomorphisms (in the obvious sense).

REMARK 37.26. We have

$$Riem(X_i, X_j, X_k, X_l) = g(\mathcal{R}(X_i, X_j)X_k, X_l) = \sum_m R^m_{kij} g_{ml} = R_{lkij}.$$

EXERCISE 37.27. Prove that the map

$$Ric : \mathfrak{X}(D) \times \mathfrak{X}(D) \to C^\infty(D), \quad Ric(Y, Z) := Tr(X \mapsto \mathcal{R}(X, Y)Z)$$

is $C^\infty(D)$-multilinear and compatible with diffeomorphisms (in the obvious sense). Note that the map $X \mapsto \mathcal{R}(X, Y)Z$ is a $C^\infty(D)$-linear map $\mathfrak{X}(D) \to \mathfrak{X}(D)$ and $\mathfrak{X}(D)$ is free of finite rank so it has a well defined trace.

REMARK 37.28. We have

$$Ric(X_j, X_k) = R_{kj}.$$

REMARK 37.29. Prove that for all X, Y we have

$$Tr(\mathcal{R}(X,Y)) := Tr(Z \mapsto \mathcal{R}(X,Y)Z) = 0.$$

Hint: One computes:

$$Tr(\mathcal{R}(X_i, X_j)) = \sum_m R^m_{mij} = \sum_{m,l} R_{lmij} g^{lm} = -\sum_{m,l} R_{mlij} g^{ml} = -\sum_{m,l} R_{lmij} g^{lm},$$

to the above equals 0.

EXERCISE 37.30. Define the map $\mathcal{R}ic : \mathfrak{X}(D) \to \mathfrak{X}(D)$ by the condition

$$g(\mathcal{R}ic\ X, Y) = Ric(X,Y), \quad X, Y \in \mathfrak{X}(D).$$

(The map $\mathcal{R}ic$ is well defined in view of the condition that $\det(g)$ is invertible and $Y \mapsto Ric(X,Y)$ is $C^\infty(D)$-linear.)

1) Prove that $\mathcal{R}ic$ is $C^\infty(D)$-linear.

2) Prove that the scalar curvature satisfies

$$Scal = Tr(\mathcal{R}ic) := Tr(X \mapsto \mathcal{R}ic\ X).$$

3) Prove that the scalar curvature is compatible with diffeomorphisms $\Phi : D \to \tilde{D}$ in the sense that that if two metrics on D and $\tilde{D}$ correspond to each other via Φ then the scalar curvature on D is obtained from that on $\tilde{D}$ by composition with Φ.

Hint for 3): It follows directly from 1) and 2).

Hint for 2): If $\mathcal{R}ic\ X_i = \sum_k \lambda_{ik} X_k$ then

$$\sum_k \lambda_{ik} g_{kj} = R_{ij}$$

hence

$$\lambda_{il} = \sum_j g^{jl} R_{ij}$$

hence

$$\sum_i \lambda_{ii} = \sum_{ij} g^{ji} R_{ij} = \sum_{ij} g^{ij} R_{ij} = Scal.$$

EXERCISE 37.31. Prove the formula

$$R^l_{kij} = X_i \Gamma^l_{kj} - X_j \Gamma^l_{ki} + \sum_m \Gamma^l_{mi} \Gamma^m_{kj} - \sum_m \Gamma^l_{mj} \Gamma^m_{ik}$$

Hint: A direct computation.

EXERCISE 37.32. (Symmetries of *Riem*). Prove that for all $X, Y, Z, V, W \in \mathfrak{X}(D)$ we have:

1) $\mathcal{R}(X,Y) = -\mathcal{R}(Y,X)$.

2) $g(\mathcal{R}(X,Y)V, W) = -g(\mathcal{R}(X,Y)W, V)$.

3) $\mathcal{R}(X,Y)Z + \mathcal{R}(Y,Z)X + \mathcal{R}(Z,X)Y = 0$.

4) $g(\mathcal{R}(X,Y)V, W) = g(\mathcal{R}(V,W)X, Y)$.

Equivalently, for all i, j, k, l,

1) $R_{ijkl} + R_{jikl} = 0$

2) $R_{ijkl} + R_{ijlk} = 0$

3) $R_{ijkl} + R_{jkil} + R_{kijl} = 0$

4) $R_{ijkl} - R_{klij} = 0$

Hint for 1): Clear from the definition of $\mathcal{R}(X, Y)$.

Hint for 2): It is enough to show that $g(\mathcal{R}(X, Y)U, U) = 0$ for all U (write $U = V + W$). By $C^\infty(D)$-linearity we may assume $X = X_i$ and $Y = X_j$ so we may assume $[X, Y] = 0$. Then we have:

$$\begin{aligned} g(\mathcal{R}(X, Y)U, U) &= g(\nabla_X \nabla_Y U, U) - g(\nabla_Y \nabla_X U, U) \\ &= X(g(\nabla_Y U, U)) - Y(g(\nabla_X U, U)) \\ &= \tfrac{1}{2} XY(g(U, U)) - \tfrac{1}{2} YX(g(U, U)) = 0. \end{aligned}$$

Hint for 3): Again we may assume $[Y, Z] = 0$. For $F(X, Y, Z)$ write

$$(Cyc\ F)(X, Y, Z) := F(X, Y, Z) + F(Y, Z, X) + F(Z, X, Y).$$

Now

$$\begin{aligned} Cyc\ \mathcal{R}(X, Y)Z &= Cyc\ \nabla_X \nabla_Y Z - Cyc\ \nabla_Y \nabla_X Z \\ &= Cyc\ \nabla_X \nabla_Y Z - Cyc\ \nabla_X \nabla_Z Y \\ &= Cyc\ \nabla_X [Y, Z] = Cyc\ 0 = 0. \end{aligned}$$

Hint for 4): Note that from 3) we have

$$R_{ijkl} + R_{jkil} + R_{kijl} = 0$$

Switching k and l one gets

$$R_{ijlk} + R_{jlik} + R_{lijk} = 0.$$

Subtracting the last two equations and using 1) and 2) one gets

$$2R_{ijkl} + R_{jkil} + R_{jlki} + R_{kijl} + R_{iljk} = 0.$$

Replacing i, j, k, l by k, l, i, j we get

$$2R_{klij} + R_{likj} + R_{ljik} + R_{iklj} + R_{kjli} = 0.$$

Subtracting the last two equations and using 1) and 2) we get $R_{ijkl} = R_{klij}$.

EXERCISE 37.33. (Symmetry of Ric). Prove that for all $X, Y \in \mathfrak{X}(D)$ we have

$$Ric(X, Y) = Ric(Y, X).$$

Equivalently, for all i, j,

$$R_{ij} = R_{ji}.$$

Hint: Using the previous Exercise we have

$$R_{ij} := \sum_{k,l} g^{lk} R_{likj} = \sum_{k,l} g^{lk} R_{kjli} = \sum_{k,l} g^{kl} R_{kjli} = \sum_{k,l} g^{lk} R_{ljki} = R_{ji}.$$

EXERCISE 37.34. Prove that

$$R_{ijkl} = R^{(2)}_{ijkl} + R^{(1)}_{ijkl},$$

where

$$R^{(2)}_{ijkl} = \frac{1}{2}(X_i X_k g_{jl} + X_l X_j g_{ki} - X_i X_l g_{kj} - X_k X_j g_{il}),$$

and $R^{(1)}_{ijkl}$ is a polynomial with coefficients in $\mathbb{R}$ in the functions

$$\{g_{st}, \det(g)^{-1}, X_m g_{st} \mid \ m, s, t \in \{1, \ldots, n\}\}.$$

EXERCISE 37.35.

1) Let $D = (0, R) \times (0, 2\pi)$ with coordinates (r, θ) and let

$$f : (0, R) \to \mathbb{R}, \quad r \mapsto f(r)$$

be a smooth function. (The notation is that of "polar coordinates.") Consider the metric on D given by

$$f(r)(dr^2 + r^2 d\theta^2).$$

Compute the Christoffel symbols of the Levi-Civita connection and the scalar curvature as functions of r.

2) Let $U = D_R(0, 0) \setminus \{(0, 0)\}$ with coordinates (x, y) and consider the metric on U given by

$$f(\sqrt{x^2 + y^2})(dx^2 + dy^2).$$

Compute the scalar curvature as a function of x, y.

Hint for 2): Use the "polar coordinates" map $\Phi(r, \theta) = (r \cos\theta, r \sin\theta)$ (plus a "rotated version" of this map) and the compatibility of the scalar curvature with diffeomorphisms.

DEFINITION 37.36. A metric g is called *Einstein* if and only if there exists $\kappa \in \mathbb{R}$ such that for all X, Y we have

$$Ric(X, Y) = \kappa \cdot g(X, Y).$$

Equivalently, for all i, j,

$$R_{ij} = \kappa \cdot g_{ij}.$$

The theory can be "globalized" as follows. Recall that by a homeomorphism of topological spaces we mean a continuous bijection whose inverse is continuous.

DEFINITION 37.37. A *smooth manifold* is a topological space M equipped with an open cover

$$M = \bigcup_{\alpha} U_\alpha$$

and with homeomorphisms $\Phi_\alpha : U_\alpha \to D_\alpha$ (where $D_\alpha \subset \mathbb{R}^n$ are open sets) such that for all indices α, β the maps $\Phi_{\alpha\beta} := \Phi_\alpha \circ \Phi_\beta^{-1}$ are smooth maps (hence diffeomorphisms) from $D_{\beta\alpha} := \Phi_\beta(U_\beta \cap U_\alpha)$ to $D_{\alpha\beta} := \Phi_\alpha(U_\alpha \cap U_\beta)$. A map $M \to \mathbb{R}$ is *smooth* if and only if its restriction to each U_α composed with Φ_α^{-1} is smooth. We denote by $C^\infty(M)$ the ring of smooth functions on M. (In a similar way one defines smooth maps between any two smooth manifolds: one asks that the map induces smooth maps between the corresponding D_αs.) A *metric* on M is a family (g_α) of metrics on the corresponding D_α such that for all α, β we have

$$\Phi_{\alpha\beta} g_\beta = g_\alpha$$

on $D_{\alpha\beta}$. By a *semi-Riemannian manifold* we understand a smooth manifold together with a metric. By the above theory, given a semi-Riemannian manifold, the family $(\mathcal{R}_\alpha)$ of curvatures attached to the corresponding Levi-Civita connections are compatible with the diffeomorphisms $\Phi_{\alpha\beta}$ and so is the family of Ricci tensors (Ric_α). We conclude that the scalar curvatures on each D_α define a smooth map $Scal \in C^\infty(M)$.

EXERCISE 37.38. Define parallel transport and geodesics on an arbitrary manifold.

REMARK 37.39. In Einstein's general theory of relativity the manifold M is the "space-time" $(n = 4)$ and for each point of $\mathbb{M}$ the metric can be represented in a suitable basis as a diagonal matrix with entries $1, -1, -1, -1$. The components of the metric are interpreted as giving the gravitational field and the equation defining the Einstein condition is Einstein's equation for the gravitational field in the absence of matter.

REMARK 37.40. An arithmetic analogue of Riemannian geometry is developed in: A. Buium, *Foundations of Arithmetic Differential Geometry*, Math. Surv. and Monographs, AMS, 2017.

CHAPTER 38

Orders

In the next two Chapters we return to number theory. Our main goal in the next Chapter is the Quadratic Reciprocity Law. The latter could not have been handled earlier because the proof we present needs some facts from analysis namely, the existence of the roots of unity (hence the exponential function or, alternatively, the Fundamental Theorem of Algebra). We start, in this Chapter, by introducing the concept of *algebraic integer* and the related concept of *order* in $\mathbb{C}$. The latter use of the word "order" has nothing to do with its use in the phrases "first order logic" or "order relations" or "order of a graph" or "order of an element in a group" or "order of a group."

DEFINITION 38.1. A complex number $u \in \mathbb{C}$ is called an *algebraic integer* if and only if there exists a monic polynomial $F \in \mathbb{Z}[x]$ such that $F(u) = 0$.

DEFINITION 38.2. A complex number $\alpha \in \mathbb{C}$ is called an *algebraic number* if and only if there exists a polynomial $F \in \mathbb{Z}[x]$ of degree ≥ 1 such that $F(\alpha) = 0$. A number $\alpha \in \mathbb{C}$ is called *transcendental* if and only if it is not algebraic.

So every algebraic integer is an algebraic number. The converse is not true.

EXAMPLE 38.3.
1) The numbers

$$2,\ \sqrt{2},\ \frac{-1+i\sqrt{3}}{2}$$

are algebraic integers because they are roots of

$$x-2,\ x^2-2,\ x^3-1,$$

respectively.
2) The numbers

$$\frac{1}{2},\ \frac{1}{2\sqrt{2}}$$

are algebraic numbers but not algebraic integers. (This follows from Propositions 38.11 and 38.15 below.)

REMARK 38.4. It is not clear that transcendental numbers exist. We will check that later.

EXERCISE 38.5. Prove that the set of algebraic numbers in $\mathbb{C}$ is countable.

EXAMPLE 38.6. Not all algebraic integers can be obtained from rational numbers by iterating the operations of addition, multiplication, and taking radicals of various orders; in order to prove the existence of algebriac integers that cannot be obtained in this way one needs "Galois theory".

In order to study the above concepts we need the following:

DEFINITION 38.7. A subset $\mathcal{O} \subset \mathbb{C}$ is called an *order* if:
1) $1 \in \mathcal{O}$
2) $u, v \in \mathcal{O}$ implies $u+v, uv, -u \in \mathcal{O}$;
3) There exist $u_1, ..., u_n \in \mathcal{O}$ such that

$$\mathcal{O} = \{m_1u_1 + ... + m_nu_n; m_1, ..., m_n \in \mathbb{Z}\}.$$

REMARK 38.8. Conditions 1 and 2 imply that $\mathcal{O}$ is a ring with respect to $+$ and $\times$.

EXERCISE 38.9. Prove that the sets

$$\mathbb{Z}[i], \ \{a + 2b\sqrt{-7}; a, b, \in \mathbb{Z}\}, \ \{a + 2b\sqrt{7}; a, b, \in \mathbb{Z}\}$$

are orders. Draw pictures of these sets.

PROPOSITION 38.10. *A complex number is an algebraic integer if and only if it is contained in an order.*

Proof. If u is an algebraic integer, root of a monic polynomial in $\mathbb{Z}[x]$ of degree n then u is contained in the order

$$\mathcal{O} := \{c_0 + c_1u + ... + c_{n-1}u^{n-1}; c_0, ..., c_{n-1} \in \mathbb{Z}\}.$$

Conversely assume u is contained in the order

$$\mathcal{O} = \{m_1u_1 + ... + m_nu_n; m_1, ..., m_n \in \mathbb{Z}\}.$$

Then for all $i = 1, ..., n$ we can write

$$uu_i = \sum_{j=1}^{n} m_{ij}u_j$$

with $m_{ij} \in \mathbb{Z}$. Set $a_{ij} = \delta_{ij}u - m_{ij}$ where δ_{ij} is 1 or 0 according as $i = j$ or $i \neq j$. Let $A = (a_{ij})$ be the matrix with entries a_{ij} and let U be the column vector with entries u_i. Since $AU = 0$ and $U \neq 0$ it follows that A is not invertible hence $det(A) = 0$. But $det(A)$ is easily seen to have the form

$$det(A) = u^n + a_1u^{n-1} + ... + a_{n-1}u + a_n$$

with $a_k \in \mathbb{Z}$ so u is an algebraic integer and we are done. □

PROPOSITION 38.11. *If u and u are algebraic integers then $u+v, uv, -u$ are also algebraic integers. In particular $\overline{\mathbb{Z}}$ is a ring with respect to $+$ and $\times$.*

Proof. Assume u belongs to the order

$$\{a_1u_1 + ... + a_nu_n; a_1, ..., a_n \in \mathbb{Z}\}$$

and v belongs to the order

$$\{b_1v_1 + ... + b_mv_m; b_1, ..., b_m \in \mathbb{Z}\}.$$

Then $u+v, uv, -u$ belong to the set

$$\left\{\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}u_iv_j; c_{ij} \in \mathbb{Z}\right\};$$

but this latter is set is clearly an order. □

DEFINITION 38.12. Denote by $\overline{\mathbb{Z}} \subset \mathbb{C}$ be the set of all algebraic integers. Denote by $\overline{\mathbb{Q}} \subset \mathbb{C}$ be the set of all algebraic numbers.

EXERCISE 38.13. Prove the following:
1) $\overline{\mathbb{Q}}$ equals the set

$$\mathbb{N}^{-1}\overline{\mathbb{Z}} := \{\frac{\alpha}{n} \mid n \in \mathbb{N},\ \alpha \in \overline{\mathbb{Z}}\}.$$

2) $\overline{\mathbb{Q}}$ is a field.
3) There exist transcendental numbers, i.e., $\overline{\mathbb{Q}} \neq \mathbb{C}$.

Hint for 1): If $\beta \in \overline{\mathbb{Q}}$ then $a_d\beta^d + a_{d-1}\beta^{d-1} + ... = 0$ with $a_d, a_{d-1}, ... \in \mathbb{Z}$. Multiplying by a_d^{d-1} we get that $(a_d\beta)^d + a_{d-1}a_d(a_d\beta)^{d-1} + ... = 0$ so $a_d\beta \in \overline{\mathbb{Z}}$, so $\beta \in \mathbb{N}^{-1}\overline{\mathbb{Z}}$. Conversely, if $\beta = \alpha/n \in \mathbb{N}^{-1}\overline{\mathbb{Z}}$ and $\alpha^d + a_{d-1}\alpha^{d-1} + ... = 0$, $a_i \in \mathbb{Z}$ then we divide by n^d the latter and conclude that $\beta \in \overline{\mathbb{Q}}$.

Hint for 2): By 1) $\overline{\mathbb{Q}}$ is a ring containing $\mathbb{Q}$. So it is enough to show that every $\alpha \in \mathbb{Z}$ is invertible in $\overline{\mathbb{Q}}$. For $\alpha \in \mathbb{Z}$ write $\alpha^d + a_{d-1}\alpha^{d-1} + ... + a_1\alpha + a_0 = 0$ with $a_0 \neq 0$; then dividing by α and expressing $\frac{a_0}{\alpha}$ in terms of the other terms we get $\frac{a_0}{\alpha} \in \overline{\mathbb{Q}}$, hence $\frac{1}{\alpha} \in \overline{\mathbb{Q}}$.

Hint for 3): $\overline{\mathbb{Q}}$ is countable while $\mathbb{C}$ is uncountable. This is Cantor's proof of the existence of transcendental numbers (which was known before him, with a different proof by Liouville).

EXERCISE 38.14. Let $\mathcal{O}$ be an order. Prove that the set

$$\mathbb{N}^{-1}\mathcal{O} := \{\frac{\alpha}{n} \mid n \in \mathbb{N},\ \alpha \in \mathcal{O}\}$$

is a field.

PROPOSITION 38.15. *A rational number which is also an algebraic integer must be an integer. In other words* $\overline{\mathbb{Z}} \cap \mathbb{Q} = \mathbb{Z}$.

Proof. Assume $\frac{a}{b} \in \mathbb{Q}$ is an algebraic integer,

$$\left(\frac{a}{b}\right)^d + a_{d-1}\left(\frac{a}{b}\right)^{n-1} + ... + a_0 = 0$$

with $a_{d-1}, ..., a_0 \in \mathbb{Z}$. Hence

$$a^d + a_{d-1}a^{d-1}b + ... + a_0b^d = 0.$$

Assume $\frac{a}{b} \notin \mathbb{Z}$. Then there exists a prime $p \in \mathbb{Z}$ with $p|b$ and $p \not| a$. But by the last equation if $p|b$ then $p|a^n$ hence $p|a$, a contradiction. $\square$

EXERCISE 38.16. Find an order containing $\sqrt{3} + \sqrt{7}$. Find a similar example involving cubic roots.

EXERCISE 38.17. Find a monic polynomial $f(x)$ in $\mathbb{Z}[x]$ such that $f(\sqrt{3}+\sqrt{7}) = 0$. Find a similar example involving cubic roots.

Some results about the integers appear as "shadows" of the arithmetic of algebraic integers. We give an example here related to the algebraic integer $i = \zeta_4$ which help prove results about the representation of primes as sums of two squares. In the next Chapter we will give an example related to roots of unity ζ_p of order a prime p which help prove the Quadratic Reciprocity Law.

DEFINITION 38.18. (Gauss) A *Gaussian integer* is a complex number of the form $a + bi \in \mathbb{C}$ with $a, b \in \mathbb{Z}$. We denote by $\mathbb{Z}[i]$ the set of Gaussian integers. It is an order in $\mathbb{C}$.

Definition 38.19. For every $u = a+bi$ the *conjugate* of u is defined as $\overline{u} = a-bi$ and the norm of u is defined as

$$N(u) = u\overline{u} = a^2 + b^2.$$

Exercise 38.20. Prove that for every $u, v \in \mathbb{Z}[i]$ we have:
1) $\overline{u+v} = \overline{u} + \overline{v}$, $\overline{u \times v} = \overline{u} \times \overline{v}$;
2) $N(uv) = N(u)N(v)$.
Hint: 1) is an easy computation. 2) follows from 1).

Definition 38.21. $u \in \mathbb{Z}[i]$ is called *invertible* if there exists $v \in \mathbb{Z}[i]$ such that $uv = 1$.

Proposition 38.22. *The invertible elements in* $\mathbb{Z}[i]$ *are* $1, -1, i, -i$.

Proof. Clearly $1, -1, i, -i$ are invertible; in fact $i(-i) = 1$. Conversely if u is invertible, hence $uv = 1$ it follows that $N(uv) = N(1) = 1$ hence $N(u)N(v) = 1$ hence $N(u) = 1$ which immediately implies u is one of $1, -1, i, -i$. $\square$

The following is an analogue of Euclid division:

Proposition 38.23. *For every* $u, v \in \mathbb{Z}[i]$ *with* $v \neq 0$ *there exist* $w, z \in \mathbb{Z}[i]$ *with* $u = vw + z$ *and* $N(z) < N(v)$. *(N.B.* w, z *are not unique.)*

Proof. Define $\mathbb{Q}(i)$ as $\mathbb{Q} \times \mathbb{Q}$ with addition and multiplication given by the same formulae as in the case of $\mathbb{Z}[i]$. Embed $\mathbb{Z}[i]$ into $\mathbb{Q}(i)$. Let $u\overline{v} = a + bi$ and let $t = \frac{a}{N(v)} + \frac{b}{N(v)}i \in \mathbb{Q}(i)$; so $tv = u$ in $\mathbb{Q}(i)$. View the points of $\mathbb{Q}(i) = \mathbb{Q} \times \mathbb{Q}$ as points in the "Euclidean plane". (The argument that follows can be made, of course, rigorous.) Then $\mathbb{Z}[i]$ can be viewed as the set of points in the plane with integer coordinates. So t will lie inside at least one square of side 1 whose vertices are in $\mathbb{Z}[i]$. There is at least one vertex of this square at distance less than 1 from t. (Any point in a square of side 1 is at distance less than 1 to one of the vertices.) We take that vertex to be w and define $z = uv - w$. Then it follows immediately that $N(z) < N(v)$. $\square$

Exercise 38.24. Make the above argument rigorous.

Hint: the vertices of the square can be defined using integral parts of rational numbers.

Definition 38.25. For $u, v \in \mathbb{Z}[i]$ we say that v *divides* u if there exists $w \in \mathbb{Z}[i]$ such that $u = vw$. A prime element in $\mathbb{Z}[i]$ is an element $\pi \in \mathbb{Z}[i]$ which is non-zero, non-invertible, and whenever $\pi = uv$ for $u, v \in \mathbb{Z}[i]$ it follows that either u or v is invertible.

The following is an analogue of Euclid's Lemma:

Proposition 38.26. *If* π *is a prime element in* $\mathbb{Z}[i]$ *and* $\pi | uv$ *with* $u, v \in \mathbb{Z}[i]$ *then either* $\pi | u$ *or* $\pi | v$.

Proof. As in the proof of Euclid's Lemma assume $\pi | uv$, $\pi \not| u$, $\pi \not| v$, and seek a contradiction. Consider the set

$$J = \{xu + y\pi; x, y, \in \mathbb{Z}[i]\}$$

and take an element $t \neq 0$ in J it whose norm is minimal. We claim that both u and π are divisible by t. This follows by dividing u and π by t with remainders as in

Proposition 38.23 and realizing the remainders belong to J hence by the minimality of the norm of t the remainders must be 0. Now since π is prime either t is invertible or t is an invertible element times π. The second case does not occur because it would imply that π divides u. So we conclude that t is invertible. We may assume $t = 1$. Then we can write $1 = xu + py$ with $x, y \in \mathbb{Z}[i]$. In exactly the same way (using v instead of u) we may write $1 = zv + w\pi$ with $z, w \in \mathbb{Z}[i]$. We get

$$1 = (xu + y\pi)(zv + w\pi)$$

and we conclude exactly as in the proof of Euclid's Lemma. □

EXERCISE 38.27. Prove that every element in $\mathbb{Z}[i]$ which is not zero and non-invertible can be written as a product of prime elements in $\mathbb{Z}[i]$.

Hint: Assume there are elements that don't have this property. Pick one of minimal norm and derive a contradiction.

Putting together Proposition 38.26 and Exercise 38.27 we get the following analogue of the Fundamental Theorem of Arithmetic:

THEOREM 38.28. *Every element u in $\mathbb{Z}[i]$ which is non-zero and non-invertible can be written as a product of prime elements in $\mathbb{Z}[i]$ such that if*

$$u = \pi_1 ... \pi_n = \pi'_1 ... \pi'_m$$

are two such representations then $n = m$ and (after a permutation of the indices) we have $\pi'_i = \epsilon_i \pi_i$ for some invertble elements ϵ_i.

EXERCISE 38.29. Write the details of the proof.

PROPOSITION 38.30. *Every prime p in $\mathbb{Z}$ with $p \equiv 3$ $(mod\ 4)$ is prime in $\mathbb{Z}[i]$.*

Proof. If $p = uv$ then $p^2 = N(p) = N(u)N(v)$ so either $N(u) = p$ or $N(u) = 1$ or $N(v) = 1$. In the last 2 cases we get u or v invertible. The case $N(u) = p$ does not occur because $N(u) = a^2 + b^2$ for integers a, b and we know that a sum of 2 squares in $\mathbb{Z}$ is never $\equiv 3$ $(mod\ 4)$. □

PROPOSITION 38.31. *If p is a prime in $\mathbb{Z}$ with $p \equiv 1$ $(mod\ 4)$ then p is not prime in $\mathbb{Z}[i]$ and in fact can be written as $p = \pi\overline{\pi} = N(\pi)$ with π a prime in $\mathbb{Z}[i]$.*

Proof. Recall that since $p \equiv 1$ $(mod\ 4)$ it follows that $p | c^2 + 1$ for some $c \in \mathbb{Z}$. Assume p is prime in $\mathbb{Z}[i]$ and seek a contradiction. Since $c^2 + 1 = (c + i)(c - i)$ it follows by Proposition 38.26 that either $p | c + i$ or $p | c - i$ in $\mathbb{Z}[i]$. But if $p | c + i$ then $c + i = p(a + bi)$ hence $c - i = \overline{c + i} = p(a - bi)$ so adding the last two equalities we get $2c = 2ap$ hence $p | c$, hence $p | 1$ a contradiction. In a similar way we get a contradiction assuming $p | c - i$. We proved that p is not prime in $\mathbb{Z}[i]$. Then, by Exerrcise 38.27 we can write

$$p = \pi_1 ... \pi_s$$

with $s \geq 2$ and π_i prime. Taking norms we get

$$p^2 = N(p) = N(\pi_1) ... N(\pi_s).$$

Since the left hand side has only 2 primes in its prime decomposition and none of the factors in the right hand side is 1 it follows that $s = 2$ and $N(\pi_1) = N(\pi_2) = p$. So $p = N(\pi_1) = \pi_1 \overline{\pi_1}$. So $\pi_2 = \overline{\pi_1}$ and we are done. □

EXAMPLE 38.32. 5 is not prime in $\mathbb{Z}[i]$ because $5 = (2 + i)(2 - i)$. 7 is prime in $\mathbb{Z}[i]$.

EXERCISE 38.33. Find the prime factorization in $\mathbb{Z}[i]$ of $29^5 \times 37^3 \times 23^7$.

EXERCISE 38.34. Prove that $2+3i$ is prime in $\mathbb{Z}[i]$.

EXERCISE 38.35. Find the prime factorization in $\mathbb{Z}[i]$ of the number $12+13i$.

Since in Proposition 38.31 $N(\pi)$ is a sum of squares in $\mathbb{Z}$ we obtain a proof of the following:

THEOREM 38.36. *(Fermat). If p is a prime in $\mathbb{Z}$ with $p \equiv 1 \ (mod\ 4)$ then $p = a^2 + b^2$ for some integers $a, b \in \mathbb{Z}$.*

The arithmetic of general orders is much more complicated than that of $\mathbb{Z}$. This was realized in the 19th century by Kummer, Dedekind, and others. In particular the fundamental theorem of arithmetic may fail in certain orders, as we will see here.

DEFINITION 38.37. An element u in an order $\mathcal{O}$ is called invertible if there exists $v \in \mathcal{O}$ such that $uv = 1$. An element $u \in \mathcal{O}$ is called irreducible if whenever $u = vw$ with $v, w \in \mathcal{O}$ it follows that either v or w is invertible. Two irreducible elements u and v in $\mathcal{O}$ are called associated in divisibility if $u = vw$ with w invertible.

One is tempted to use the word *prime* instead of *irreducible*; but in view of pathologies to be put forward soon one prefers the work *irreducible*.

EXERCISE 38.38. Prove that in the order $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5}; a, b \in \mathbb{Z}\}$ (called the ring of Kummer integers) the following hold. (Morally the Fundamental Theorem of Arithmetic fails in this order.)

1) The only invertible elements in $\mathbb{Z}[\sqrt{-5}]$ are 1 and -1;

2) The elements $2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}$ are irreducible and no two of them are associated in divisibility;

3) The element 6 has the following 2 decompositions:

$$6 = 2 \times 3 = (1+\sqrt{-5})(1-\sqrt{-5}).$$

Hint: Define the norm $N(u) = u\overline{u} = a^2 + 5b^2$ for $u = a + b\sqrt{-5}$. Prove that u is invertible if and only if it has norm 1 which proves 1). To prove 2) assume one of these elements u can be written as $u = vw$ with v, w non-invertible, take norms to get $N(v)N(w)$ is $4, 6$, or 9, conclude that $N(v)$ is 2 or 3, and derive a contradiction. 3) is clear.

REMARK 38.39. One of the interesting orders that were carefully studied in the 19th century (especially by Kummer) and are still an object of study is the smallest order containing a given root of unity $\zeta_N = \exp(2\pi i/N) \in \overline{\mathbb{Z}}$, $N \geq 3$, namely:

$$\mathbb{Z}[\zeta_N] := \{\sum_{i=0}^{N-1} a_i \zeta_N^i \mid a_0, ..., a_{N-1} \in \mathbb{Z}\}.$$

For $N = 4$ we have $\mathbb{Z}[\zeta_N] = \mathbb{Z}[i]$, the ring of Gauss integers. For $N = 3$ one still has an analogue of the Fundamental Theorem of Arithmetic. However the Fundamental Theorem of Arithmetic fails in $\mathbb{Z}[\zeta_N]$ for N sufficiently big.

CHAPTER 39

Reciprocity

In this Chapter we prove the Quadratic Reciprocity Theorem of Gauss; this was partially proved before Gauss by Legendre and essentially conjectured before Legendre by Euler. The proof we present here is based on the study of roots of unity. In this Chapter we let p be a prime $\neq 2$ in $\mathbb{Z}$.

DEFINITION 39.1. If $a \in \mathbb{Z}$ define the *Legendre symbol*

$$\left(\frac{a}{p}\right) = N_p(x^2 = a) - 1,$$

i.e. the Legendre symbol is $-1, 0, 1$ according as $x^2 \equiv \ a \ (mod\ p)$ has 2 solutions, one solution (this is the case if and only if $p|a$), or no solution respectively, in $\{0, ..., p-1\}$.

REMARK 39.2. The map $s : \mathbb{F}_p^\times \to \mathbb{F}_p^\times$, $x \mapsto x^2$, is a group homomorphism with kernel $\{\pm 1\}$. Then $\left(\frac{a}{p}\right) = 1$ if and only if the image of a in $\mathbb{F}_p$ is in the image of the map s.

EXERCISE 39.3. Prove that

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Hint: Use Remark 39.2.

LEMMA 39.4. *(Euler).*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \ (mod\ p).$$

Proof. Since by Exercise 39.3 both sides are "multiplicative" in a it is enough to check this for a a primitive root mod p which follows from Remark 39.2. $\square$

COROLLARY 39.5. *(Euler).* $\left(\frac{-1}{p}\right) = 1$ *if and only if* $p \equiv 1 \ (mod\ 4)$.

LEMMA 39.6. *(Euler)* $\left(\frac{2}{p}\right) = 1$ *if and only if* $p \equiv 1, 7 \ (mod\ 8)$.

Proof. First we claim that $\left(\frac{2}{p}\right) = (-1)^\mu$ where μ is the number of integers in the set $\{2, 4, 6, ..., p-1\}$ congruent mod p to a negative integer between $-\frac{p}{2}$ and $\frac{p}{2}$. Indeed let $r_1, ..., r_{\frac{p-1}{2}}$ be the integers between $-\frac{p}{2}$ and $\frac{p}{2}$ that are congruent mod p to $2, 4, 6, ..., p-1$. Then it is easy to check that

$$\{|r_1|, ..., |r_{\frac{p-1}{2}}|\} = \{1, 2, 3, ..., \frac{p-1}{2}\},$$

where $|r|$ is the absolute value of r i.e. r or $-r$ according as r is positive or negative. Taking products we get

$$1 \times 2 \times 3 \times ... \times \frac{p-1}{2} \equiv (-1)^{\mu} \times 2^{\frac{p-1}{2}} \times 1 \times 2 \times 3 \times ... \times \frac{p-1}{2} \quad (mod\ p)$$

which proves our claim.

Now note that if an integer a between $-\frac{p}{2}$ and 0 is congruent mod p to one of the numbers $2, 4, 6, ..., p-1$ then $2x \equiv a\ (mod\ p)$ for some $x \in \{1, 2, 3, ..., \frac{p-1}{2}\}$. Writing $a = 2x + mp$ we get $-\frac{p}{2} < 2x + mp < 0$ hence $\frac{p}{2} < 2x + (m+1)p < p$ which forces $m = -1$ hence $\frac{p}{2} < 2x < p$. Conversely if the latter holds then $a = 2x - p$ is between $-\frac{p}{2}$ and 0. So if $p = 8k + r$, $0 \leq r < 7$, we have

$$\begin{aligned} \mu &= |\{x \in \mathbb{Z}; \tfrac{p}{2} < 2x < p\}| \\ &= |\{x \in \mathbb{Z}; \tfrac{p}{4} < x < \tfrac{p}{2}\}| \\ &= |\{x \in \mathbb{Z}; 2k + \tfrac{r}{4} < x < 4k + \tfrac{r}{2}\}| \\ &= |\{x \in \mathbb{Z}; \tfrac{r}{4} < x < 2k + \tfrac{r}{2}\}| \end{aligned} .$$

and we conclude by inspecting the values $r = 1, 3, 5, 7$. $\square$

Exercise 39.7.

$$\sum_{a=1}^{p-1} \left(\frac{a}{p}\right) = 0.$$

Hint: Use Remark 39.2.

Exercise 39.8. Let $E(\mathbb{F}_p)$ be the elliptic curve over $\mathbb{F}_p$ attached to the cubic equation $y^2 = x^3 + \overline{a}x + \overline{b}$ where $a, b \in \mathbb{Z}$. Prove that the cardinality (order) of the group $E(\mathbb{F}_p)$ is given by

$$|E(\mathbb{F}_p)| = p + 1 + \sum_{x=0}^{p-1} \left(\frac{x^3 + ax + b}{p}\right).$$

The main result pertaining to the Legendre symbol is the following theorem of Gauss (partially proved before him by Legendre):

Theorem 39.9. *(Quadratic Reciprocity Law). For every two distinct primes p and q different from 2 we have*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Theorem 39.9 plus Lemma 39.6 imply:

Corollary 39.10. *If $a \in \mathbb{N}$ and p_1, p_2 are primes such that*

$$p_1 \equiv p_2\ (mod\ 4a)$$

then

$$N_{p_1}(x^2 = a) = N_{p_2}(x^2 = a).$$

Exercise 39.11. Prove Corollary 39.10.

Remark 39.12. Gauss gave several proofs of Theorem 39.9 and a lot of the 19th century and early 20th century number theory was devoted to trying to generalize (and hence better understand) this statement. The essence of this statement is now understood to be captured by Corollary 39.10 which is viewed as a "periodicity" result. Indeed if we fix the polynomial $f(x) = x^2 - a$ then the value of $N_p(f)$ only depends on the remainder $r_N(p)$ when p is divided by N for an integer N depending on f (in our case $N = 4a$). This can be viewed as saying that the function

$$p \mapsto N_p(f)$$

is "periodic" with period N. Such a statement fails, in general, for polynomials f of arbitrary degree (although there are examples of polynomials of higher degree, all related, in a certain sense, to the roots of unity, for which such a statement holds).

To prove Theorem 39.9 recall first the p-th root of unity,

$$\zeta_p = \exp\left(\frac{2\pi i}{p}\right) \in \overline{\mathbb{Z}}.$$

Exercise 39.13. Prove that if c is an integer then

$$\sum_{b=1}^{p-1} (\zeta_p^c)^b$$

equals $p-1$ or -1 according as $p|c$ or $p \not| c$.

Definition 39.14. Define the *Gauss sum*

$$G = \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \zeta_p^a \in \overline{\mathbb{Z}}.$$

Lemma 39.15. *(Gauss).*

$$G^2 = (-1)^{\frac{p-1}{2}} p.$$

Proof. We have

$$G^2 = \sum_{a=1}^{p-1} \sum_{b=1}^{p-1} \left(\frac{ab}{p}\right) \zeta_p^{a+b}.$$

Denote by $r_p(x)$ the remainder when an integer x is divided by p. If (a,b) runs through the set of indices of the above sum then clearly $(r_p(ab), b)$ runs through the same set of indices so substituting a by ab and noting that

$$\zeta_p^{ab} = \zeta_p^{r_p(ab)}$$

we get that the above sum equals

$$\sum_{a=1}^{p-1} \sum_{b=1}^{p-1} \left(\frac{ab^2}{p}\right) \zeta_p^{ab+b} = \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \sum_{b=1}^{p-1} (\zeta_p^{a+1})^b.$$

In view of Exercises 39.13 and 39.7 the above sum equals

$$\left(\frac{-1}{p}\right)(p-1) - \sum_{a=1}^{p-2} \left(\frac{a}{p}\right) = \left(\frac{-1}{p}\right) p$$

and we are done by Lemma 39.4. $\square$

Exercise 39.16. Prove that every square root of a non-zero integer is a sum of roots of unity in $\mathbb{C}$.

DEFINITION 39.17. For $u, v \in \overline{\mathbb{Z}}$ and q a prime in $\mathbb{Z}$ let us write $u \equiv v$ $(mod\ q)$ in $\overline{\mathbb{Z}}$ if there exists $w \in \overline{\mathbb{Z}}$ such that $qw = v - u$.

EXERCISE 39.18. Prove that if $u \equiv v$ $(mod\ q)$ in $\overline{\mathbb{Z}}$ and $u, v \in \mathbb{Z}$ then $u \equiv v$ $(mod\ q)$ in $\mathbb{Z}$.

Hint: This follows directly from Proposition 38.15.

EXERCISE 39.19. (Freshman's Dream) Prove that

$$(u_1 + ... + u_n)^p \equiv u_1^p + ... + u_n^p \ (mod\ p) \text{ in } \overline{\mathbb{Z}}$$

for $u_1, ..., u_n \in \overline{\mathbb{Z}}$ and p a prime in $\mathbb{Z}$.

Proof of Theorem 39.9. By Lemma 39.15 and then Lemma 39.4

$$G^q = G(G^2)^{\frac{q-1}{2}} = G(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}} \equiv G(-1)^{\frac{p-1}{2}\frac{q-1}{2}} \left(\frac{p}{q}\right) \ (mod\ q) \text{ in } \overline{\mathbb{Z}}.$$

On the other hand by "Freshman's Dream" we get

$$G^q = \left(\sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \zeta_p^a\right)^q \equiv \sum_{a=1}^{p-1} \left(\frac{a}{p}\right)^q \zeta_p^{aq} = \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \zeta_p^{aq}$$

$$= \left(\frac{q}{p}\right) \sum_{a=1}^{p-1} \left(\frac{aq}{p}\right) \zeta_p^{aq} = \left(\frac{q}{p}\right) G \ (mod\ q) \text{ in } \overline{\mathbb{Z}}.$$

The two expressions of G^q above give

$$G(-1)^{\frac{p-1}{2}\frac{q-1}{2}} \left(\frac{p}{q}\right) \equiv \left(\frac{q}{p}\right) G \ (mod\ q) \text{ in } \overline{\mathbb{Z}}$$

Assume

$$(-1)^{\frac{p-1}{2}\frac{q-1}{2}} \left(\frac{p}{q}\right) \neq \left(\frac{q}{p}\right),$$

and let us derive a contradiction. Since the two numbers above are ± 1 we get that one is 1 and the other is -1 so we get

$$G \equiv -G \quad (mod\ q) \text{ in } \overline{\mathbb{Z}}$$

hence

$$2G \equiv 0 \quad (mod\ q) \text{ in } \overline{\mathbb{Z}}$$

Squaring we get

$$4p \equiv 0 \quad (mod\ q) \text{ in } \overline{\mathbb{Z}}$$

and hence, by Exercise 39.18,

$$4p \equiv 0 \quad (mod\ q) \text{ in } \mathbb{Z}$$

which is a contradiction. $\square$

CHAPTER 40

Categories

Categories are one of the most important unifying concepts of Mathematics. In particular they allow to create bridges between various parts of Mathematics. Categories were introduced by Eilenberg and MacLane partly motivated by work in homological algebra. A further input in the development of the concept was given by work of Grothendieck in algebraic geometry. Here we will only explore the definition of categories and we give some examples.

Definition 40.1. A *category* is a tuple $\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$ where $X^{(0)}$ and $X^{(1)}$ are sets, $\sigma, \tau : X^{(1)} \to X^{(0)}$ are maps and, upon considering the sets

$$\begin{array}{rcl} X^{(2)} & = & \{(a,b) \in X^2 \mid \tau(b) = \sigma(a)\} \\ X^{(3)} & = & \{(a,b,c) \in X^3 \mid \tau(c) = \sigma(b), \tau(b) = \sigma(a)\}, \end{array}$$

and the maps $p_1, p_2 : X^{(1)} \to X^{(0)}$, $p_1(a,b) = a$, $p_2(a,b) = b$, we have that $\mu : X^{(2)} \to X^{(1)}$ and $\epsilon : X^{(0)} \to X^{(1)}$ are maps satisfying the following properties:

1) We have $\sigma \circ \epsilon = \tau \circ \epsilon = I$, the identity of $X^{(0)}$.

2) The following diagrams are commutative:

$$\begin{array}{ccc} X^{(2)} & \stackrel{\mu}{\to} & X^{(1)} \\ p_1 \downarrow & & \downarrow \tau \\ X^{(1)} & \stackrel{\tau}{\to} & X^{(0)} \end{array}, \quad \begin{array}{ccc} X^{(2)} & \stackrel{\mu}{\to} & X^{(1)} \\ p_2 \downarrow & & \downarrow \sigma \\ X^{(1)} & \stackrel{\sigma}{\to} & X^{(0)} \end{array}, \quad \begin{array}{ccc} X^{(3)} & \stackrel{\mu \times 1}{\to} & X^{(2)} \\ 1 \times \mu \downarrow & & \downarrow \mu \\ X^{(2)} & \stackrel{\mu}{\to} & X^{(0)} \end{array}.$$

3) The compositions

$$X^{(1)} \xrightarrow{1 \times (\epsilon \circ \sigma)} X^{(2)} \stackrel{\mu}{\to} X^{(1)}$$

and

$$X^{(1)} \xrightarrow{(\epsilon \circ \tau) \times 1} X^{(2)} \stackrel{\mu}{\to} X^{(1)}$$

are the identity of $X^{(1)}$.

Here

$$(1 \times (\epsilon \circ \sigma))(a) = (a, \epsilon(\sigma(a))),$$
$$((\epsilon \circ \tau) \times 1)(a) = (\epsilon(\tau(a)), a).$$

Remark 40.2. The set $X^{(0)}$ is called the set of *objects* of the category and is also denoted by $Ob(\mathcal{C})$. The set $X^{(1)}$ is called the set of *arrows* or *morphisms* and is sometimes denoted by $Mor(\mathcal{C})$. The map μ is called *composition* and we write $\mu(a,b) = a \star b$. The maps σ and τ are called the source and the target map, respectively. The map ϵ is called the identity. We set $\epsilon(x) = 1_x$ for all x. The first commutative diagram says that the target of $a \star b$ is the target of a. The second diagram says that the source of $a \star b$ is the source of b. In the third diagram (called *associativity diagram*) the map $\mu \times 1$ is defined as $(a,b,c) \mapsto (a \star b, c)$ while the map

$1\times\mu$ is defined by $(a,b,c)\mapsto(a,b\star c)$; the diagram then says that $(a\star b)\star c = a\star(b\star c)$. For $x,y\in X^{(0)}$ one denotes by $Hom(x,y)$ the set of all morphisms $a\in X^{(1)}$ with $\sigma(a)=x$ and $\tau(a)=y$. Instead of $a\in Hom(x,y)$ we also write $a : x\to y$. We say $a\in Hom(x,y)$ is an isomorphism if there exists $a'\in Hom(y,x)$ such that $a\star a' = 1_y$ and $a'\star a = 1_x$. (Then a' is unique and is denoted by a^{-1}.) A category is called a *groupoid* if all morphisms are isomorphisms.

Remark 40.3. What we called a *category* in the above definition is sometimes called a *small* category; since our presentation will involve universes (see below) rather than "classes" (as in Eilenberg-MacLane) we do not need to make any distinction between small categories and categories.

Definition 40.4. Given a category

$$\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$$

the *opposite category* is defined as

$$\mathcal{C}^{\circ} = (X^{(0)}, X^{(1)}, \tau, \sigma, \mu\circ S, \epsilon)$$

where $S(a,b)=(b,a)$.

Definition 40.5. Given a category

$$\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$$

and a subset $Y\subset X^{(0)}$ one defines the *full subcategory* associated to Y by

$$\mathcal{C}_Y = (Y^{(0)}, Y^{(1)}, \tau_Y, \sigma_Y, \mu_Y, \epsilon_Y)$$

where $Y^{(0)} = Y$, $Y^{(1)} = \{a\in X^{(0)} \mid \sigma(a)\in Y, \tau(a)\in Y\}$, and $\tau_Y, \sigma_Y, \mu_Y, \epsilon_Y$ defined as the restrictions of τ,σ,μ,ϵ, respectively.

In what follows we give some basic examples of categories. For the various examples below that involve *universes* the correctness of the definitions depends on certain claims (typically that certain sets are maps, etc.) Those claims cannot be proved a priori from ZFC and one typically needs an extra axiom to make things work. So we will add to ZFC the following:

Axiom 40.6. (Grothendieck's axiom of the universes) There exists u such that all the axioms of ZFC hold if we add to their statements the proviso that the variables belong to u. A witness $\mathcal{U}$ for this axiom is called a *universe*.

Example 40.7. Let $\mathcal{U}$ be a universe. Define the sets and maps:

$X^{(0)} = \mathcal{U}$,
$X^{(1)} = \{(A,B,F)\in\mathcal{U}^3 \mid F\in Fun(A,B)\}$,
$\sigma(A,B,F) = A$, $\quad\tau(A,B,F) = B$.
$\epsilon(A) = I_A$, (identity of A)

We get a category; indeed by the axiom of the universes if $A,B\in\mathcal{U}$ we have $Fun(A,B)\in\mathcal{U}$ (check!) which insures that composition of morphisms is well defined. In the examples below we will encounter from time to time the same kind of phenomenon (where the correctness of definitions depends on the axiom of universes; we will not repeat the corresponding discussion but we will simply add everywhere the words "in a given universe."

EXAMPLE 40.8. If in the example above we insist that all Fs are bijections we get a category called

$$\{\text{sets} + \text{bijections}\}.$$

This category is a groupoid.

EXAMPLE 40.9. Let A be a set and consider the category denoted by

$$\{\text{bijections of } A\}$$

defined as follows. We let $X^{(0)} = \{x\}$ be a set with one element, we let $X^{(1)}$ be the set of all bijections $F : A \to A$, we let σ and τ be the constant map $F \mapsto x$, we let μ be defined again by $\mu(F, G) = F \circ G$ (compositions of functions), and we let $\epsilon(F) = I_A$. This category is a groupoid.

EXAMPLE 40.10. Define the category

$$\{\text{ordered sets}\}$$

as follows. We take $X^{(0)}$ the set of all ordered sets $(A, \leq)$ with A in a given universe, we take $X^{(1)}$ to be the set of all triples $((A, \leq), (A', \leq'), F)$ with $(A, \leq), (A', \leq') \in X^{(0)}$ and $F : A \to A'$ increasing, we take μ to be again, composition, and we take $\epsilon(A, \leq) = I_A$.

EXAMPLE 40.11. Let $(A, \leq)$ be an ordered set. Define the category

$$\{(A, \leq)\}$$

as follows. We let $X^{(0)} = A$, we let $X^{(1)}$ be the set $\leq$ viewed as a subset of $A \times A$, we let $\sigma(a, b) = a$, $\tau(a, b) = b$, we let $\mu((a, b), (b, c)) = (a, c)$, and we let $\epsilon(a) = (a, a)$.

EXAMPLE 40.12. Equivalence relations give rise to groupoids. Indeed let A be a set with an equivalence relation $R \subset A \times A$ on it which we refer to as $\sim$. Define the category

$$\{(A, \sim)\}$$

as follows. We let $X^{(0)} = A$, we let $X^{(1)} = R$, we let $\sigma(a, b) = a$, $\tau(a, b) = b$, we let $\mu((a, b), (b, c)) = (a, c)$, and we let $\epsilon(a) = (a, a)$. This category is a groupoid.

EXAMPLE 40.13. We fix the type of algebraic structures below. (For instance we may fix two binary operations, one unary operation, and two given elements.) Define the category

$$\{\text{algebraic structures}\}$$

as follows. $X^{(0)}$ is the set of all algebraic structures $(A, \star, ..., \neg, ..., 1, ...)$ of the given type with A in a given universe, $X^{(1)}$ is the set of all triples

$$((A, \star, ..., \neg, ..., 1, ...), (A', \star', ..., \neg', ..., 1', ...), F)$$

with F a homomorphism, σ and τ are the usual source and target, and ϵ is the usual identity.

EXAMPLE 40.14. Here is a variant of the above example. Consider the category of rings

$$\{\text{commutative unital rings}\}$$

as follows. The set of objects $X^{(0)}$ is the set of all commutative unital rings $(A, +\times, 0, 1)$ (usually referred to as A) in a given universe and the set of arrows $X^{(1)}$ is the set of all triples (A, B, F) where A, B are rings and $F : A \to B$ is a ring homomorphism. Also the target, source, and identity are the obvious ones; the composition map is the usual composition.

EXAMPLE 40.15. Define the category

$$\{\text{topological spaces}\}$$

as follows. We let $X^{(0)}$ be the set of all topological spaces X in a given universe; we take $X^{(1)}$ to be the set of all triples (X, X', F) with $F : X \to X'$ continuous, we let μ be given by usual composition of maps, σ and τ the usual source and target maps, and ϵ the usual identity.

EXAMPLE 40.16. Here is a variation on the previous example. Define the category

$$\{\text{pointed topological spaces}\}$$

as follows. We let $X^{(0)}$ be the set of all pairs (X, x) where X is a topological space in a given universe and $x \in X$; we take $X^{(1)}$ to be the set of all triples $((X, x), (X', x'), F)$ with $F : X \to X'$ continuous, and $F(x) = x'$; we let μ be given by usual composition of maps, σ and τ the usual source and target maps, and ϵ the usual identity.

EXAMPLE 40.17. Define the category of groups

$$\{\text{groups}\}$$

defined as follows. The set $X^{(0)}$ of objects is the set of all groups whose set is in a given universe, and the set $X^{(1)}$ is the set of all the triples consisting of two groups G, H and a homomorphism between them. The source, target, and identity are the obvious ones.

EXAMPLE 40.18. If in the above example we restrict ourselves to groups which are Abelian we get the category of Abelian groups

$$\{\text{Abelian groups}\}.$$

EXAMPLE 40.19. A fixed group can be viewed as a category as follows. Fix a group $(G, \star, ', e)$. Then one can consider the category

$$\{G\}$$

where $X^{(0)} = \{x\}$ is a set consisting of one element, $X^{(1)} = G$, $\mu(a, b) = a \star b$, the source and the target are the constant maps, and $\epsilon(x) = e$.

EXAMPLE 40.20. Define the category

$$\{\text{vector spaces}\}$$

as follows. The set of objects $X^{(0)}$ is the set of all vector spaces, in a given universe, over a fixed field; the set $X^{(1)}$ of morphisms consists of all triples (V, W, F) with $F : V \to W$ a linear map; source, target, and identity are defined in the obvious way.

EXAMPLE 40.21. Define the category

$$\{\text{complex affine algebraic varieties}\}$$

as follows. The objects of the category (called *complex affine algebraic varieties*) are pairs $(\mathbb{C}^n, X)$ where X is a subset $X \subset \mathbb{C}^n$ for which there exist polynomials $f_1, ..., f_m \in \mathbb{C}[x_1, ..., x_n]$ such that

$$X = \{(a_1, ..., a_n) \in \mathbb{C}^n \mid f_1(a_1, ..., a_n) = ... = f_m(a_1, ..., a_n) = 0\}.$$

(Lines, conics, and cubics introduced earlier are examples of complex affine algebraic varieties if one takes $R = \mathbb{C}$.) A morphism between $(\mathbb{C}^n, X)$ and $(\mathbb{C}^{n'}, X')$ is a map $F : X \to X'$ such that there exist polynomials $F_1, ..., F_{n'} \in \mathbb{C}[x_1, ..., x_n]$ with the property that for every $(a_1, ..., a_n) \in X$ we have

$$F(a_1, ..., a_n) = (F_1(a_1, ..., a_n), ..., F_{n'}(a_1, ..., a_n)).$$

Composition of morphisms is composition of maps.

Example 40.22. Define the category

$$\{\text{ordinary differential equations}\}$$

as follows. The set of objects consists of all pairs $(\mathbb{R}^n, V)$ where $V : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ is a derivation which is $\mathbb{R}$-linear. (Such a V is called a *vector field* on $\mathbb{R}^n$.) A morphism $(\mathbb{R}^n, V) \to (\mathbb{R}^m, W)$ is a map $u : \mathbb{R}^n \to \mathbb{R}^m$ with smooth components such that for all $f \in C^\infty(\mathbb{R}^m)$ we have $V(f \circ u) = W(f) \circ u$. The reason why this is viewed as the right category for the theory of ordinary differential equations will be seen later.

Exercise 40.23. Check that, in all examples above, the conditions in the definition of a category are satisfied. (This is long and tedious but straightforward.)

Each category describes, in some sense, the paradigm for one area of Mathematics. The bridges between the various areas are realized by functors, cf. the definition and examples below.

Definition 40.24. A *functor* $\Phi : \mathcal{C} \to \tilde{\mathcal{C}}$ between two categories

$$\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon), \quad \text{and} \quad \tilde{\mathcal{C}} = (\tilde{X}^{(0)}, \tilde{X}^{(1)}, \tilde{\sigma}, \tilde{\tau}, \tilde{\mu}, \tilde{\epsilon})$$

is a pair of maps $(\Phi^{(0)}, \Phi^{(1)})$,

$$\Phi^{(0)} : X^{(0)} \to \tilde{X}^{(0)}, \quad \Phi^{(1)} : X^{(1)} \to \tilde{X}^{(1)}$$

such that the following diagrams are commutative:

$$\begin{array}{ccc} X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} \\ \sigma \downarrow & & \downarrow \tilde{\sigma} \\ X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} \end{array}, \quad \begin{array}{ccc} X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} \\ \tau \downarrow & & \downarrow \tilde{\tau} \\ X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} \end{array}, \quad \begin{array}{ccc} X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} \\ \epsilon \uparrow & & \uparrow \tilde{\epsilon} \\ X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} \end{array}$$

$$\begin{array}{ccc} X^{(2)} & \xrightarrow{\Phi^{(2)}} & \tilde{X}^{(2)} \\ \mu \downarrow & & \downarrow \tilde{\mu} \\ X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} \end{array}$$

where $\Phi^{(2)} : X^{(2)} \to \tilde{X}^{(2)}$ is the naturally induced map. One usually denotes both $\Phi^{(0)}$ and $\Phi^{(1)}$ by Φ. So compatibility with μ and $\tilde{\mu}$ reads

$$\Phi(a \star b) = \Phi(a) \star \Phi(b),$$

for all $(a, b) \in X^{(2)}$.

Here are a few examples of functors. We start with some "forgetful" functors whose effect is to "forget" part of the structure:

EXAMPLE 40.25. Consider the "forgetful" functor

$$\Phi : \{\text{commutative unital rings}\} \to \{\text{Abelian groups}\}$$

defined as follows. For $(R, +, \times, -, 0, 1)$ a commutative unital ring we let

$$\Phi(R, +, \times, -, 0, 1) = (R, +, -, 0),$$

which is an Abelian group. For every ring homomorphism F we set $\Phi(F) = F$, viewed as a group homomorphism.

EXAMPLE 40.26. Consider the "forgetful" functor

$$\Phi : \{\text{commutative unital rings}\} \to \{\text{Abelian groups}\}$$

defined as follows. For $(R, +, \times, -, 0, 1)$ a commutative unital ring we let

$$\Phi(R, +, \times, -, 0, 1) = (R^\times, \times, (\ )^{-1}, 1),$$

where $R^\times$ is the group of invertible elements of R, which is an Abelian group, and x^{-1} is the inverse of x. For every ring homomorphism F we let $\Phi(F)$ be the restriction of F to the invertible elements.

EXAMPLE 40.27. Consider the functor

$$\Phi : \{\text{Abelian groups}\} \to \{\text{groups}\}$$

defined as follows. For $(G, \star, ', e)$ an Abelian group we let

$$\Phi(G, \star, ', e) = (G, \star, ', e).$$

For every group homomorphism F we let $\Phi(F) = F$.

EXAMPLE 40.28. Consider the "forgetful" functor

$$\Phi : \{\text{groups}\} \to \{\text{sets}\}$$

defined as follows. For $(G, \star, ', e)$ a group we let

$$\Phi(G, \star, ', e) = G.$$

For every group homomorphism F we let $\Phi(F) = F$, as a map of sets.

EXAMPLE 40.29. Consider the following functor that is the prototype for some important functors in areas of Mathematics called *functional analysis* and *algebraic geometry*. The functor is

$$\Phi : \{\text{topological spaces}\} \to \{\text{commutative unital rings}\}^\circ,$$

it takes values in the opposite of the category of commutative unital rings, and is defined as follows. For X a topological space we let

$$\Phi(X) = (C^0(X), +, \cdot, -, 0, 1),$$

where the latter is the following ring. The set $C^0(X)$ is the set of all continuous functions $f : X \to \mathbb{R}$, the addition $+$ and multiplication $\cdot$ are the pointwise operations, and $0(x) = 0$, $1(x) = 1$. For every continuous map $F : X \to X'$ we let $\Phi(F) = F^*$ where, for $f : X' \to \mathbb{R}$, $F^*(f) = f \circ F$.

EXERCISE 40.30. Prove that all Φs in the examples above are functors. (This is tedious but straightforward.)

EXAMPLE 40.31. Consider the following functor that, again, is the prototype for some important functors in functional analysis and algebraic geometry. The functor is

$$\Phi : \{\text{commutative unital rings}\}^{\circ} \to \{\text{topological spaces}\},$$

and is defined as follows. Consider any commutative unital ring R. By an ideal in R we understand a subset $I \subset R$ such that I is a subgroup of R with respect with addition (i.e., $0 \in I$, $a + b \in I$, and $-a \in I$ for all $a, b \in I$), and $ab \in I$ for all $a \in R$ and $b \in I$. An ideal P in R is called a *prime ideal* if $P \neq R$ and whenever $ab \in P$ with $a \in R$ and $b \in R$ it follows that either $a \in P$ or $b \in P$. We let *Spec* R be the set of all prime ideals in R. For any ideal I in R we let $D(I) \subset \mathit{Spec}\ R$ be the set of all prime ideals P such that $I \not\subset P$. Then the collection of all subsets of the form $D(I) \in \mathcal{P}(\mathit{Spec}\ R)$ is a topology on *Spec* R called the *Zariski topology*. With this topology *Spec* R becomes a topological space and we define

$$\Phi(R) = \mathit{Spec}\ R.$$

If $F : R \to R'$ is a ring homomorphism we define $\Phi(F) = F^* : \mathit{Spec}\ R' \to \mathit{Spec}\ R$ by $F^*(P') = F^{-1}(P)$.

EXERCISE 40.32. Prove that the collection $\{D(I) \mid I \text{ an ideal in R}\}$ is a topology on *Spec* R. Prove that F^* is continuous. Prove that Φ is a functor.

EXERCISE 40.33. Prove that the ideals of $\mathbb{Z}$ are exactly the subgroups of $\mathbb{Z}$; hence they are of the form $\langle n \rangle$ for $n \in \mathbb{Z}$, $n \geq 0$. Prove that $\langle n \rangle$ is a prime ideal if and only if n is prime or $n = 0$. Prove that *Spec* $\mathbb{Z}$ is not a Hausdorff space.

EXERCISE 40.34. Consider the following functor that plays a key role in algebraic geometry. The functor is

$$\Phi : \{\text{complex affine algebraic varieties}\} \to \{\text{topological spaces}\},$$

and is defined as follows. If $(\mathbb{C}^n, X)$ is a complex affine algebraic variety, $X \subset \mathbb{C}^n$, then one can give X the topology induced from the Euclidean topology of $\mathbb{C}^n$ which we call the *Euclidean topology* on X; then X becomes a Hausdorff topological space and we let $\Phi(\mathbb{C}^n, X) = X$, with the Euclidean topology. For every morphism of complex affine algebraic varieties $F : X \to X'$ we let $\Phi(F) = F$ (which is continuous for the Euclidean topologies).

EXERCISE 40.35. Check that if $F : X \to X'$ is a morphism of complex affine algebraic varieties then F is continuous for the Euclidean topologies.

EXAMPLE 40.36. Consider the following functor that plays a key role in an area of Mathematics called *algebraic topology*. The functor is

$$\Phi : \{\text{pointed topological spaces}\} \to \{\text{groups}\}$$

and is defined as follows. For (X, x) a pointed topological space we let

$$\Phi(X, x) = (\pi_1(X, x), \star,', e),$$

where the latter is the following group (called the *fundamental group* of (X, x)). To define the set $\pi_1(X, x)$ we first define the set $\Pi(X, x)$ of all continuous maps $\gamma : [0, 1] \to X$ such that $\gamma(0) = \gamma(1) = x$; the elements γ are called *loops*. Next one defines a relation $\sim$ on $\Pi(X, x)$ called *homotopy*: two loops $\gamma_0, \gamma_1 : [0, 1] \to X$ are called *homotopic* (and write $\gamma_0 \sim \gamma_1$) if there exists a continuous map $F : [0, 1] \times [0, 1] \to X$ such that $F(t, 0) = \gamma_0(t)$, $F(t, 1) = \gamma_1(t)$, $F(0, s) = x$, $F(1, s) = x$, for all $t, s \in [0,1]$. One proves that $\sim$ is an equivalence relation on $\Pi(X,x)$ and one defines the set $\pi_1(X,x)$ as the set of equivalence classes:

$$\pi_1(X,x) = \Pi(X,x)/\sim .$$

The class of a loop γ is denoted by $[\gamma] \in \pi_1(X,x)$. On the other hand there is a natural "composition map"

$$\Pi(X,x) \times \Pi(X,x) \to \Pi(X,x), \quad (\gamma_1, \gamma_2) \mapsto \gamma_1 \star \gamma_2,$$

defined by $(\gamma_1 \star \gamma_2)(t) = \gamma_1(2t)$ for $0 \leq t \leq 1/2$ and $(\gamma_1 \star \gamma_2)(t) = \gamma_2(2t-1)$ for $1/2 \leq t \leq 1$. (Note that $\gamma_1 \star (\gamma_2 \star \gamma_3) \neq (\gamma_1 \star \gamma_2) \star \gamma_3$ in general.) However one can prove that

$$\gamma_1 \star (\gamma_2 \star \gamma_3) \sim (\gamma_1 \star \gamma_2) \star \gamma_3. \tag{40.1}$$

Define a binary operation on $\pi_1(X,x)$ by

$$[\gamma_1] \star [\gamma_2] = [\gamma_1 \star \gamma_2].$$

This makes $\pi_1(X,x)$ a group with identity $e = [\gamma_x]$ where $\gamma_x(t) = x$ for all t; associativity follows. For every continuous map $F : X \to X'$ we let $\Phi(F) = F_*$ where, for $\gamma \in \Pi(X,x)$, we let $F_*([\gamma]) = [F \circ \gamma]$.

EXERCISE 40.37. Prove that $\sim$ is an equivalence relation on $\Pi(X,x)$. Show that the operation $\star$ is well defined on $\pi_1(X,x)$ and gives a group structure on $\pi_1(X,x)$. Check that the data above define a functor.

One can ask if there is a way to summarize the main objectives of modern Mathematics. We would like this summary to transcend the particularities of the various fields in which the questions are being raised. The language of categories seems to be well adapted for this purpose, as we shall see in the examples below.

EXAMPLE 40.38. (Equations and solutions) Let $\mathcal{C}$ be a category as above. Let us define an equation to be a morphism $b \in Hom(y,z)$ and let $a \in Hom(x,z)$. Let us define the set of solutions in a of the equation b as the set

$$Sol(a,b) = \{c \in Hom(x,y) \mid a = b \star c\}.$$

A large part of Mathematics is devoted to "finding the set of solutions of given equations" in the sense above. Algebraic equations and differential equations can be put, for instance, into this setting.

In order to put algebraic equations into the framework above let us consider a simple situation in which $\mathcal{C}$ is the dual of the category of commutative unital rings. Let $f \in A = \mathbb{Z}[x_1, ..., x_n]$ be a polynomial in variables $x_1, ..., x_n$ with coefficients in $\mathbb{Z}$ and define an equivalence relation $\equiv_f$ on A by declaring that $u \equiv_f v$ for $u, v \in A$ if and only if there exists $w \in A$ such that $u - v = fw$. Write $A/(f)$ for the set $A/\equiv_f$ of equivalence classes. Then $A/(f)$ becomes a commutative unital ring with operations induced by the operations on A. Now let $b \in Hom(A/(f), \mathbb{Z})$ be the morphism corresponding to the natural ring homomorphism $\mathbb{Z} \to A/(f)$ and let $a \in Hom(K, \mathbb{Z})$ be the morphism corresponding to the natural homomorphism $\mathbb{Z} \to K$ where K is a field. Then we claim that there is a natural bijection

$$\psi : Sol(a,b) \to \{(c_1, ..., c_n) \in K^n \mid \; f(c_1, ..., c_n) = 0\}$$

given as follows: for every solution $c \in Hom(K, A/(f))$ corresponding to a ring homomorphism $c : A/(f) \to K$ we can attach the tuple $\psi(c) = (c([x_1]), ..., c([x_n]))$ where $[x_i] \in A/(f)$ is the equivalence class of x_i. This shows that the concept of solution in the category $\mathcal{C}$ corresponds to the usual concept of solution of an algebraic equation.

In order to put differential equations into a categorical framework start with a differential equation of the form

$$\frac{d^r F}{dx^r} = Q(x, F(x), \frac{dF}{dx}(x), ..., \frac{d^{r-1}F}{dx^{r-1}}(x)) \tag{40.2}$$

where $F \in C^\infty(\mathbb{R})$ and $Q \in C^\infty(\mathbb{R}^{r+1})$. Let now $\mathcal{C}$ be the category of ordinary differential equations. Consider the object $X = Z = (\mathbb{R}^1, D)$ with $DF = F'$ the usual derivative. Consider also the object $Y = (\mathbb{R}^{r+1}, V_Q)$ where for $f \in C^\infty(\mathbb{R}^{r+1})$, $f = f(x, y_0, ..., y_{r-1})$, we set

$$V_Q(f) = \frac{\partial f}{\partial x} + y_1 \frac{\partial f}{\partial y_0} + y_2 \frac{\partial f}{\partial y_1} + ... + y_{r-1} \frac{\partial f}{\partial y_{r-2}} + Q(x, y_0, ..., y_{r-1}) \frac{\partial f}{\partial y_{r-1}}. \tag{40.3}$$

We let $a : X \to X = Z$ be the identity and $b : Y \to X$ be defined by the first projection $\mathbb{R}^{r+1} \to \mathbb{R}$, $(x, y_0, ..., y_{r-1}) \mapsto x$. Then there is a natural bijection

$$\Psi : Sol(a, b) \to \{F \in C^\infty(\mathbb{R}) \mid \ F \text{ is a solution to 40.2}\}$$

given as follows. For every solution $c \in Sol(a, b)$ given by a map $c : \mathbb{R} \to \mathbb{R}^{r+1}$, $c(x) = (x, c_0(x), c_1(x), ..., c_{r-1}(x))$, we let $\Psi(c) = F$ with $F(x) = c_0(x)$. This shows that the concept of solution in the category $\mathcal{C}$ corresponds to the usual concept of solution of a differential equation.

Exercise 40.39. In the notation above define the operations on $A/(f)$ and check that $A/(f)$ is a ring. Prove that ψ is well defined and a bijection.

Exercise 40.40. In the notation above prove that Ψ is well defined and a bijection.

Example 40.41. (Symmetries) Let $\mathcal{C}$ be a category and $x \in Ob(\mathcal{C})$. We denote by $Aut(x)$ the set of isomorphisms in $Hom(x, x)$. Then $(Aut(x), \star, (\ )^{-1}, 1_x)$ is a group; it is referred to as the automorphism group of x and is viewed as the "group of symmetries of x." Many problems of modern Mathematics boil down to computing this group.

Here is an example. Let $\mathcal{C}$ be the category of commutative unital rings. Let $f \in \mathbb{Q}[x]$ be a polynomial in one variable and let $Z = \{\alpha_1, ..., \alpha_n\}$ be the set of all roots of f in $\mathbb{C}$. Let

$$K_f = \{P(\alpha_1, ..., \alpha_n) \mid \ P \in \mathbb{Q}[x_1, ..., x_n]\} \in Ob(\mathcal{C}).$$

(This is actually a field, called the *splitting field* of f.) Then $Aut(K_f)$ is the group of all ring isomorphisms $g : K_f \to K_f$; this group is called the *Galois group* of f over $\mathbb{Q}$ and is sometimes denoted by G_f; it plays a central role in number theory. For $g \in G_f$ we have that $g(Z) = Z$ so g induces a permutation σ_g of $\{1, ..., n\}$ such that $g(\alpha_i) = \alpha_{\sigma_g(i)}$ for all i. We get an injective homomorphism $G_f \to S_n$, $g \mapsto \sigma_g$.

Here is another example. Let $\mathcal{C}$ be the category of ordinary differential equations. For the object $(\mathbb{R}^{r+1}, V_Q)$ with V_Q as in 40.3, the group $Aut(\mathbb{R}^{r+1}, V_Q)$ is viewed as the group of symmetries of the Equation 40.2. For $V = 0$ the group $Aut(\mathbb{R}^n, V)$ is called the *diffeomorphism group* of $\mathbb{R}^n$ and is usually denoted by *Diff*$(\mathbb{R}^n)$.

Here is another example. Let $\mathcal{C}$ be the category of all vector spaces over a field K. Then there is a natural isomorphism $Aut(K^n) \to GL_n(K)$.

Exercise 40.42. Look at other examples of categories and analyze the *Aut* groups of their objects.

Example 40.43. (Classification problem and invariants) Let $\mathcal{C}$ be a category. Then there is an equivalence relation $\simeq$ on $Ob(\mathcal{C})$ defined as follows: for $x, y \in Ob(\mathcal{C})$ we have $x \simeq y$ if and only if there exists an isomorphism $x \to y$. One can consider the set of equivalence classes

$$Ob(\mathcal{C})/\simeq .$$

"Describing" this set is referred to as the classification problem for the objects of category $\mathcal{C}$ and many important problems in modern Mathematics boil down to the classification problem for an appropriate category.

For some categories this is trivial; for instance if $\mathcal{C}$ is the full subcategory of the category of vector spaces over a field K consisting of the finite dimensional vector spaces then there is a natural bijection

$$Ob(\mathcal{C})/\simeq \ \to \ \mathbb{N} \cup \{0\}, \ [V] \mapsto \dim\ V.$$

However for other categories $\mathcal{C}$ such as the category of topological spaces or the category of complex affine algebraic varieties no description is available for $\mathcal{C}/\simeq$; partial results (e.g., results for full subcategories of these categories or variants of these categories) are known and some are very deep.

Another way to formulate (or weaken) the problem is via systems of invariants for objects of a category $\mathcal{C}$. If S is a set a system of invariants for $\mathcal{C}$ is a map

$$I : Ob(\mathcal{C}) \to S$$

such that for every $x, y \in Ob(\mathcal{C})$ with $x \simeq y$ we have $I(x) = I(y)$. Any system of invariants defines a map

$$\overline{I} : Ob(\mathcal{C})/\simeq \ \to \ S, \ \overline{I}([x]) = I(x).$$

If the latter is an injection we say I is a complete system of invariants. So the classification problem is the same as the problem of (explicitly) finding a complete system of invariants I and finding "all possible invariants" (i.e., the image of I). For instance if $\mathcal{C}$ is the category of finite dimensional vector spaces over a field K then $I = \dim$ is a complete system of invariants and I is surjective. A weaker problem for a general category is to find an "interesting" (not necessarily complete) system of invariants.

Some important theorems in Mathematics claim the equality of a priori unrelated invariants $I' : Ob(\mathcal{C}) \to S$ and $I'' : Ob(\mathcal{C}) \to S$. As corollaries one sometimes obtains interesting equalities between (integer or real) numbers.

Finally note that functors can produce systems of invariants as follows. If $\Phi : \mathcal{C} \to \mathcal{C}'$ is a functor then we have a natural map

$$I : Ob(\mathcal{C})/\simeq \ \to \ Ob(\mathcal{C}')/\simeq' .$$

So any system of invariants for $\mathcal{C}'$ induces a system of invariants for $\mathcal{C}$. This is one of the main ideas of algebraic topology (respectively algebraic geometry) for which $\mathcal{C}$ is the category of topological spaces (respectively complex affine algebraic varieties or variants of it) and $\mathcal{C}'$ is the category of groups, rings, vector spaces, etc.

Remark 40.44. It is far from being the case that all main questions of Mathematics have a structural that fits into the categorical framework explained above. What *is* the case, however, is that one expects that behind many of the important results of Mathematics there is a more general, structural "explanation" that *does* fit into the categorical viewpoint; looking for such "explanations" is a modern trend in Mathematics called *categorification*.

CHAPTER 41

Models

We briefly indicate here how one can try to create a "mirror" of Logic within Mathematics. What results is a subject called *Mathematical Logic* (also referred to as *Formal Logic*). This mirroring process can be thought of as a "second formalization" (or a formalization of a formalization of natural language). This mirror of Logic inside Mathematics is, however, far from accurate. An important discrepancy between (pre-mathematical) Logic and Mathematical Logic is the following. On the one hand (pre-mathematical) Logic involves "time" in its conception: theories are "finite" at every moment in time and "evolve in time" by addition of new symbols and sentences, becoming new theories. On the other hand theories in the sense of of Mathematical Logic can be viewed as attempting to "mirror" objects that are "eternal" and "infinite" (rather than "temporal" and "finite"). So if Mathematical Logic can be viewed as mirroring something then whatever it is that it mirrors is a realist entity quite different from (pre-mathematical) Logic. In the nominalist view this mirroring process is therefore highly problematic because it contains implicit traces of realism.

The main concept in our discussion of Mathematical Logic here will be that of *model*; cf. the definitions below.

All definitions and theorems below are in Set Theory T_{set} so we place ourselves under ZFC (no need here to add to ZFC the axiom of universes.) Recall that $\mathbb{N}$ and also its elements are sets, i.e. constants in L_{set}, that were defined in the chapter on the integers.

Definition 41.1. For $i \in \mathbb{N}$ define sets

$$c, f_i, r_i, x_i^*, \wedge^*, \vee^*, \neg^*, \rightarrow^*, \leftrightarrow^*, \forall^*, \exists^*, =^*, (^*,)^*, T, V, W$$

as follows:

$c = 0$
$f_i = (1, i)$,
$r_i = (2, i)$,
$x_i^* = (3, i)$,
$\wedge^* = 1$,
$\vee^* = 2$,
$\neg^* = 3$,
$\rightarrow^* = 4$,
$\leftrightarrow^* = 5$,
$\forall^* = 6$,
$\exists^* = 7$,
$=^* = 8$,
$(^* = 9$,

$)^* = 10,$
$C = 0,$
$T = \{c\} \cup \{f_1, f_2, f_3, ...\} \cup \{r_1, r_2, r_3, ...\},$
$V = \{x_1^*, x_2^*, x_3^*, ...\},$
$W = \{\vee^*, \wedge^*, \neg^*, \rightarrow^*, \leftrightarrow^*, \forall^*, \exists^*, =^*, (^*,)^*\}.$

V is called the set of *variables*; W is called the set of *logical symbols*. We sometimes write $x^*, y^*, z^*, ...$ instead of $x_1^*, x_2^*, x_3^*, ...$ By a T-partitioned set we mean in what follows a set S together with a map $S \to T$. We let $S_t \subset S$ the preimage of $t \in T$. Let S be a T-partitioned set; the elements of S_c are called *constant symbols*; the elements of S_{f_n} are called n-ary *function symbols*; the elements of S_{r_n} are called n-ary *relation symbols*. For any such S we consider the set

$$\Lambda_S = V \cup W \cup S$$

(referred to as the formal language attached to S).Then one considers the set of words $\Lambda_S^\star$ with letters in Λ_S. One defines (in an obvious way, imitating the metaaxioms in the Chapters on "pre-mathematical" Logic) what it means for an element $\Phi \in \Lambda_S^\star$ to be an S-formula or an S-formula without free variables (the latter are referred to as sentences). One denotes by $\Lambda_S^f \subset \Lambda_S^*$ the set of all S-formulas and by $\Lambda_S^s \subset \Lambda_S^f$ the set of all S-sentences.

REMARK 41.2. Note that symbols "$\wedge, \vee, ...$" are, of course, connectives in L_{set} while "$\wedge^*, \vee^*, ...$" are sets, i.e. constants in L_{set}. Also "$=$" is equality in L_{set} while "$=^*$" is a constant in L_{set}. Also "$(,)$" are separators in L_{set} while "$(^*,)^*$" are constants in L_{set}. Similarly "$x, y, z, ...$" are the variables in L_{set} while "$x^*, y^*, z^*, ...$" are constants in L_{set}.

EXAMPLE 41.3. Assume $\rho \in S_{r_3}$, $a \in S_c$, and $x^*, z^* \in V$. Define the word

$$\varphi = \forall^* x^* \exists^* z^* (^* \rho (^* x^*, z^*, a)^*)^* = (\forall^*, x^*, \exists^*, z^*, (^*, \rho, (^*, x^*, z^*, a,)^*,)^*).$$

Words are sets so φ is a set. Also φ is an S-formula i.e. $\varphi \in \Lambda_S^f$ is a theorem in T_{set}. For simplicity we write

$$\varphi = \{\forall x \exists z (\rho(x, z, a))\}^*.$$

Note that $\{, \}$, and * (the latter taken by itself) are not symbols in L_{set} but rather in metalanguage; so, whenever one encounters $\{...\}^*$ in a text in L_{set} one needs to replace $\{...\}^*$ by the corresponding word in L_{set}^*. Now the word $\{\exists a \forall z(\rho(x, z, a))\}^*$ is not an S-formula because constants cannot have quantifiers $\forall, \exists$ in front of them. Also the word $\{\forall x \exists z (\rho(x, a))\}^*$ is not an S-formula because ρ is "supposed to have 3 arguments." If $\Box \in S_{f_2}$ and a, x, z are as above then the word

$$\varphi = \{\forall x \exists z (\Box(z, a) = z)\}^*$$

is an S-formula.

REMARK 41.4. One can naturally define binary operations $\wedge^*$ and $\vee^*$ on Λ_S^s and a unary operation $\neg^*$ on Λ_S^s. Note that we have

$$\varphi \wedge^* (\psi \wedge^* \eta) \neq (\varphi \wedge^* \psi) \wedge^* \eta,$$

so Λ_S^s is not a Boolean algebra with respect to these operations. For simplicity, and if no confusion arises, we write $\wedge, \vee, \neg$ in place of $\wedge^*, \vee^*, \neg^*$.

Let M be a set. We let $S_c(M) = M$. For $n \in \mathbb{N}$ we set $S_{r_n}(M) = \mathcal{P}(M^n)$, the set of n-ary relations on M and $S_{f_n}(M) \subset \mathcal{P}(M^{n+1})$ the set of maps $M^n \to M$. We consider the T-partitioned set $S(M)$, union of the above. Then we can consider the formal language $\Lambda_{S(M)}$. An assignment in M is a map

$$\mu : V \to M.$$

For every set M and assignment μ one can prove (by recursion) that there exists a map $v_{M,\mu} : \Lambda^f_{S(M)} \to \{0,1\}$ which is a homomorphism with respect to $\vee, \wedge, \neg$, is compatible (in an obvious sense) with $\forall, \exists$, and satisfies obvious conditions with respect to relational and functional symbols, and also with μ. If φ has no free variables we write $v_M(\varphi)$ in place of $v_{M,\mu}(\varphi)$. Here are two examples of the "obvious conditions" referred to above. On example is the condition

$$\forall x \forall y(((x \in M) \wedge (y \in \mathcal{P}(M)) \to Q)$$

where

$$Q = \text{“}((v_M(y(^*x)^*) = 1)) \leftrightarrow (x \in y)\text{”}$$

Another example of such a condition is

$$\forall y(y \in \mathcal{P}(M) \to P)$$

where

$$P := \text{“}((v_{M,\mu}(\forall^* z^* (^* y(^* x^*, z^*)^*)^*) = 1) \leftrightarrow (\forall z((z \in M) \to ((\mu(x), z) \in y))))\text{”}.$$

Here

$$y(^*x)^* = (y, (^*, x,)^*),$$
$$\forall^* z^* (^* y(^* x^*, z^*)^*)^* = (\forall^*, z^*, (^*, y, (^*, x^*, z^*,)^*,)^*),$$

and the definition of Λ^f_S is such that the following is a theorem:

$$\forall x \forall y((x \in M) \wedge (y \in \mathcal{P}(M)) \to (y(^*x)^* \in \Lambda^f_S))$$

(similarly for the second condition). In general the conjunction of all such "obvious conditions" can be expressed in the language of Set Theory using recursion (this was done by Tarski); we will not go here into this.

Next we discuss "semantics" of formal languages. The word "semantics" here is taken in a metaphorical sense, justified by the fact that translations are involved.

Definition 41.5. By a *translation* of Λ_S into $\Lambda_{S'}$ we understand a map

$$\mathfrak{m} : S \to S'$$

which is compatible with the partitions (in the sense that $\mathfrak{m}(S_t) \subset S'_t$ for all $t \in T$). For every $\varphi \in \Lambda^f_S$ one can form, in an obvious way, a formula $\mathfrak{m}(\varphi) \in \Lambda^f_{S'}$ obtained from φ by replacing the constants and relational and functional symbols by their images under $\mathfrak{m}$, respectively. So we get a map

$$\mathfrak{m} : \Lambda^f_S \to \Lambda^f_{S'}$$

which is a homomorphism with respect to $\vee, \wedge, \neg$, compatible with $\forall, \exists$.

An S-structure (or simply a structure if S is understood) is a pair $\mathcal{M} = (M, \mathfrak{m})$ where M is a set and $\mathfrak{m}$ is a translation

$$\mathfrak{m} : S \to S(M).$$

So we get a map

$$\mathfrak{m} : \Lambda^f_S \to \Lambda^f_{S(M)},$$

which is a homomorphism with respect to $\vee, \wedge, \neg$, compatible with $\forall, \exists$. Fix an assignment μ and set $v_{M,\mu} = v_M$. We have a natural map

$$v_M : \Lambda^f_{S(M)} \to \{0, 1\}$$

which is again a homomorphism compatible with $\forall, \exists$. So we may consider the composition

$$v_{\mathcal{M}} = v_M \circ \mathfrak{m} : \Lambda^f_S \to \{0, 1\}.$$

We say that a sentence $\varphi \in \Lambda^s_S$ is satisfied in the structure $\mathcal{M}$ if $v_{\mathcal{M}}(\varphi) = 1$. This concept is independent of μ. This is a variant of Tarski's *semantic definition of truth*: one can define in Set Theory the predicate *is true in* $\mathcal{M}$ by the definition:

$$\forall x((x \text{ is true in } \mathcal{M}) \leftrightarrow ((x \in \Lambda^s_S) \wedge (v_{\mathcal{M}}(x) = 1))).$$

(We will continue to NOT use the word *true* in what follows, though.) If $v_{\mathcal{M}}(\varphi) = 1$ we also say that $\mathcal{M}$ is a *model* of φ and we write $\mathcal{M} \models \varphi$. (So "is a model" and "$\models$" are predicates that are being added to L_{set}.) We say a set $\Phi \subset \Lambda^s_S$ of sentences is satisfied in a structure (write $\mathcal{M} \models \Phi$) if all the formulas in Φ are satisfied in this structure. We say that a sentence $\varphi \in \Lambda^s_S$ is a semantic formal consequence of a set of sentences $\Phi \subset \Lambda^s_S$ (and we write $\Phi \models \varphi$) if φ is satisfied in any structure in which Φ is satisfied. Here the word semantic is used because we are using translations; and the word formal is being used because we are in Mathematical Logic rather than in pre-mathematical Logic. A sentence $\varphi \in \Lambda^s_S$ is valid (or a formal tautology) if it is satisfied in any structure, i.e., if $\emptyset \models \varphi$. We say a sentence φ is satisfiable if there is a structure in which it is satisfied. Two sentences φ and ψ are semantically formally equivalent if each of them is a semantic formal consequence of the other, i.e $\varphi \models \psi$ and $\psi \models \varphi$; write $\varphi \approx \psi$. Note that the quotient $\Lambda^s_S / \approx$ is a Boolean algebra in a natural way. Moreover each structure $\mathcal{M}$ defines a homomorphism $v_{\mathcal{M}} : \Lambda^s_S / \approx \to \{0, 1\}$ of Boolean algebras.

Example 41.6. Algebraic structures can be viewed as models. Here is an example. Let

$$S = \{\star, \iota, e\}$$

with e a constant symbol, $\star$ a binary function symbol, and ι a unary function symbol. Let Φ_{gr} be the set of S-formulas $\Phi_{gr} = \{\varphi_1, \varphi_2, \varphi_3\}$ where

$$\begin{aligned}
\varphi_1 &= \forall x \forall y \forall z (x \star (y \star z) = (x \star y) \star z) \\
\varphi_2 &= \forall x (x \star e = e \star x = x) \\
\varphi_3 &= \forall x (x \star \iota(x) = \iota(x) \star x = x).
\end{aligned}$$

Then a group is simply a model of Φ_{gr} above. We also say that Φ_{gr} is a set of axioms for the formalized theory of groups.

More generally:

Definition 41.7. A *formalized (or formal) system of axioms* is a pair (S, Φ) where S is a T-partitioned set and Φ is a subset of Λ^s_S; then define

$$\Phi^{\models} = \{\varphi \in \Lambda^s_S \mid \Phi \models \varphi\};$$

$\Phi^{\models}$ is called the *formal theory* generated by the system of axioms (S, Φ).

REMARK 41.8. One thinks of $\Phi^{\models}$ as semantically defined because its definition involves models and hence translations. One may define a "syntactic" version of this set namely the set

$$\Phi^{\vdash} = \{\varphi \in \Lambda_S^s \mid \Phi \vdash \varphi\};$$

where $\vdash$ is the predicate added to L_{set} meaning "φ provable from Φ" in an obvious sense that imitates the definition of proof in pre-mathematical Logic. Gödel proved that $\Phi^{\vdash} = \Phi^{\models}$. He also proved his remarkable incompleteness theorems which we will not review here. From a nominalist viewpoint all these theorems, being theorems in Set Theory, have no reference and in particular "they say nothing **about** Mathematics itself."

Appendix: Philosophy

In order for the student to better understand the implicit ontological position taken in this course we will offer, in the first part of this Appendix, some remarks (footnotes) on a number of philosophical systems (or groups of such systems) with special emphasis on their logical structure. We will follow the historical order in which these systems were put forward by their authors and we will emphasize their polemical nature. Our presentation is necessarily a grossly oversimplified account; all that is intended is to touch upon *some* ideas that have influenced the views on our subject. In particular our account below should not be viewed as a "Philosophy 101"-type crash course: such a course would require a much more systematic review than ours and a careful study of the primary sources. In the second part of this Appendix we will provide a number of comments on specific logical and philosophical issues of special relevance to our course.

Vocabulary. Before we proceed it is useful to recall that philosophy considers questions about existence (ontology), knowledge (epistemology), and value (axiology). Axiology comprises aesthetics (which deals with "the Beautiful") and ethics (which deals with "the Good" and "justice".) We will mostly ignore, in what follows, axiology and concentrate on aspects of ontology and epistemology. Ontology distinguishes between particulars (concrete entities such as individual bodies, individual events, all of these "real" or "imaginary", etc.) and universals (abstract entities such as properties, kinds, etc.). One can distinguish two ontological positions on particulars: materialism (some particulars exist outside mind) and idealism (all particulars exist in mind only). Also, one can distinguish three ontological positions on universals: realism (all universals exist outside mind), conceptualism (all universals exist in mind only), and nominalism (universals do not exist in any form). As to the main epistemological positions they are: rationalism (knowledge is rooted in reason), empiricism (knowledge is rooted in sensations), and skepticism (knowledge is impossible). There are innumerable nuances blurring these rough divisions. We have deliberately ignored these subtleties here: our purpose in this paragraph was to brutally fix a vocabulary before a more careful revisiting of the matter is possible. So the statements above should be viewed as definitions based on undefined terms such as "existence, knowledge, value, body, event, property, kind, reason, sensation," undefined predicates such as "exists in mind, exists outside mind, is rooted in" etc. The clarification of the undefined terms is provided by (and sometimes varies with) the philosophical system in question. Also, in our discussion below, we shall be using the word "logic" liberally. So we will use the words "first order Logic" to mean "Logic as presented in this course" and we will use the words "first order theory" to mean "a theory in the sense of first order Logic."

Ways to use the remarks below. There are two ways in which our remarks below can be of interest to the reader.

1. One way offers a glimpse into ontologies and epistemologies other than the ones adopted in this course.

2. Another way suggests to see each of the ontological systems below as a possible template for a first order theory and each of the epistemological systems below as a template for a discourse in Metalanguage.

So the first way contrasts philosophical systems with first order theories while the second way suggests a reinterpretation of the ontological systems in terms of first order theories. According to the second way, the various words that these ontological systems operate with can be viewed as either constants or predicates or functions in a first order theory. (Examples of such words are: existence, attribute, substance, senses, experience, thinking, doubting, infinity, space, time, causality, necessity, freedom, quantity, quality, intuition, absolute, spirit, negation, universal law, human nature, alienation, justice, the good, beauty, etc.) Descriptions or even definitions for these words are given by the authors of these systems, basic assumptions are being postulated which can be viewed as axioms, and some deductions are carried out that can be viewed as "proofs of theorems." No classical philosophical text (with the possible exception of Spinoza's Ethics) attempts to follow this pattern rigorously. It was one of Leibniz's dreams, however, and one of Gödel's dreams as well, that philosophy can be made rigorous in this way. All of the above amounts to turning ontological systems into "mock metaphysics" expressed in first order (object) languages and theories. On the other hand, as already mentioned, one can turn the epistemologies of these philosophical systems into rules of Metalanguage; examples are the various classifications of sentences (judgements) into "analytic," "synthetic," "a priori," "a posteriori," (in Leibniz and Kant), the philosophies of logic in Hegel and Russell, the philosophy of language in Wittgenstein, the philosophy of science in Carnap, etc. Needless to say ontologies and epistemologies are not clearly separated in the history of philosophy so reinterpreting these as first order theories in object languages versus discourses in Metalanguage exhibits a certain arbitrariness. It is for this reason that we will not attempt, in what follows, to reformulate the ontologies of various philosophical systems as first order theories; and we will not attempt to reformulate the corresponding epistemologies as discourses in Metalanguage. However an attempt to do this in special cases would be an interesting exercise that we recommend to the reader.

Remarks on Aristotle. The syntax of first order Logic is close to (but does not coincide with) that presented in Aristotle's Categories. When analyzing a sentence like "Snow is white" first order Logic views "snow" as a "constant" and "is white" as a "predicate." Aristotle, however, refers to "snow" and "white" as both being "things," he calls "snow" a "subject," and he says that "white" is predicable of the subject "snow." (This corresponds to our saying that "white" is an attribute of "snow" or that "snow" has the attribute "white"; or saying that "snow" is an object, "white" is a concept, and the object "snow" falls under the concept "white.") More generally the above analysis applies to sentences of the form "A is B" (sentences in subject-predicate form with predicate unary, involving a copula). Aristotle also says that some things A are "present in a subject" B

by which he means what we would probably express as "A is a property of B." (He insists that "A is present in B" not as a part in the whole because A is assumed to be "incapable of existence apart from B"; this latter concept is not syntactic but ontological.) So "whiteness" is not predicable of any subject (one cannot say that "this thing is whiteness"). Also whiteness is present in snow. (This is to be contrasted with Plato's realist theory of ideas. Plato would say that "whiteness" is an idea, it exists apart from snow and indeed its kind of existence is more robust than that of snow. Similarly, the objects of Mathematics, as well as Beauty, Justice, etc. are ideas existing independently and more robustly than the things of the sensible world.) Aristotle defines "substance" as a thing which is not predicable of any subject and is not present in any subject; for instance Socrates is a substance but whiteness is not. Substances have no contraries or variation in degrees. Other categories apart from substances are discussed such as: quantity, quality, priority/simultaneity, movement, etc. He says that quantity has no contrary: "small" and "big" are not contrary to each other because if A is bigger then B and B is bigger than C then B is both big (because bigger then B) and small (because smaller than A) so B would have two contradictory properties; this contrasts with the concept of relation in modern Logic and can be viewed (as Russell does) as a confusion between "is big" (a unary predicate) and "is bigger than" (a binary predicate/relation). A discussion of quantification (as in "all A are B," "some A are B," "no A is B," etc) and syllogisms (as in "from A and B it follows that C") is made in Aristotle's Prior Analytics. A detailed analysis of proofs (axioms, definitions, and various types of proofs, including proofs by contradiction) and a discussion of knowledge in general (causes, truth) is made in Posterior Analytics.

From the viewpoint of first order Logic Aristotle's language L_{Arist} contains constants such as "Socrates," "white," "whiteness," etc. One has binary predicates "is predicable of" and "is present in." One defines the unary predicate "is a substance" by saying that "for all x, x is a substance if and only if for all y different from x we have that x is not predicable of y and x is not present in y." One can attempt to continue this type of analysis of Aristotle's system (and of other systems described below) by turning it into a "mock metaphysics" but we shall not pursue this any further.

Remarks on rationalism. This is mainly represented by Descartes, Spinoza, Leibniz, and can be viewed as the peak of the "classical metaphysics" originated by Aristotle. They based their ontologies on modified versions of Aristotle's notion of substance. Motivated by different types of arguments (and sensibilities) they postulated different theories of substance as follows.

Descartes postulated that there exist two and only two substances called matter (which possesses the attribute of extension) and mind (which possesses the attribute of thinking). The two substances must interact (we know this from our day to day life) but the problem of how that happens (the mind-body problem) was never solved by Descartes: how can a non material thing (mind) move a material thing (the body)? (The only interaction admitted by Descartes is mechanical interaction by immediate contact/collision/pressure; but non-material things cannot produce movement by collision/pressure.) At one point Descartes tried to identify the place of the interaction of mind and matter in the pineal gland. (Descartes' mechanistic theory of interaction was subsequently demolished by Newton's theory of forces acting at a distance; Newton's paradigm was viewed by his contemporaries, as well as by himself, as a reintroduction of the "occult sympathies" of medieval philosophy but the practical success of Newton's theory was so overwhelming that it led to the dethroning of Descartes' physics.) The existence of individual mind in Descartes is more certain than the existence of material bodies. However doubting plus the "cogito" can be used to "prove" that the individual mind exists; for a first order theory giving this see Example 7.2. There is a (somewhat problematic) deduction of the existence of God from the existence of individual mind (based on the postulated fact that the idea of perfection in mind must come from a being that has that perfection; without that being, it is maintained, the idea cannot occur because, it is implied, ideas are mirror images of beings). Finally from the existence of God the existence of the World is deduced.

Spinoza postulated that there exists only one substance (which possesses infinitely many attributes two of which are extension and thought; the rest of the attributes are not knowable by Man); he called this substance God and viewed objects and men as "wrinkles" in this substance (a view very much like that of contemporary theoretical physics). His Ethics is written in the form of Euclid's Elements: the claims are Propositions for which proofs are provided.

Leibniz postulated that there exist infinitely many substances called monads or souls or minds (which possess attributes such as perception and appetite but not extension). Hence monads have no parts; they can be visualized as (but not identified with) geometric points. Each monad A perceives (more or less clearly) the other monads (the image in Leibniz is that of rays coming from all directions and converging to one geometric point, A) but no two monads interact. The apparent interaction of various parts of reality is due to a "pre-established harmony" of monads (the metaphor is that of the clocks that show the same time without interacting). No two monads have the same qualities (Leibniz' principle of identity of indiscernibles). Some monads are more conscious than others. The souls of men are more conscious than the monads that compose the body of men. God is the most conscious monad. Space consists in the distribution of monads and does not exist independently of the monads; there is no empty space. (This is in deep contrast with Newton's physics where empty space exists independently of the matter in it and functions like God's "sensorium.") The world we live in is the best of all possible worlds (a view ridiculed by Voltaire in Candide, for instance). A world is "the best" if it allows the largest variety of things to co-exist: it has the largest amount of "compossibles." So "the best" is defined in logical terms. In particular one can justify by this the presence of evil (since evil creates the possibility of good). Everything happens due to a reason (the principle of "sufficient reason"). Infinite chains of reasons can only be apprehended by God; Man's inability to apprehend infinite chains of reasons creates the illusion of chance. Propositions (sentences) that are true can be divided into truths of reason (analytic) and truths of fact (synthetic). Analytic propositions are those which can be found to be true by virtue of analysis of the terms involved; an example is "All men are animals" which is true because because "man" is defined as an animal with a specific list of properties. Analytic propositions follow from Logic only, independently of "facts.". Synthetic propositions are those which are not analytic; an example is "There is a dog under my table"; the truth of this cannot be found by analyzing the words "dog," "table," etc., but rather one needs to analyze the "facts" as well. Synthetic propositions do not appear as necessary (they are contingent) to Man but they appear as necessary to God who can apprehend the infinite chains of reasons for them to be true. There should exist a symbolic "calculus" that could decide all philosophical questions; Leibniz' attempt at such a calculus can be viewed as a precursor of Mathematical Logic. It is no coincidence that Leibniz also invented the "calculus" of Mathematics (integration and derivation) as a general machinery for solving mathematical problems.

Remarks on empiricism. In opposition to rationalism which is implicitly based on the assumption that reason is the main reliable source of knowledge (Descartes) classical British empiricism (Locke, Hume, Berkeley) held that the main reliable source of knowledge are sensations (acquired by sight, hearing, smell, etc.)

Locke was the first to clearly distinguish between primary qualities of objects (shape, movement, etc.) and secondary qualities (color, sound, smell, etc.); primary qualities were postulated to belong to the objects whereas secondary qualities were postulated to "appear" to us without being entirely "attributable" to objects. (This distinction was later abolished by Kant who postulated that there are no primary qualities. A similar view was actually expressed by Leibniz in correspondence but he did not elaborate on this.)

Hume was the first to subject causality to a careful analysis. His radical idea was that causality (the claim that event A is the cause of event B) does not correspond to a "thing" that can be detected by our senses (that has a shape, sound, smell, etc.) Therefore one needs to view causality as a superfluous concept.

Berkeley pushed empiricism to its limit which is essentially an idealist skepticism. Since the only knowledge we have comes from senses the "outside" world has no "reality" when it is not perceived. The only way "reality" subsists is as an "idea in the mind of God."

Remarks on Kant. His system is a synthesis of rationalism and empiricism. Such a synthesis may be viewed as surprising because rationalism and empiricism seem, at first sight, incompatible; but in fact Kant's synthesis also implied the rejection of various elements of both these doctrines. In particular Kant rejected both the classical metaphysics from Aristotle to the rationalists (e.g., their reliance on the concept of substance) and the skepticism/idealism of "extreme" empiricism (found, for instance in Berkeley). Kant, who was originally a Leibnizian, says that his reading of Hume woke him up from his "dogmatic slumber." To explain Kant's synthesis of rationalism and empiricism one can proceed as follows. Before Kant (e.g., in Leibniz) the division between analytic and synthetic propositions was taken to coincide with the division between a priori and a posteriori propositions. A priori propositions are those that are independent of experience. A posteriori propositions are those that depend on experience. Also recall from Leibniz that analytic propositions are those which can be found to be true by virtue of analysis of the terms involved while synthetic propositions are those which are not analytic. So the view before Kant was that a proposition is analytic if and only if it is a priori. Kant calls propositions by the name of judgements. Kant's thesis is that, although all analytic judgements are a priori it is not the case that all a priori judgements are analytic: there exist also synthetic a priori judgements. In fact Kant maintained that all judgements of geometry (respectively arithmetic) are synthetic a priori because, according to him, their truth is independent of experience but also depends on an extra-logical intuition of space (respectively time) which is, we would say, "wired into our brains." (The intuition of time is encapsulated in the sequence $1, 2, 3, \ldots$ which "measures" the passing of time.) Similarly, according to Kant, the laws of physics (and, implicitly, causality) are synthetic a priori because they are not purely logical (so they are synthetic) but they are universal (admit no exception, so a priori). Kant's explanation (in his Critique of Pure Reason or his Prolegomena) as to how synthetic a priori judgements are possible (and in particular why geometry, arithmetic, and physics agree with experience) is as follows. There is a "noumenal" world (of things as they are in themselves) and a "phenomenal" world (of things as they appear to us, i.e., of which we have an experience). Experience is constructed from perceptions (which "correspond," in an unknowable way, to noumena which are themselves unknowable) with the help of two procedures: one regulated by intuition and the other regulated by the categories of understanding; both are "wired in us". The forms of intuition are space and time. Among categories is causality. So space, time, and causality are "wired in us" and constitute some of the very conditions for the existence of experience itself; as a result experience is built based on the principles of geometry, arithmetic, and physical law and hence must conform to these automatically. Note that intuition and categories only apply to phenomena (as conditions for the existence of phenomena) and never to noumena. In particular judgements made about the World as a whole (viewed as a noumenon) or about God (viewed as a noumenon) lead to contradictions (antinomies): e.g., the problem of the finiteness of the World in time or space is an "ill posed" problem because time and space only apply to phenomena. The World as a whole and God are, for Kant, what he calls ideas and can be viewed as "limits" that can be approached but never attained by reason. As to categories, they are "parallel" to the forms of judgement (hence to syntactical concepts). Causality corresponds to hypothetical judgements (if...then...). Other categories are: unity, plurality, reality, negation, possibility, existence, etc., corresponding to universal judgements (for all x...), particular judgements (involving constants), affirmative judgements, negative judgements (it is not the case that...), problematical judgements (it is possible that...), existential judgements (there exists x...), etc., respectively. The forms of judgement essentially go back to Aristotle; Kant's original thesis is that they "correspond" to categories, hence to conditions of possibility for experience.

A standard argument against Kant's view on geometry (which in his time meant Euclidean geometry) is based on the discovery, after Kant's time, of non-Euclidean geometries. The argument is that since there exist several incompatible geometries they cannot be all part of our intuition. However this argument does not seem to invalidate Kant. Indeed one can argue that our brains are wired consistently with the laws of Euclidean geometry while the non-Euclidean geometry of physics, although it provides an appropriate description of experience, apparently violates our intuition hence cannot be part of our intuition; we cannot "intuit" curved space except through embedding it into a higher dimensional Euclidean space. By the way one can note that the visual data obeys the laws of perspective (hence of projective geometry, and not of Euclidean geometry); the brain appears to reorder this projective geometric data according to the laws of Euclidean geometry. This seems to vindicate Kant.

Another argument against Kant's view on geometry is that there are two types of geometries: one type is mathematical (all these geometries, Euclidean or not, are analytical and a priori) and the other type is physical (which is synthetic and a posteriori). This view (to be found in Russell and Carnap) assumes space is in the noumena (it assumes the objective reality of space) so this argument against Kant has actually nothing to do with the existence of non-Euclidean geometries but rather with the relation between space and noumena. The distinction between mathematical and physical geometries was already made (implicitly) by Gauss and (explicitly) by Hilbert, Poincaré, and Einstein (but not necessarily as a rejection of Kant).

Yet another argument made against Kant is that his view on the synthetic a priori nature of Mathematics appears today to be wrong in that Mathematics has been reduced today to Logic (so it is totally analytic). This argument misses an important aspect of Mathematics. Indeed, although Mathematics, viewed as a first order theory, seems to be entirely analytic it seems that, at the same time, Mathematical Practice (the invention/discovery and application of Mathematics) has an essential synthetic aspect: Mathematical Practice requires the Kantian "intuitions" and this (rather than analyticity) is, arguably, the essence of Mathematics. This seems to vindicate Kant's view on Mathematics and (by similar arguments) his view on physics. But note that such a position needs an ontology that is more generous than that of Formalism.

Here is an argument against Kant's view on causality (cf., for instance, Russell). Perceptions, according to Kant, are assumed, the argument goes, to be caused by noumena. But causality is a category and categories only apply to phenomena, a contradiction. This contradiction, Russell says, seems to not have been grasped by Kant. Yet there seems to be an easy way out of this contradiction by assuming that there are two (entirely different) types of causality: one that is a category and one that is a relation between noumena and perceptions; the two types of causality could/should be called by two different names and the contradiction disappears.

A propos of noumena and phenomena, the problem of free will in Kant is resolved as follows. Man can be viewed from two different points of view, as noumenon and as phenomenon. If viewed as phenomenon Man is subject to the law of causation (like every phenomenon) so Man is not free; but when viewed as noumenon he is not subject to the law of causation (no noumenon is) so Man is free. This concept of freedom, as something residing in noumenon (rather than in mind, e.g., in reason), played an essential role in the romantic conception of Man in the 19th century (cf. Goethe, Schiller, etc.) On the other hand Kant's ethics (presented in his Critique of Practical Reason or his Metaphysics of Morals) is entirely "non-romantic": a good action is defined as one that can be willed to become a universal law. The prescription "Do the Good" with Good defined in logical terms as above is called by Kant the categorical imperative. This is to be contrasted with hypothetical imperatives which are of the form "If you want to go to Heaven then do the Good." So the Good consists in obeying a rational, self-imposed, disinterested, logically necessary law and has nothing to do with our freedom as noumena or with the conditional "If... then....".

Remarks on Hegel. Hegel's Logic is both a Logic and an Ontology. It is posited that only one thing exists and that thing is referred to as the Absolute. However the Absolute is not a substance like Spinoza's unique substance because the Absolute undergoes changes: its attributes change. This change can be said to introduce time in the picture and this is considered to be one of Hegel's important contributions to philosophy. The Absolute seems to possess a mind/soul/spirit which is referred to as Geist. (Hegel's philosophy is viewed as the the apex of German Idealism originating in Kant; however Kant rejected the charge of "idealism" made by some of his contemporaries.) Geist thinks and evolves (towards self-awareness and self-knowledge) through thinking. It thinks through sentences. No sentence is entirely true or false because it can only reflect a part of the Absolute. The movement of the sentences follows the following pattern (Hegel's dialectic). To every sentence which can be called "thesis" one can oppose another sentence called an "antithesis" (which is a form of partial negation of the thesis); they are both, in different ways, "provisionally true." Then the "thesis" and the "antithesis" are "reconciled" into a third sentence called "synthesis." The "synthesis" becomes a new thesis for which a new antithesis is found after which a new synthesis is found. (One can see the contrast with classical Logic where a sentence and its negation cannot be simultaneously "affirmed." For Hegel there is no proof of an "eternal" sentence. There is only a quest, achieved by successive negations and syntheses, for more and more embracing judgements. Such a Logic must be contrasted with that employed by Mathematics viewed as a first order theory, which aims for "eternal" sentences. However Hegel's Logic may apply to the Mathematical Practice (including the history/evolution of mathematical ideas.) The beginning of Logic starts with the thesis affirming Being. The antithesis affirms Nothingness. And the synthesis affirms Becoming. In this way concepts come into existence and acquire a more and more complex content. It is then claimed (cf. Hegel's Philosophy of History) that the history of mankind runs a course that is parallel to that of Geist. Geist's direction is towards self-consciousness. The history's direction is towards "freedom." In the early empires only one was free (the emperor); in classical Greece some were free (the non-slaves); in modern societies (constitutional monarchies of the 19th century, especially in Prussia) all men are free. However by "freedom" Hegel understands something close to Kant's ethic concept of Good rather than Kant's concept of noumenal freedom: as Russell puts it Hegel's freedom is the freedom to obey the State. In fairness to Hegel, the State is viewed by him as the realization of the rational Good. Unlike Kant's philosophy which was greatly influenced by the successes of mathematical physics of the 18th century and influenced in its turn the scientific thinking of the 19th and 20th centuries, Hegel's Logic seems to be disconnected from, and of no relevance to, sciences and Mathematics viewed as first order theories. Hegel's Logic arguably applies, however, to the history of Mathematics and sciences and more generally to history as a whole. Hegel's dialectic, interpreted as an identity game that is being played between contraries does play a role in first order Logic but only at the stage where one sets up the definitions and axioms of a theory (cf. the discussion in Wang 1996); once axioms and definitions are fixed the resulting theories are viewed as "frozen for eternity" and immune to dialectic.

It is largely as a reaction to Hegel that Marx in mid 19th century developed dialectical materialism. Roughly speaking, he considered a variant of Hegel's system in which he replaced Hegel's Geist (an idealist concept) by Praxis (a materialist concept) while maintaining Hegel's dialectic method. Like Hegel, Marx applied his theory to history but not to science and Mathematics viewed as first order theories. History, according to Marx, is governed by a universal law: the direction of history is towards the freedom of Man from alienation; alienation, originally a Hegelian term, is used here in the sense of alienation from one own's nature. The nature of Man involves self-realization; this view is, in some sense, influenced by Romanticism. However the universal law referred to above is inspired by the science of the time, in particular by Darwin's theories. As noted by Russell, among others, Marx's materialist ontology is not a simple repetition of the naive Greek materialism (Democritus, for instance). In Marx the "really existent" is a relation between Nature and Man.

Remarks on Nietzsche. His ontology is extravagant: it proposes, among other things, the thesis of the Eternal Return (everything that happens will happen again infinitely many times, a thesis that he tried to justify "scientifically" by invoking an argument in the style of Poincaré's Recurrence Theorem) and the thesis of the Superman (an exalted/messianic vision of the future of Man as a free, audacious, and creative agent: the Man, or dominant Man, of the future is either a warrior aristocrat, or a genius artist, or a combination of these). His most important work is in axiology or rather in the negation of all classical axiology: cf. his work on the "revaluation of all values," an effort, in fact, to demolish (without any real effort to replace) all the values (the "idols," as he calls them) of our civilization including science, religion, and morality. A superb stylist and an unapologetic iconoclast, profoundly influential in the 20th century culture and sometimes grossly misinterpreted (his legacy hijacked by Nazi Germany). Uninterested in epistemology, and inconsequential in Logic or Mathematics.

Remarks on phenomenology. This includes work by Husserl, Heidegger, Sartre and others in the 20th century. With Heidegger and Sartre this led to their brands of existentialism. The starting idea is that since (accepting Kant) the noumenon is not accessible one should "bracket" it (Husserl's terminology) and concentrate on the phenomenon. So the main task of philosophy is to have consciousness analyze itself. A mathematical-like analysis of consciousness was undertaken by Husserl. An analysis of the various possible types of existence (with the implicit realization that not everything exists in the same way) was undertaken by Heidegger (his concepts of Being and Dasein) and Sartre (being in itself and being for itself). For Sartre "being in itself" is, for instance, the being of a rock: in such a case, he says, essence precedes existence in the sense that what the rock "is" (its essence) dictates its fixed being in the world (its existence). On the other hand "being for itself" is the being of Man: in such a case, he says, existence precedes essence in the sense that what each man "is" (its essence) is dictated by man's free choices (which "are" his existence). Implicit in this doctrine is the absolute freedom of Man (within a situation, another basic concept in Sartre). Man can pretend to be non-free (i.e., entirely determined by situation): what results is "bad faith," a form of self-deception which, according to Sartre, is a betrayal of Man's true nature. All this discussion makes ontology have direct axiological implications. Note that one of Sartre's basic epistemological assumptions is that Man's thoughts are all available to consciousness (and this is what makes Man responsible for choosing). This is to be contrasted with the psychoanalytic view (Freud) that most psychological events are not available to consciousness and therefore most Man's actions are not free. It is important to note that the concepts of freedom appearing in Kant, Hegel, Marx, Freud, Heidegger, and Sartre are all different. Sartre, in his later work, attempted a synthesis between Marxism and his own earlier existentialism.

Phenomenology and its corollary, existentialism, were the main sources of what today is called the "continental philosophy" (whose main representatives are French and German). A rather different kind of philosophy, referred to as "analytic philosophy," has its origins in logical positivism which we discuss next and whose main representatives are British and American.

Remarks on logical positivism. Another reaction to Hegel (more generally to German philosophy from Kant to Hegel) is that of logical positivism (logical empiricism) which is a return to issues of relevance to sciences and Mathematics viewed as first order theories. This includes (to various degrees and not always in agreement) work by Mach, Wittgenstein, Carnap, Ayer, Quine, etc., in the 20th century. According to most logical positivists propositions that are neither analytic nor empirically verifiable are nonsensical. Further, one identifies empirically verifiable propositions with a posteriori propositions so one is led to the claim that all synthetic a priori propositions are nonsensical: this is a complete rejection of Kant and also of classical metaphysics (who had also been rejected by Kant but on different grounds). In particular Wittgenstein's Tractatus qualified most of classical philosophy as nonsensical. (Wittgenstein, however, thought that the most important questions faced by Man are neither analytic nor empirically verifiable but "mystical." He concluded in the Tractatus that these questions cannot be addressed by philosophy.)

As an illustration here are some comments on Carnap's logical definition of cause, explanation, prediction. Consider the following sentences

1. $\forall x(P(x) \to Q(x))$
2. $P(a)$
3. $Q(a)$.

Clearly 1 and 2 imply 3. His example:

1. "$\forall x \forall y$ (if x takes y from a place then y disappears from that place)."
2. "Jones took the watch from the table."
3. "The watch disappeared from the table."

His next example:

1. "$\forall x \forall y$(if x wants to borrow y then x takes y)."
2. "Jones wanted to borrow the watch."
3. "Jones took the watch (from the table)."

Carnap calls 2 and 3 facts and says 1 is an example of a universal law. (This is an empirical universal law; there are other universal laws called theoretical universal laws.) He says (by definition) that to find an explanation for fact 3 is to find a fact 2 and an universal law 1 that fit as above, i.e., $1 \wedge 2 \to 3$; one then calls 2 the cause of 3. He also says that 3 is a prediction from 2 based on 1. So the cause of "The watch disappeared from the table" is that "Jones took the watch from the table." And the cause of "Jones took the watch from the table" is that "Jones wanted to borrow the watch." The chain of causes may continue indefinitely provided universal laws are found. According to Carnap universal laws do not exist in nature, independently of our minds, e.g., expressing some natural necessity. Up to this point the position is the same as both Kant's and Hume's (who, on this point, agree). However, whereas Kant maintained that the universal laws are, as we would say today, "wired into our brains" and hence are immutable (so "universal") preconditions of our experience, Carnap sees these laws as simply sentences involving a universal quantifier which fit into the above scheme and are provisionally adopted based on repeated empirical testing (observation of regularities in nature). For Carnap causality is not necessity in nature since necessity is neither empirically observable nor analytic so it is not allowed in discourse; this was also Hume's argument and Carnap explicitly sides with Hume in this matter.

For an explanation to be a good explanation it needs to be the case that the fact 2 be empirically verifiable and that the universal law leads (starting from empirical verifiable facts 2) to as many empirically verifiable facts 3 as possible but not to their negation. So universal laws need to be few and simple: a proliferation of universal laws (or indefinitely complicated laws) is tantamount to no laws at all because every universal statement can be replaced by a "practically infinite" list of particular statements that could qualify as laws.

A variant of the verification principle referred to above was Popper's requirement that every universal law (or system of such laws, e.g., scientific theories) must be "falsifiable" in the sense that it should be possible to imagine an empirical test that could in principle violate the law (or the system of laws in question). (There were significant differences between the position of Popper and the positions of the logical positivists, by the way; also Ayer (Ayer 1952) did not see Popper's criterion as distinct from the verification criterion. But Popper did.) According to Popper if an instance is empirically found when a universal sentence $\forall x P(x)$ is false then the sentence needs to be rejected as a law. If the sentence is empirically verified every time this is being checked the sentence qualifies as a law and is provisionally accepted until new tests are made. But if there is no way to imagine an empirical test in which the sentence is false then the sentence does not qualify as a law. An example of bad explanation is:

1. "For all x if God wants x then x happens."
2. "God wants the planets to revolve around the Sun."
3. "The planets revolve around the Sun."

1 plus 2 form a bad explanation of 3 for any of the following reasons:

a) 2 is not empirically verifiable

b) the sentence 1 does not qualify as a law because it is not falsifyable: one cannot imagine an empirical test that could find something that God wants but God cannot make happen.

The physical (quantitative) laws are the best universal laws from the point of view of the above criterion. As example of such a law is Newton's universal law of gravitation that can be coupled with facts as follows:

1. "For every bodies $x, y, z, ...$ the trajectories of these bodies behave as if there was a force between each pair of them given by the formula F"
2. "The Sun, Earth and Moon are bodies."
3. "The trajectories of Sun, Earth and Moon are given by the functions S."

The formula F is the usual formula in Newton's theory:

$$\text{force} = \text{constant} \times \frac{\text{mass}_1 \times \text{mass}_2}{\text{distance}^2}$$

where "constant" is a universal constant (independent of the bodies involved). The law 1 above simply means that the trajectories satisfy a universal differential equation which is a consequence of the postulation of a hypothetical entity called force satisfying the formula F. "Force" is not an empirically verifiable/measurable entity; it is rather a concept used to deduce the universal differential equation. The trajectories are here assumed empirically measurable. The functions S are a specific solution of the 3 body problem. The remarkable fact about the law 1 is that it can be applied to any bodies whatsoever, e.g. to x the Earth and y an apple. The formula F above can be called a theoretical law as opposed to the law 1 which is an empirical law; what distinguishes them is that the law 1 only involves empirically observable/measurable entities whereas the "theoretical law" F involves the entity "force" which is not empirically observable/measurable. The distinction between empirical and theoretical laws is not clear cut, of course: for instance the empirical observation/measurement of trajectories may be rather indirect and depends itself on some background theory (about optical or radiation phenomena) with its own theoretical laws.

On a different note let us consider the following example analyzed by Carnap:

1. "For every iron rod x if x is heated then the Earth rotates."

This is not a universal law NOT because the heating of a rod seems to have nothing to do with the rotation of the Earth but because it is not falsifiable: the Earth will rotate no matter what. On the other hand consider the sentence:

1. "For every iron rod x and any body y on which x lies if x is heated then y rotates."

The latter qualifies as a law because it is in principle falsifiable by an experiment: for some heated x lying on some y, y might not rotate (as we, by the way, expect, although our expectation has nothing to do with falsifiability).

The view of logical positivism on natural law has some interesting consequences.

First note that logical positivism accepts free will. Indeed since the laws of nature do not imply necessity in nature (being simply parts of explanations of facts and/or principles allowing prediction) it follows that there is nothing objectively necessary about the behavior of Man. In particular Man has free will (where the free decisions are viewed as facts 3 explained/caused by psychological laws 1 and prior facts 2).

Secondly, it is interesting to mention here a physico-mathematical argument (found, for instance, in (Weyl 1963)) that seems to support the (somewhat puzzling) thesis that "physical laws must exist for tautological reasons." Following modern physics, Weyl argues, a physical law is a relation (not involving the space-times variables) between the values of some observables quantities $f_1, f_2, ..., f_n$ (that depend on space-time variables) and the values of their derivatives with respect to the space-time variables. Assume for simpicity that we only have one such quantity

$f_1 = f = f(x,y,z,t)$ with x,y,z space variables and t the time variable; think of f as a scalar field. Then we have 5 functions

$$f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial t}$$

of 4 variables so there must be one functional relation

$$F\left(f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial t}\right) = 0$$

between these 5 functions (which does not involve space-time variables explicitly). (Indeed, if the Hessian of f is non-singular then, at least locally, one can take $F(y_0, y_1, y_2, y_3, y_4) = y_0 - G(y_1, y_2, y_3, y_4)$ where G is the composition of f with the inverse of the map $(t,x,y,z) \mapsto (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial t}.)$ One can interpret this functional relation as a "natural law" governing our scalar field f. The conclusion is that *any* scalar field obeys some "natural law" tautologically! This demystifies, in some sense the concept of "natural law": there is no mystery in the fact that nature "obeys laws" for no matter how nature evolves in space-time there will be a "law" governing this evolution (which does not involve space-time variables so it is universal with respect to time and place). If there is a "miracle" it is not that there is a law F but that the law F is often found to be of a very simple nature. This simplicity, however, might be a consequence of the appropriate choice of the observable f. The latter argument was part of the early criticism against modern science: according to this criticism mathematical laws are found only because one isolates the phenomena to be studied from the whole of nature (in which case the function f will not reflect the complexity of nature). This kind of argument was invoked, for instance, by Goethe (in his theory of color) against Newton's theory of color.

⋆ ⋆ ⋆

We continue the discussion in our Appendix with some additional comments on the philosophy of Language, Logic, and Mathematics. We ignore, in what follows, historical development and concentrate on specific problems rather than on "systems."

Dogmatism. Our presentation of Logic and Mathematics is largely dogmatic in that it ignores alternative approaches and does not justify the (apparently) arbitrary (meta)definitions and (meta)axioms that are being introduced. Dogmatism definitely helped keeping our introduction minimal. But the question of a priori justifying our (meta)definitions and (meta)axioms remains. The quest for such a priori justifications is, in some sense, analogous to Hegel's requirement (Hegel 1975) that the "list of categories", for instance, be a priori deduced rather than "dogmatically given"; Hegel criticizes Kant for precisely such a dogmatic approach. In our case, for instance, one can ask: *what is the justification for having exactly* 5 *connectives* $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$ *and what is the justification for their particular "truth tables"*? One answer could be that our starting point were *natural* languages which we started *analyzing* before we *synthesized* our more *formal* languages. And, in analyzing natural languages, we just happened to *empirically find* 5 connectives that are being used in such a way that, if actual *truth* were an allowed predicate, then the "truth tables" would coincide with the corresponding actual truth tables. But this justification is not acceptable. First it is an a posteriori, rather than an a priori, justification. Secondly this justification is at odds with our insistence that the *truth* predicate be banned from our discourse; our "truth tables" are, rather, part of the inference machine. Here is, however, an a priori justification, in line with our ban on *truth*. There are 16 possible "truth tables" for a binary connective and 4 possible "truth tables" for unary connectives. So there are $16 + 4 = 20$ possible binary+unary connectives. So there are 15 "new" connectives in addition to the "old" 5 that we have been using. (We can call them $\triangle, \square, ...$, etc. and we can even translate these 15 connectives into English as, say, *gook*, *gonk*,... etc.) Now it turns out that *any of the* 15 *new connectives is expressible as a Boolean combination in terms of the old* 5 *connectives* (and indeed in terms of, say, $\neg$ and $\wedge$ alone); this is an easy exercise. So, after all, the old 5 connectives are enough to define *all* 20 *possible (unary+binary) connectives*. This somewhat dispels the mystery of the 5 connectives and their special "truth tables"; for they "generate" *all possible* connectives, i.e. all possible "truth tables". It would be interesting to find a similar a priori justification for the background axioms, the specific axioms of Set Theory (ZFC), and the various definitions within Mathematics. The ZFC axioms are an especially puzzling case: their shape is far from intuitive and hence far from easily justifiable a priori.

Plurality of Logics and Mathematics. One of the corollaries of our dogmatic approach is that only *one* pre-mathematical Logic was considered; and within it only *one* Mathematics (namely ZFC) was put forward; and within this particular Mathematics only *one* mathematical Logic was set up. But lacking an a priori justification for the choices made in developing the theory one should point out towards various alternatives. Indeed consideration of a pre-mathematical Logic involves setting up a set of syntactic rules; let $S_1, S_2, ...$ be the various possible sets of syntactic rules. Belonging to each set S_n of syntactic rules there are various languages $L_{n1}, L_{n2}, ...$ For each n one can select a language, say L_{n1}, that we want to use for Mathematics. In this language L_{n1} one can formulate various sets $A_{n11}, A_{n12}, ...$ of axioms (which are variants of "axioms for Mathematics"). For each n and m we can then give various metaaxioms $M_{n1m1}, M_{n1m2}, ...$ related to the concept of *theory*. All of these lead to corresponding theories $T_{n1m1}, T_{n1m2}, ...$ (which are variants of "Mathematics"). Within each T_{n1mk} one can set up various systems of definitions $D_{n1mk1}, D_{n1mk2}, ...$ each of which allows us to create a mirror of T_{n1mk} inside itself; for each D_{n1mkl} we get a different "Mathematical Logic." Then the syntactic rules we put forward in Part 1 of our course identify with one of our sets of syntactic rules, say with S_1. The language L_{set} introduced in Part 2 of our course identifies then with L_{11}. The ZFC axioms identify with one of the possible sets of axioms, say with A_{111}. Our specific metadefinition of the concept of *theory* corresponds to some M_{1111}, say. Set theory T_{set} identifies then with T_{1111}. And the definitions used to set up Mathematical Logic identify, say, with D_{11111}. So one can see that, in setting up pre-mathematical Logic, Mathematics, and Mathematical Logic, one actually has a whole spectrum of choices at each point. This perspective was fully understood and thoroughly explored by the classics of contemporary Logic, in particular by Gödel who made sustained efforts to enrich ZFC (i.e. enrich A_{111}) in such a way that the continuum hypothesis CH can be "decided" within the enriched Set Theory. Gödel also looked carefully at how his incompleteness theorems change if one "weakens" L_{11} and A_{111}. He noted, cf. (Wang 1996), that his arguments for the incompleteness theorem of arithmetic are *not* finitistic, as Hilbert had required for "metamathematics". In fact, it was pointed out by Gödel that his arguments of the incompleteness of arithmetic require the full power of the language/axioms of Set Theory and not merely the language/axioms of the integers; in other words, for the proof of these theorems, one cannot substantially weaken L_{11} and A_{111}. This was viewed by Gödel as an argument in favor of the "metaphysical reality" of sets.

Metaphysics. Metaphysics (in its classical form from Aristotle to Medieval to Leibniz) is another name for ontology: it deals with problems such as "being qua being" (existence in itself), existence of the world, of substances, what exists, how many substances there are, "first causes" (theology, rationality of the world, "logos"), "universals" (abstract terms, the unchanging), etc. Kant claimed to have "killed" classical metaphysics by shifting attention from ontology to (what he called) the "transcendental" viewpoint which is that of analyzing the conditions that make experience and understanding possible. Logical positivism of 20th century claimed to have "killed" metaphysics once again by rejecting the traces of metaphysics still present in Kant (e.g., the "synthetic a priori," "the thing in itself," etc.). Our presentation of Logic, and in particular of Mathematics, is as free as possible from metaphysics. This is mostly done by ignoring the metaphysical problems that are inherent in Logic (as in the theory of meaning, reference, truth) and more generally in linguistics. These problems are real; but, as shown here, their solution does not seem to be required for a presentation of Mathematics. Remarkably, Gödel believed (along Leibniz's suggestions and contrary to Kant's critique) that traditional metaphysics can be made into an exact science; here is a quote from Gödel reproduced in (Wang 1996): "Philosophy as an exact theory should do for metaphysics as much as Newton did for physics." Again, according to Wang, Gödel believed that complicated abstract concepts (such as sets, which he viewed as concepts) can be perceived as clearly as simpler abstract concepts (such as syntactic combinatorics, which he also viewed as a system of concepts); he also seems to have confessed that he could not eventually find a satisfactory system of primitive notions for metaphysics.

Mock metaphysics. One can turn the tables and view metaphysics as part of Logic by introducing a language $L_{metaphys}$ whose constants are *existence, world, God, this or that object*, etc., and whose predicates are *exists, predicates, is one, is infinite*, etc. This is doable and can be said to lead to a "mock metaphysics" which is a "mirror" of metaphysics inside Logic. One can iterate this move and create a mirror of the "mock metaphysics" (a "mock mock metaphysics") inside Mathematics (i.e. inside Set Theory) in the same way in which Logic itself has a mirror inside Mathematics which is Mathematical Logic. Now recall Alain Badiou's maxim that the metaphysics of existence qua existence *is nothing but* Set Theory. (By *Set Theory* Badiou seems to understand Cantor Set Theory or any metaphysically realist version of Set Theory.) This could then be transposed inside Mathematics by saying instead that, "the mock mock metaphysics [that we discussed above] is a chapter of the Set Theory T_{set}" (where now *Set Theory* T_{set} is in the sense of our course, in particular T_{set} is just a text). The moral of the above discussion is that all these mock-ifications of metaphysics are not metaphysics itself and say nothing about metaphysics itself (in the same way in which the theorems of Mathematical Logic say nothing about Logic itself, simply because, being sentences in object language, they say nothing at all). The only way "mock mock metaphysics" can be relevant to metaphysics is to have its sentences translated into a natural language such as English. (In the same way Mathematical Logic can only be relevant to Mathematics after translation into English, say.) Such translations are suspect and easily lead to contradictions; this is why we made it a rule in our course to ignore translations from object languages. On the other hand if metaphysics is to exist at all it has to be a prerequisite of (or to incorporate) Logic. Our course is trying to avoid making metaphysics a prerequisite of Logic, though, and this seems to work if the project is modest enough. Introducing Mathematics seems to be one of these modest projects.

Theories of truth. There are three main ways to theorize about truth: the correspondence theory, the coherence theory, and the deflationary theory. Each of these theories comes in a number of flavors. Correspondence theory originates with Aristotle and defines truth as agreement between sentences in a language and "states of affairs in the real world." The main argument against this theory is that the agreement is said to be between two terms ("language" and "state of affairs") whereas, in fact, one of the terms (the "state of affairs") cannot be made intelligible without the other term ("language"). So there is only one self-standing term ("language") and the theory is circular. Coherence theory (one of whose proponents was Hegel) is that truth is agreement among a body of sentences (or among a body of ideas expressed by sentences). The main argument against this theory is that there exist bodies of sentences B_1 and B_2 with the property that no two sentences in B_1 contradict each other, no two sentences in B_2 contradict each other, but some (or all) sentences in B_1 contradict some (or all) sentences in B_2 (for instance B_2 may consist of the negations of all the sentences in B_1); so there is no criterion to tell which of the bodies is "true." So truth cannot be defined. Deflationary theory essentially affirms that truth is a redundant predicate: saying that a sentence is true is the same as affirming that sentence. (We adopted this viewpoint in this course but as a philosophical position this seems too minimalistic to be useful.) To these theories one can add Tarski's "semantic" theory of truth which is, in some sense, a "synthesis" of the above 3 theories. In Tarski's theory one has an object language L (usually a formal language) and one considers translations (or interpretations) of L into a form of realist (Cantorian) set theory. Such translations are called *models.* A sentence P in L is then said to be *true in a model* M (intuitively in a "possible world" M) if the sentence becomes "true" as a sentence "about the elements of M." This definition of truth in a realist model does not solve the problem of defining truth; it just defines truth in L in terms of truth about Cantor (or other realist) sets. But truth about Cantor sets is left undefined! One way around this dilemma (which we adopted in this course) is to ascend one step by reproducing the whole Tarski scheme inside Mathematical Logic, as opposed to pre-mathematical Logic. The result is a "formal" version of Tarski's scheme. But being formal, this version has no semantic content, hence no direct report with the non-formal version of Tarski's scheme. This, once again, leaves truth undefined.

Varieties of nominalism. For the nominalist all that exists in connection with an abstract entity such as set, or number, or force, or mass, or God, is the corresponding word (as a physical mark on paper). One can arguably further subdivide nominalism into two types which we could call T and NT; type T recognizes objective truth as a meaningful concept (mostly through a correspondence theory) whereas NT (which is a rather rare and radical variety) denies the meaningfulness of objective truth. The present course is written from the viewpoint of an NT type nominalism. The classical books of Quine (Quine 1980) and Putnam (Putnam 1971, Putnam 1981) reject nominalism based on their commitment to a concept of "objective truth." Quine says (Quine 1980, p. 121) that he is adopting a "liberal ontology" (admitting classes, hence sets) while Putnam says (Putnam 1971, p. 23) that "reference to 'classes' [...] is indispensable to the science of logic." In particular Putnam makes a strong case against what we referred to as type T nominalism. However he does not consider type NT nominalism as a possibility and his arguments against type T do not seem to apply to type NT. There is one caveat in the assertion above that our approach is from the viewpoint of an NT nominalism; namely, although sentences in a language can be viewed as concrete objects (signs on a piece of paper), the theories $T, T', \ldots$ considered in this course may be accused of being non-nominalist entities (because they are invoked but not explicitly given as a physical object). This accusation has some merit and the way this course implicitly responds to it is as follows. The more abstract our discussion about theories becomes (i.e. the more one starts using words such as *consistency, completeness*, etc.) the less is one allowed to metaprove metasentences involved in such discussion. In the limit, when discussion about theories becomes utterly abstract, the discussion transforms itself into "un-metaprovable babbling." This is the very reason why this course defers all matters related to consistency and completeness to Mathematical Logic and essentially bans them from pre-mathematical Logic. What saves the day here is the ban on truth predicates *true/false*. The "babbling" referred to above is neither true nor false.

Variation on an argument of Mach. Here is an argument against realism in Mathematics. Assume one adopts a correspondence theory of truth in Mathematics (based on a realist position) and one considers a sentence such as, "*For any positive integer n there is a prime bigger than n.*" In order to ascertain the truth of such a sentence one needs to "check it" against "reality". Assuming the "reality" of the integers the only way to check the above is to perform an experiment. Since we assume the actual infinity of the integers the only possible type of experiment is a "thought experiment": one fixes, mentally, a positive integer n and one puts forward a procedure that constructs a prime bigger than n. But an objection to this can be raised similar to Mach's objection to thought experiments in physics. Recall that Mach's objection (originally addressing Newton's "bucket experiment") is that it is meaningless to talk about the outcome of a thought experiment if the conditions for the practical implementation of that experiment are inherently "impossible" to achieve. One should not accept thought experiments, for instance, that allow the size of the experimenter's apparatus to be the size of the Earth itself because there is no way to know what would happen if this were the case. In the same way we do not know if the procedure to find a prime bigger than a given n would function as expected in case n is a number bigger, say, than the "number of atoms in the universe." It is conceivable that the "rules of Logic" do not apply as expected to such an n. Similarly a thought experiment that starts with an "impossible" premise such as, "*Assume the whole universe, except the Earth, were to disappear...,*" should not be accepted; and in a similar vein mathematical proofs by contradiction (that usually start with an "impossible" premise as well) should not be accepted from a realist standpoint.

Variation on a maxim of Badiou. In the preface to (Badiou 2009) Alain Badiou says that, "today, natural belief is condensed in a single statement: *There are only bodies and languages.*" He then proposes to replace the 2 by 3 by introducing a "universal exception" as follows: *There are only bodies and languages, except there are truths.* Philosophy, he says, does not create truths; it merely organizes truths (which arise from the "truth-procedures" of art, science, politics, and love) using ontology (which is identified by him with a philosophical reformulation of Set Theory). What our approach amounts to is to the replace the 2 by 1 by introducing another "universal exception" as follows: *There are only bodies and languages, except all bodies are made of language.* The latter point of view is sufficient for pure Mathematics but it is not appropriate, of course, for more ambitious philosophical endeavors. Indeed ontology is a matter of commitment and commitment is a pragmatic attitude. In this course we say that the only things we accept as existent are the signs on paper that can be assembled into languages (according to rules that are also expressed within yet another language). By this we simply mean that, when dealing with Mathematics, we *choose* to ignore everything else except language. This is a commitment based on a practical decision to develop a certain specific project. Such a position leaves room for more elaborate ontologies oriented towards more ambitious projects. One motivation for a more inclusive ontology would be that, with a minimal ontology that only recognizes the reality of language, there is no room for non-linguistic reference. Discussion of non-linguistic reference may become relevant, however, when we ask the following question: *What does Mathematical Logic say about Mathematics itself?*

Variation on the field metaphor of Quine. Here is an example of a more inclusive ontology that can be used to accommodate the interaction between Mathematics and natural sciences. It is the ontology implicit in this course and to present it we shall elaborate on a metaphor of Quine's (cf. Quine 2008, p. 43) using some Kantian jargon. One can image a sphere the things outside of which are called the *noumenal* world. The things inside the sphere are called *symbols.* Both the noumenal world and the symbolic world are viewed as empirical/physical: the symbols are for instance written or spoken; we just choose to view them as irreducible/simple entities rather than complexes of images, sounds, etc. Symbols are articulated into systems called languages. The articulation itself is encoded into other languages and hence into other systems of symbols. There are various dynamical reports between the various languages (translations, reference, etc.) which are encoded into yet another language, etc. The noumenal world acts like a *boundary condition* (in a sense similar to that encountered in the theory of partial differential equations). The symbolic world acts like a *field.* The field adjusts itself according to the boundary conditions; in particular Logic itself is part of the field and adjusts itself to the noumenal world. The law according to which this adjustment takes place is one that seeks equilibrium (i.e. simplicity/intelligibility) very much like Maupertuis' law of minimum action. But this law does not correspond to a "noumenal law"; it is rather a law obeyed by the "symbolic world". The surface of the sphere is the *phenomenal* world. The strings of symbols close to the surface are *concepts.* The strings of symbols closer to the center of the sphere are the *ideas.* There is no direct relation (in particular semantic relation) between specific things outside the sphere such as the "real" planets and specific things inside the sphere such as the word *force*; the way the outside influences the inside is purely global (as in the case of boundary conditions influencing a field). The above metaphor is still "nominalist" but its ontology is more generous than the one adopted in this book. An interesting question related to this metaphor is: what are the "field equations"? Can they be written in one of the languages inside the sphere? This corresponds, in the metaphor above, to one of the main projects of critical philosophy.

Similarity and reference. Similarity theory of reference holds that reference is based on the similarity between our thoughts and reality. It was thoroughly criticized by empiricists like Locke and, at a deeper level by Kant and all the resulting German idealist tradition, down to the internalist project of Putnam, say. For an example showing that similarity does not imply reference see (Putnam 1981); Putnam's example is that of an ant tracing a line in the sand that is a recognizable caricature of Winston Churchill: here we have similarity without reference. And of course reference does not imply similarity as one can see in the example of the "*three feet long table*" whose mental image is *not* three feet long and indeed has no length at all. Interestingly, the similarity theory of reference is upheld by Wittgenstein in his Tractatus. He says "*There must be something identical in a picture and what it depicts, to enable the one to be a picture of the other at all.*" (Wittgenstein, 2.161). Here reference is viewed as based on similarity of form between language and reality, an extreme version of the correspondence theory of truth. The first sentence of the Tractatus reads, "*The world is all that is the case.*" This seems to identify the world with what is being said about the world, an extreme form of correspondence theory. Wittgenstein abandoned his view on reference and correspondence in his later work.

Indeterminacy of meaning. In our presentation the meaning of a sentence is *defined* to be the totality of all available translations of the sentence. The catch word here is "available"; it implies that the translations are viewed as given (via a dictionary and the grammar of paraphrases). But of course, for the philosophy of language, the problem that needs to be answered is: if meaning is defined in terms of translations then how are translations (e.g. dictionaries, grammars) possible if meaning is not yet present? For an illuminating discussion of these issues see (Quine 1980, Chapter III) and (Quine 1964). One of the main ideas in loc.cit. is that meaning is underdetermined (i.e. there is no way to fix it through the concept of logical truth). Cf. especially Quine's careful discussion of radical translation (i.e. translation from/to an unknown language) in (Quine 1964), in particular his description of what would take to create a Jungle-to-English dictionary (where Jungle is the language of a remote tribe none of whose speakers speak English). The creation of such a dictionary would require a detailed interaction with the members of the tribe such as sequences of questions that would identify their words for *yes* and *no* as well as their words for particular objects such as rabbits. This dictionary work would actually not guarantee that the Jungle word that we translate as *rabbit* in English actually refers to *rabbit* and does not refer to *a part of the rabbit* or *an instance of rabbit-hood*, etc. Quine's discussion is mostly relevant to natural (as opposed to artificial/formal) languages. For the (artificial) language of Set Theory (i.e. of Mathematics) the issue of meaning can be, in principle, completely avoided. Nevertheless, the metalanguage needed to operate with the language of Set Theory is treated in our course as a natural language and its own meaning (that was not discussed in the book) requires further clarification.

Reference versus meaning. This issue goes back to Frege's breakthrough in the philosophy of language. Frege's famous example is: the "morning star" and the "evening star" were discovered at some point, to have the same reference (to refer to the same physical star). Now the sentence "The morning star is identical to the evening star" makes the claim that the two stars are one; and the meaning of this sentence is this claim. But this claim goes beyond the reference of its terms. So meaning and reference are not the same thing. Put differently, if meaning and reference were the same thing then the sentence above would be a tautology which it is not because it accounts for an empirical discovery. In our course references to things such as stars are ignored because we ignore non-linguistic reference. However linguistic reference is being considered in the form of metalanguage referring to language. Also meaning is being considered in the form of translations between object languages. The main difference between reference and meaning comes, in our course, from the fact that translations (in their simplest word for word form) attach constants to constants, predicates to predicates, etc., (hence they preserve logical categories) while linguistic reference attaches sentences, and even sequences of sentences, in object languages to constants in metalanguage, connectives and quantifiers in object language to functional symbols in metalanguage, etc. (hence it shifts logical categories). In the case of not-word-for-word translations logical categories are also shifted but only in the small and never in the large; for instance sentences in one language are always sent into sentences. Also we postulate that translations preserve reference; i.e. if L and L' have a reference and if P is a sentence in L whose translation in L' is P' then P and P' have the same reference.

About the phylogeny of reference and meaning: this is mostly an anthropological (rather than philosophical) problem. One modern answer (Cassirer 1951) involves the insights of the history of mythical thought, from its first stage of "momentary" deities (perceived as spontaneous apparitions accompanying concrete objects such as this or that tree) to more permanent deities (functioning as names for abstract concepts such as the general concept of storm, crop, etc.). Another modern answer (put forward by many including Bronowski, rejected by many, including Putnam) involves natural selection. The meaning of the concept of "cause," for instance, could have been fixed as follows. Say that hominid A throws a rock at hominid B and, as a result, hominid B dies. Hominid C witnesses the scene and his mental apparatus produces the following description of the events: "*The cause of death of hominid B was A's intension to kill B with a rock*". Let us also assume that another hominid, D, witnesses the same scene and *his* mental apparatus produces the following description of the events: "*The cause of death of hominid B was the property of B's head to attract rocks present in A's hand*". Now hominid C will survive by being careful to avoid hominids such as A; then C will have offspring who will inherit his particular mental wiring. Hominid D, on the other hand, will probably die soon at the hands of the likes of A; he will have no offspring and his type of mental wiring will not be inherited by anybody. After generations most hominids will have a mental wiring similar to C's; the category of "causality" with its more or less fixed meaning will have emerged in this way. This would explain not only the meaning and reference of the abstract term "causality" but also the origin of the a priori form of understanding (in Kant's sense) of "causality." Whether or not such an account has merit is beyond our scope here.

About the ontogeny of reference and meaning: the problem is how reference and meaning are fixed in a given individual. There are rival theories, for which we refer, for instance, to (Quine 1964) and (Chomsky 2006) respectively.

Grammatical analysis. Earlier we said that logical analysis of sentences in English is quite different from grammatical analysis. Let us take a quick look at the latter in the following simple example. Consider the following sentence in English:

"The father of Hamlet is a king."

The grammatical (as opposed to logical) categories here are:

- nouns: *Hamlet, father, king*
- verbs: *is*
- determinators: *a, the*

The sentence (S) above is constructed from a noun phrase (NP) "*the father of Hamlet*" followed by a verb phrase (VP) "*is a king*." The noun phrase "*the father of Hamlet*" is constructed from the noun phrase "*the father*" and the prepositional phrase (PP) "*of Hamlet*." The noun phrase "*the father*" is constructed from a determinator (D) "*the*" and a noun phrase which itself consists of a noun (N) "*father*." The prepositional phrase "*of Hamlet*" is constructed from a preposition (P) "*of*" and a noun phrase which itself consists of a noun, "*Hamlet*." The verb phrase "*is a king*" is constructed from a verb (V) "*is*," and a noun phrase "*a king*." The latter is constructed from a determinator (D) "*a*," and a noun (N) "*king*." One can represent the above grammatical analysis as an array:

$$
\begin{array}{lllllll}
S & & & & & & \\
NP & & & & VP & & \\
NP & & PP & & V & NP & \\
D & N & P & N & V & D & N \\
\text{the} & \text{father} & \text{of} & \text{Hamlet} & \text{is} & \text{a} & \text{king}
\end{array}
$$

The above may be referred to as a *grammatical sentence formation*; such a formation is something quite different from the (logical) *formula/sentence formations* based on logical analysis. One can add edges to the array above as follows: each entry X in a given row is linked by an edge to the closest entry Y in the previous row that is above or to the left of X. In this way we get an inverted tree. Such inverted trees are a basic tool in the work of Chomsky on generative grammar, for instance. Alternatively one can encode the information contained in a grammatical sentence formation as follows. One enriches English by adding separators $[_S$ and $]_S$ for sentences, $[_{NP}$ and $]_{NP}$ for noun phrases, etc., and one encodes the above sentence formation as a string of words in this enriched English:

$$[_S[_{NP}[_{NP}[_D\text{the}]_D[_N\text{father}]_N]_{NP}[_{PP}[_P\text{of}]_P[_N\text{Hamlet}]_N]_{PP}]_{NP}[_{VP}...]_{VP}]_S.$$

Grammatical sentence formations are obtained by applying substitution rules symbolically written, for instance, as

$S \to NP \;\; VP$
$NP \to NP \;\; PP$
$VP \to V \;\; NP$
$NP \to D \;\; NP$
$NP \to N$

$PP \to P \quad NP$
$N \to$ father
$N \to$ Hamlet
etc.

More complicated rules are, of course, present in English. On the other hand very simple "super-rules" that generate these complicated rules in virtually all languages have been discovered. This way of looking at natural languages such as English has deep consequences in psychology and the philosophy of mind; however this approach does not seem appropriate for the study of languages such as Formal or other languages of interest to science and Mathematics. Mathematics requires logical (rather than grammatical) analysis.

Skepticism, nihilism, historicism. It has often been claimed that skepticism, nihilism, and historicism are self-defeating philosophical positions. However these claims can be refuted if one consistently applies the distinction between language and metalanguage. Here are the details.

A person A who maintains that skepticism is a self-defeating philosophical position usually argues as follows: if the skeptic B says "I am a skeptic" then B should be skeptic about his/her own statement so his/her position cannot be upheld, hence is "self-defeating." However a person C who accepts the distinction between languages and metalanguage could refute A as follows: "I am a skeptic," says C, should be paraphrased as a collection $\widehat{S}$ of metasentences $\widehat{P}$ in the metalanguage $\widehat{L}$ (one for each constant P in $\widehat{L}$ which is the name of a sentence in a language L) of the form:

$\widehat{P}$=“*I am skeptical about P*”.

What A actually does, says C, is to claim that the string of symbols

$\widehat{\widehat{P}}$=“*I am skeptical about* $\widehat{P}$”

belongs to $\widehat{S}$; but this is an error on the part of A, says C, because $\widehat{P}$ is not the name of a sentence in L but the name of sentence in $\widehat{L}$. So $\widehat{\widehat{P}}$ is not syntactically correct hence, according to C, A's argument does not work.

Similarly it has been maintained that nihilism is a self-defeating philosophical position. (This was repeatedly used to make a case against Nietzsche, for instance.) A person A who maintains that nihilism is a self-defeating philosophical position usually argues as follows: if the nihilist B says "Nothing has value" then B should agree that his/her own opinion has no value which makes that opinion self-defeating. However a person C who accepts the distinction between languages and metalanguage could refute A as follows: "Nothing has value" should be paraphrased, says C, as a collection S of sentences $\widehat{c}$ in the language L (one for each constant c in L) of the form:

$\widehat{c}$=“*c has no value*”.

What A actually does, says C, is to claim that the string of symbols

$\widehat{\widehat{c}}$=“$\widehat{c}$ *has no value*”

belongs to S; but this is an error on the part of A, says C, because $\widehat{c}$ is not a constant in L but a constant in $\widehat{L}$. So $\widehat{\widehat{c}}$ is not syntactically correct hence, according to C, A's argument does not work.

Similarly it has been maintained that historicism in philosophy is a self-defeating position. (This was repeatedly used to make a case against Hegel, Marx, Foucault, etc., for instance.) A person A who maintains that historicism in philosophy is a self-defeating position usually argues as follows: if the historicist B says "Every philosopher merely expresses the view of his/her society/class at one historical moment" then B should agree that his/her own opinion is an expression of his/her society/class at one historical moment which makes that opinion non-objective and hence self-defeating. However a person C who accepts the distinction between languages and metalanguage could refute A as follows: "Every philosopher merely expresses the position of his/her society/class at one historical moment" should be paraphrased, says C, as a collection $\widehat{S}$ of sentences $\widehat{P}$ in the metalanguage $\widehat{L}$ (one for each constant P in $\widehat{L}$ which is the name of a sentence in L) of the form:

$\widehat{P}$="*The sentence P merely reflects the position of its author's society/class at one historical moment*".

What A actually does, says C, is to claim that the string of symbols

$\widehat{\widehat{P}}$="*The sentence $\widehat{P}$ merely reflects the position of its author's society/class at one historical moment*"

belongs to $\widehat{S}$; but this is an error on the part of A, says C, because $\widehat{P}$ is not the name of a sentence in L but the name of a sentence in $\widehat{L}$. So $\widehat{\widehat{P}}$ is not syntactically correct hence, according to C, A's argument does not work.

Bibliography

[1] Aristotle. 2001. *The Basic Works of Aristotle*, Modern Library Classics.
[2] Ayer, A. J. 1952. *Language, Truth, and Logic*. New York. Dover.
[3] Badiou, A. 2009. *Logics of Worlds*. Bloomsbury, London, New York.
[4] Cantor, G. 1955. *Contributions to the Founding of the Theory of Transfinite Numbers*. New York: Dover.
[5] Carnap, R. 1966. *An Introduction to the Philosophy of Science*, Dover.
[6] Cassirer, E. 1953. *Language and Myth*. New York: Dover.
[7] Chomsky, N. 2006. *Language and Mind*. Cambridge University Press.
[8] Cohen, P. 2008. *Set Theory and the Continuum Hypothesis*, Dover.
[9] Gödel, K. 1992. *On formally undecidable propositions of Principia Mathematica and related systems*. New York: Dover.
[10] Hegel, G.W.F. 1975. *Hegel's Logic: Being Part One of the Encyclopaedia of the Philosophical Sciences*. Oxford University Press.
[11] Hilbert, D. 1994. *Foundations of Geometry*. La Salle: Open Court.
[12] Hilbert, D., Bernays, P. 1934. *Grundlagen der Mathematik*. Berlin.
[13] Kant, I. 1991. *Critique of Pure Reason*. London: J. M. Dent & Sons.
[14] Manin, Yu. I. 2009. *A Course in Mathematical Logic for Mathematicians*. New York: Springer.
[15] Putnam, H. 1971. *Philosiphy of Logic*. New York, San Francisco, London: Harper and Row.
[16] Putnam, H. 1981. *Reason, Truth, and History*, Cambridge Univ. Press.
[17] Quine, W. V. 1964. *Word and Object*, MIT Press.
[18] Quine, W. V. 1980. *From a Logical Point of View*, Harvard University Press.
[19] Quine, W. V. 1986. *Philosophy of Logic*. Cambridge: Harvard University Press.
[20] Russell, B. 1993. *Introduction to Mathematical Philosophy*. New York: Dover.
[21] Tarski, A. 1995. *Introduction to Logic and to the Methodology of Deductive Sciences*, New York: Dover.
[22] Wang, Hao. 1996. *A Logical Journey: from Gödel to Phylosophy*, Cambridge, MIT Press.
[23] Weyl, H. 1963. *Philosophy of Mathematics and Natural Sciences*. New York: Atheneum.
[24] Wittgenstein, L. 2001. *Tractatus Logico-Philosophicus*. Routledge: Taylor and Francis.

Index