

Introduction to mathematical thinking

Alexandru Buium

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NEW MEX-
ICO, ALBUQUERQUE, NM 87131, USA

E-mail address: `buium@math.unm.edu`

INTRODUCTION

Mathematics evolved as 3 intertwined sciences: Geometry (dealing with figures), Algebra (dealing with numbers), and Calculus (dealing with functions). It eventually came to be seen as part of Logic (which deals with the rules of thought).

- Geometry and Algebra emerged around the 4th century BC in ancient Greece and evolved at first separately due to the discovery (due to the Pythagoreans) that integer numbers cannot account for relations between geometric quantities such as the side and the diagonal of the square.

- Geometry was reduced to Algebra mainly through the work of Descartes and Fermat (17th century) who invented analytic geometry and through subsequent work of Euler, Lagrange (18th century), and Plücker, Möbius, Riemann (19th century).

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

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

- 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 sense of these notes and his concepts were soon discovered to lead to contradictions (as in Russell's paradox).

- Logic was first developed by Aristotle (4th century BC). 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 sense of these notes; in this theory Russell's paradox does not seem to arise anymore (although it is possible that a similar paradox 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 attempts at axioms (cf. von Neumann, Gödel, etc.) were made which we will ignore.

- A 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.

In these notes we follow a "reversed historical order" of these developments (which is at the same time the necessary order of exposition from the logical viewpoint). We begin with Logic (as a prerequisite for mathematics which could be called "Pre-Mathematical Logic"). Then we introduce Set Theory as one example of a theory within Logic and we identify Set Theory with Mathematics. We then show how the concepts of Algebra can be defined within Set Theory. Next we show how the concepts of Geometry and Calculus can be defined using 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.

Although the above account of mathematics is generally accepted today there are different ways of interpreting this account from the viewpoint of the philosophy of mathematics and language.

The viewpoint of these notes roughly fits into what is referred to as the *nominalist* (or *formalist*) position in the philosophy of mathematics. Nominalism is the ontological position according to which there is nothing beyond abstract concepts (such as “sets,” “numbers,” “figures,” “functions,” “infinity,” “justice,” “beauty”, etc.) except the words themselves and the role of these words in discourse. At the opposite end of the ontological spectrum one finds what is referred to as *realism* (or *Platonism*). Realism holds that abstract concepts do not reduce to the words that are their names; rather, these concepts correspond to entities existing in a/the “real world.” 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. There is yet another school in the philosophy of mathematics referred to as *intuitionism*; however the theorems of intuitionist mathematics are different from those of formalist or Platonist mathematics and they are not regarded today as “main stream” mathematics.

Formalism was initiated by Hilbert (who cites Kant as a precursor). Platonism is a widespread position among working mathematicians of all times and some of the great logicians (e.g., Gödel) are associated with it. Intuitionism was initiated by Brouwer (and also invokes the Kantian tradition). There is more than one way to present the formalist view. In particular the present notes differ from the classical viewpoint of Hilbert in a number of key aspects. We will make no attempt to compare our approach with other approaches.

Modern philosophy of mathematics roughly described above 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 which, however, possesses a whole hierarchy of “types” (of arbitrary “order”). This position underwent substantial changes in subsequent developments of logic. In particular the formalist approach of these notes will postulate a whole variety of languages (each of which is of “first order”) and a whole variety of relations between such languages.

Note. These notes are a (drastically) simplified version of my book: *Mathematics: a minimal introduction*, CRC press, 2013. However one should be aware of the fact that there are some key conceptual differences between the present notes and the book, especially when it comes to the material that pertains to logic.

Contents

Part 1. Logic	7
Chapter 1. Languages	9
Chapter 2. Metalanguage	19
Chapter 3. Syntax	25
Chapter 4. Tautologies	31
Chapter 5. Proofs	37
Chapter 6. Theories	47
Part 2. Mathematics	57
Chapter 7. ZFC	59
Chapter 8. Sets	65
Chapter 9. Maps	71
Chapter 10. Relations	75
Chapter 11. Operations	81
Chapter 12. Integers	87
Chapter 13. Induction	91
Chapter 14. Rationals	95
Chapter 15. Combinatorics	97
Chapter 16. Sequences	99
Chapter 17. Reals	103
Chapter 18. Topology	105
Chapter 19. Imaginaries	109
Chapter 20. Arithmetic	111
Chapter 21. Groups	115

Chapter 22. Vectors	121
Chapter 23. Matrices	123
Chapter 24. Determinants	127
Chapter 25. Polynomials	131
Chapter 26. Congruences	135
Chapter 27. Geometry	139
Chapter 28. Limits	147
Chapter 29. Series	151
Chapter 30. Trigonometry	155
Chapter 31. Calculus	157
Chapter 32. Categories	161
Chapter 33. Models	173
Commentary	177
Bibliography	187

Part 1

Logic

CHAPTER 1

Languages

In these notes Mathematics will be identified with *Set Theory*. Theories are viewed as (dynamical/growing) sequences of sentences. The formation of sentences and theories is governed by (general or pre-mathematical) logic (not to be confused with mathematical logic which we view as part of mathematics). Logic starts with the logical analysis of natural languages. To explain this we will introduce here two languages: English (which is one of many natural languages such as Chinese, Japanese, French, German, etc.) and Formalese (which is one of many possible formalized languages; the term *Formalese* is not standard and was concocted for the purposes of these notes only). We shall then examine the interconnections between these two languages. The approach we present was pioneered by Bertrand Russell although there will be differences between Russell's presentation and ours.

Let us start with a discussion of English. 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 Socrates 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) "*Socrates is a man*"
- 2) "*Caesar killed Brutus*"
- 3) "*The killer of Caesar is Brutus*"
- 4) "*Brutus killed Caesar and Socrates is a man*"
- 5) "*Brutus is not a man or Caesar is a killer*"
- 6) "*If Brutus killed Caesar then Brutus is a killer*"
- 7) "*Brutus did not kill Caesar*"
- 8) "*A man killed Caesar*"
- 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 from a surrounding text we put them between quotation marks (and sometimes we write them in italics). So quotation marks do not belong to the language but rather they lie outside the language; they belong to *metalinguage*, 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, a person, ...*” (sometimes replaced by “*x, y, z, ...*”)
- constants: “*Socrates, Brutus, Caesar, ...*”
- 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.* In general objects are named by constants or variables. (So constants and variables roughly correspond to proper nouns.)

Constants are names for specific objects 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.).

Quantifiers specify quantity and are always followed by variables.

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, and functions applied to them (these are the objects that one is talking about), and finally one identifies the functions (which say something about the objects). The above type of analysis (called *logical analysis*) is quite different from the *grammatical analysis* based on the grammatical categories of *nouns, verbs, etc.*

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

- variables: “ x, y, \dots ”
- constants: “ S, B, C, \dots ”
- functions: “ \downarrow, \dots ”
- 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 dictionary) that attaches to each symbol in Formalese a symbol (or possibly several symbols) in English as follows:

- “ x, y, z, \dots ” are translated as “*something, a thing, an entity, \dots*” (or simply by “ x, y, z, \dots ”)
- “ S, B, C ” are translated as “*Socrates, Brutus, Caesar*”
- “ \downarrow, \dots ” are translated as “*the killer of, \dots*”
- “ 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*”

Then the English sentences 1-10 are translations of the following Formalese sentences; equivalently the following Formalese sentences are translations (called *formalizations*) of the corresponding English sentences:

- 1') “ mS ”
- 2') “ $C \dagger B$ ”
- 3') “ $\downarrow (C) = B$ ”
- 4') “ $(B \dagger C) \wedge mS$ ”
- 5') “ $(\neg(mB)) \vee kB$ ”
- 6') “ $(B \dagger C) \rightarrow (kB)$ ”
- 7') “ $\neg(B \dagger C)$ ”
- 8') “ $\exists x(x \dagger C)$ ”
- 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 “ $m(S)$ ” instead of “ mS .” Also one can write “ $\dagger(C, B)$ ” or “ $\dagger CB$ ” instead of “ $C \dagger B$.” One also says that 1) is *analyzed* as 1'), etc.

REMARK 1.1.

a) The above formalizations are obtained in two steps. The first step (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 Caesar*” or simply “*There exists x such that x is a man and x killed Caesar.*”

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 "

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 "; these schemes were introduced by Russell.) The second step in translation 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 " $P \rightarrow Q$ " where "if" and "then" are replaced by \rightarrow 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 Formalese.

b) One could have paraphrased 1) by

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

or even

1'") "*Socrates 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 Formalese by " \in " and one could treat "*the totality of all (things that are) men*" as a constant which could be translated in Formalese by " M ." Then the translation of 1") or 1'") into Formalese 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.

c) 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."

d) 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).

e) All of our discussion of English and Formalese 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 *metalanquages*). The very latter sentence is written in Metalanguage; and indeed the whole course is written in Metalanguage.

REMARK 1.2. (Naming) It is useful to give names to sentences. For instance if we want to give the name P to the English sentence “*Socrates is a man*” we can write the following sentence in Metalanguage:

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

So neither P nor the word *equals* nor the quotation marks belong to English; and “*Socrates 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 Formalese 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 1.3. (Definitions) A language may be enlarged by definitions. More precisely one can add new predicates or constants to a language by, at the same time, recording certain sentences, called definitions. As an example for the introduction of a new predicate in English we can add to English the predicate *is an astrochicken* by recording the following sentence:

DEFINITION 1.4. 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 1.5. An astrochicken is something which is a chicken and also a space ship.

DEFINITION 1.6. Something is called astrochicken if it is a chicken and also a space ship.

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

$$\text{DEFINITION 1.7. } \forall x(\epsilon(x) \leftrightarrow ((x \in c) \wedge (x \in s))).$$

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.

In the above discussion we encountered 2 examples of languages that we described in some detail (English and Formalese) 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 Remark 1.3 because definitions in that sense 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 Remark 1.3.

REMARK 1.8. (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 Formalese 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”), “programming 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, \dots to the sentences in L ; these names P, \dots 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 Formalese 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 Formalese; 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 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 notes this approach will be sufficient.

REMARK 1.9. (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 “*Socrates*” refers to the “*physical man Socrates*”; the words “*the word Socrates*” refers to the “*physical word Socrates*” 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 Socrates” should be replaced by the “phenomenal” (as opposed to the “noumenal” Socrates) or by the “idea” of Socrates (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 Formalese; 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 Formalese 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 declarative sentences based on other declarative sentences that are already accepted; see the comments below on declarative 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 $1, \dots, 10, 1', \dots, 10'$ are “true” or “false.” We will not define *truth/falsehood* for sentences in any of our languages. Indeed a theory 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.

REMARK 1.10. (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).

EXAMPLE 1.11. (Fixed number of constants) English and Formalese 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''', \dots starting from a constant c ; one can form variables x', x'', x''', \dots from a variable x ; and one can do the same with functions, predicates, etc.; we will not pursue this in what follows.

EXAMPLE 1.12. (Alternative translations) We already gave an example of translation of Formalese 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 “*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 sometimes “*for any,*” although the latter may lead to confusion: e.g., the sentence “if for any of you rock and roll is pure noise then...” is actually paraphrased as “if there exists x such that x belongs to your group and rock and roll is pure noise for x then...”; so “*for any*” sometimes needs to be paraphrased as “*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 .”)

Alternative translations of \exists into English are “*for some*” or “*there is an/a.*”

English has many 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 “*many,*” “*most,*” “*quite a few,*” “*half,*” “*for at least three,*” etc.); we will ignore these other quantifiers.

REMARK 1.13. (Texts) Let us consider the following types of objects:

- 1) symbols (e.g., $x, y, B, C, \dagger, k, \dots, \in, \downarrow, \dots, \wedge, \vee, \neg, \rightarrow, \leftrightarrow, \forall, \exists, =, (,)$);
- 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.**

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

EXERCISE 1.14. 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 1.15. 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*.”

REMARK 1.16. 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 6.12 for more on this. The sentence 4 is a version of the “cosmological argument” (Aquinas).

REMARK 1.17. (Declarative/imperative/interrogative sentences) All sentences considered so far were declarative (they declare their content). Natural languages have other types of sentences: imperative (giving a command like: “Lift this weight!”) and interrogative (asking a question such as: “Is the electron in this portion of space-time?”). In principle, from now on, we will only consider declarative sentences in our languages. An exception to this is the language called Argot; see below.

EXAMPLE 1.18. For a language L (such as Formalese) 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 one more category of symbols,

- *commands*: “*consider,*” “*assume,*” “*let...be,*” “*let us...,*” 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 $\rho(c)$.*”

We will not insist on explaining the syntax of Argot which is rather different from that of both English and Formalese. Suffices to note that the symbols in Formalese 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.

CHAPTER 2

Metalinguage

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 *metalinguage*. (The term *metalinguage* was used by Tarski in his theory of truth; but our metalinguage 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.)

Metalinguages 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 metalinguage are called *metasentences*. If we treat English and Formalese as object languages then all our discussion of English and Formalese was written in a metalinguage (which is called Metalinguage) 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 Formalese (or several object languages L, L', \dots). In what follows we introduce a metalinguage \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 Formalese):

- 1) x is a variable in the sentence “ $\forall x(x \in a)$.”
- 2) P equals “*Socrates 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, \dots, U is a proof of V .
- 8) V is a theorem.
- 9) If P, \dots, 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 **there are no variables and no quantifiers in metalanguage**. Next we have:

- constants: “Socrates,” “Socrates is a man,” “ \wedge ,” “=,” $L, L^*, L^s, \dots, P, Q, R, \dots$,
- functions: the variables in, the translation of, the proof of, $\wedge, \vee, \neg, \rightarrow, \leftrightarrow, \exists x, \forall x, \dots$
- 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, \dots ” used as constants, the symbols \wedge, \dots 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 *MetaFormalese* 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 Formalese, then the following is a metasentence:

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

Allowing the above to be a metasentence helped Tarski define *truth in a language* (the Tarski *T* scheme); we will not do 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 we keep less precise than that of 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.

This “double standard” approach 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 designate.

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”).

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 that we explained above are introduced in order to avoid circularity and/or infinite regression in the world of languages and is 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 $\wedge, \vee, \dots, \exists x$ are functions in metalanguage one can form syntactically correct strings

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

in metalanguage, etc. If

P equals “ $p\dots$ ”

Q equals “ $q\dots$ ”

where p, \dots, q, \dots are symbols in the object language then we introduce the rule (“metaaxiom”)

$P \wedge Q$ equals “ $(p\dots) \wedge (q\dots)$ ”

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 “ $p\dots$ ”

Then we say P is a *name* for the sentence “ $p\dots$ ” 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 “ $p\dots x\dots$ ”

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:

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

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 “*Socrates is a man*” and “*all men are mortal*” it follows that “*Socrates is mortal*.”

Its disquotation is the following sentence in English:

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

Note that 1 refers to some sentences in English whereas 2 refers to something (somebody) called Socrates. 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 “*Socrates is a man*,” Q equals “*all men are mortal*,” and R equals “*Socrates 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*” 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. (Declarative/imperative/interrogative) All metasentences considered so far were declarative (they declare their content). There are other types of metasentences: 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 declarative 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 Formalese (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, \dots , 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, \dots , a functional symbol f , and a binary predicate \square then here are examples of strings in L^* :

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

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, \dots are terms and f is a functional symbol then $f(t, s, \dots)$ 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, \dots are constants, x, y, \dots are variables, f is a unary functional symbol, and g is 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, \dots are terms and ρ is a predicate then $\rho(t, s, \dots)$ 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, \dots appear in a formula P we sometimes write in metalanguage $P(x, y, \dots)$ 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 $\rho(t, s, \dots)$. If x does not appear in the terms t, s, \dots and if ρ is a predicate then x is bound in the formula $\rho(t, s, \dots)$.
- 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 $\exists xP$ and $\forall xP$ are formulas and x is bound in $\exists xP$ and $\forall xP$.
- 5) If P is a formula and x is free in P then x is free in $\exists yP$ and in $\forall yP$. If P is a formula and x is bound in P then x is bound in $\exists yP$ and in $\forall yP$.

Formulas of the form (i.e., which are equal to one of) $\forall xP$, $\forall xP(x)$ are referred to as universal formulas. Formulas of the form $\exists xP$, $\exists xP(x)$ 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 $\exists xq(x, y)$ 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 $(\forall y(\exists x q(x, y))) \rightarrow \forall(\neg p(x, y, z))$. One cannot metaprove the metasentence “the above is a formula” the way we proceeded in the previous example (the problem being with the symbol \forall). 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*.

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, \dots are variables and t, s, \dots are terms, we may replace all free occurrences of these variables by t, s, \dots to get a formula $P \frac{ts\dots}{xy\dots}$. 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 $P(t, s, \dots)$ instead of $P \frac{ts\dots}{xy\dots}$. We will constantly use this $P(t), P(t, s, \dots)$, 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 “Socrates” then $P(a)$ equals “Socrates 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 “Socrates” then $P(a)$ equals “Socrates 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.”

The following metadefinition makes the concept of *definition* in L more precise:

METADEFINITION 3.19. A definition in L is a sentence of one of the following types:

- 1) (Definition of a new constant c): “ $P(c)$ ” where $P(x)$ is a formula with one free variable; in this case c is called an existential witness for $P(x)$.
- 2) (Definition of a new unary predicate ϵ): “ $\forall x (\epsilon(x) \leftrightarrow E(x))$ ” where E is a formula with one free variable. More generally for several variables, “ $\forall x \forall y (\epsilon(x, y) \leftrightarrow E(x, y))$ ” is a definition of ϵ , etc.
- 3) (Definition of a new unary functional symbol f): “ $\forall x \forall y ((y = f(x)) \leftrightarrow F(x, y))$ ” where F is a formula with 2 free variables; more generally one allows several variables.

If any type of symbols is missing from the language we disallow, of course, the corresponding definitions.

A language together with a collection of definitions is referred to as a language with definitions.

REMARK 3.20. Given a language L with definitions and a term t without variables one can add to the language a new constant c and one can add to the definitions the definition $P(c)$. We will say that c is (a new constant) defined by $P(c)$.

Similarly given a language L with definitions and a formula $E(x)$ in L with one free variable x one can add to the predicates of L a new unary predicate ϵ and one

can add to the definitions the definition $\forall x(\epsilon(x) \leftrightarrow E(x))$. We will say that ϵ is (a new predicate) defined by $\forall x(\epsilon(x) \leftrightarrow E(x))$.

One may introduce new functions f by adding a symbol f and definitions of type 3 or of type 1 (such as $c = f(a, b, \dots)$, for various constants a, b, c, \dots).

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.

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 “falseness”; 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 ”.

METAXIOM 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 $P \rightarrow (Q \vee \neg R)$ and $P \wedge R$:

$$\begin{aligned} &P \rightarrow (Q \vee \neg R) \\ &P \wedge R \\ &(P \wedge R) \leftrightarrow (P \rightarrow (Q \vee \neg R)) \end{aligned}$$

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	\dots	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, \dots . Then the column of V is obtained from the columns of V' and V'' using the tables of the corresponding connective \wedge, \vee, \dots , respectively. The above metadefinition is, however, not correct because it implicitly involves quantifiers. So one has to reject this as a metadefinition and one proceeds, instead, as follows. 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.

EXAMPLE 4.9. Consider the following Boolean string generated by P and Q :

$$\begin{aligned} &P \\ &Q \\ &\neg P \\ &\neg P \wedge Q \end{aligned}$$

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 4.10. 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 T s in the V column then V is a tautology. Same metaaxiom for more or less than 3 sentences.

REMARK 4.11. We do not ask (and in view of the fact that quantifiers in metalanguage are forbidden, we cannot ask), in the metadefinition above, that for every Boolean string ending in S and for every generators the last column of the truth table have only T s; we only ask this to hold for “one Boolean string” and “one system of generators.”

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

EXAMPLE 4.12. $P \vee \neg P$ is a tautology. To metaprove this consider the Boolean string generated by P ,

$$\begin{array}{l} P \\ \neg P \\ P \vee \neg P \end{array}$$

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

$$\begin{array}{l} P \\ \neg P \\ P \vee \neg P \end{array}$$

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 T s 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 4.13. Let S be the sentence $(P \wedge (P \rightarrow Q)) \rightarrow Q$. This sentence is a tautology which is called *modus ponens*. To metaprove this consider the following Boolean string generated by P, Q :

$$\begin{array}{l}
P \\
Q \\
P \rightarrow Q \\
P \wedge (P \rightarrow Q) \\
(P \wedge (P \rightarrow Q)) \rightarrow Q
\end{array}$$

Its truth table is:

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

EXERCISE 4.14. Explain how the table above was computed.

EXERCISE 4.15. Give a metaproof of the fact that each of the sentences below is a tautology:

- 1) $(P \rightarrow Q) \leftrightarrow (\neg P \vee Q)$.
- 2) $(P \leftrightarrow Q) \leftrightarrow ((P \rightarrow Q) \wedge (Q \rightarrow P))$.

EXERCISE 4.16. Give a metaproof of the fact that each of the sentences below is a tautology:

- 1) $(P \wedge Q) \rightarrow P$.
- 2) $P \rightarrow (P \vee Q)$.
- 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 4.17.

- 1) $Q \rightarrow P$ is called the converse of $P \rightarrow Q$.
- 2) $\neg Q \rightarrow \neg P$ is called the contrapositive of $P \rightarrow Q$.

EXERCISE 4.18. Give a metaproof of the fact that each of the sentences below is a tautology:

- 1) $((P \vee Q) \wedge (\neg P)) \rightarrow Q$ (modus ponens, variant).
- 2) $(P \rightarrow Q) \leftrightarrow (\neg Q \rightarrow \neg P)$ (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) $((P \rightarrow R) \wedge (Q \rightarrow R)) \rightarrow ((P \vee Q) \rightarrow R)$ (case by case argument).
- 6) $\neg(P \rightarrow Q) \leftrightarrow (P \wedge \neg Q)$ (negation of an implication).
- 7) $\neg(P \leftrightarrow Q) \leftrightarrow ((P \wedge \neg Q) \vee (Q \wedge \neg P))$ (negation of an equivalence).

REMARK 4.19. 2) in Exercise 4.18 says that the contrapositive of an implication is equivalent to the original implication.

REMARK 4.20. (VERY IMPORTANT) Give an example of sentences P, Q in a language such that the sentence

$$(P \rightarrow Q) \leftrightarrow (Q \rightarrow P)$$

is not a tautology. In other words the converse of an implication is not equivalent to the original implication. Here is an example in English suggestive of the above:

“*If c a human then c mortal*” is not equivalent to “*If c is mortal then c is a human.*”
(Hint: take P to be $a = b$ and Q to be $\neg(a = b)$).

METADefinition 4.21. A sentence P is a contradiction if and only if $\neg P$ is a tautology.

CHAPTER 5

Proofs

In this chapter we discuss inference (which is the process by which we construct proofs). As usual this chapter is written in metalanguage. Recall our language Argot L_{Argot} which is obtained from the symbols of English L_{Eng} , plus the symbols of a language L (such as Formalese), plus command symbols (such as “let,” “consider,” etc.)

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

METADEFINITION 5.1. A sentence in Argot is a string of symbols of one of the following forms:

- 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 c is such that $P(c)$.
 - 13) By P there exists c such that $Q(c)$.
 - 14) 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. 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”, etc.

METAAXIOM 5.2. Let A, B, \dots be sentences in a language L ; call A, B, \dots axioms. A sequence T of axioms 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 (called *definitions*, *axioms*, or *theorems*) subject to certain rules to be discussed below then T' is a theory; we say T' is an extension of T .

Here are some clarifications of the above metaaxiom.

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 every terms t, t', t'' , functions f , and predicates p :

$$\begin{aligned} t &= t \\ (t = t') &\rightarrow (t' = t) \\ ((t = t') \wedge (t' = t'')) &\rightarrow (t = t'') \\ (t = t') &\rightarrow (f(t) = f(t')) \\ (t = t') &\rightarrow (p(t) \leftrightarrow p(t')) \end{aligned}$$

When a definition is added to T one also needs to add to L the newly introduced corresponding symbol.

If $\exists xP(x)$ is a theorem or an axiom one can add a NEW constant c to L and one can add to the theory the Definition “ $P(c)$ ” or, in Argot, “Let c be such that $P(c)$ ”. Then c is called an *existential witness* for the formula $P(x)$ or for the sentence $\exists xP(x)$. If P equals “ $x = t$ ” where t is a term without variables then one can add the definition “ $c = t$ ” without having previously included a theorem of the form “ $\exists x(x = t)$ ”; indeed “ $\exists x(x = t)$ ” will automatically be a theorem as we will see from the discussion of proofs below.

Apart from the above procedure of addition of theorems a theorem can be added to T only if a proof for it is also being supplied. 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 $H \rightarrow C$. The sentence H is called the hypothesis and C is called the conclusion.

Rule 1. 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 2. If at one point in a proof P and Q are accepted sentences and if $(P \wedge Q) \rightarrow R$ 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 3. If at one point in a proof “ $\forall xP(x)$ ” is an accepted sentence then at any point later in the proof one is allowed to add the sentence “Since $\forall xP(x)$ 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 $\forall xP(x)$ 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 $\forall xP(x)$ Since $\forall xP(x)$ it follows that $P(c)$ \square

Rule 4. If at one point in a proof “ $\exists xP(x)$ ” 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 $\exists xP(x)$ 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 $\exists xP(x)$ Let c be such that $P(c)$ \square

Rule 5. 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 $\exists xP(x)$ ” to the proof, where c can be any constant new or not, and one declares “ $\exists xP(x)$ ” to be an accepted sentence. Schematically we have

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

Rule 6. If the theorem that is to be proved has the form “ $\forall xP(x)$ ” then one proceeds as follows. One begins the proof with the sentence “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 $P(c)$. Schematically we have

Proof. Let c be arbitrary; we want to prove that $P(c)$ So $P(c)$. \square

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.

EXAMPLE 5.3. An example of a theory may look as follows:

(Constants: a, b)
 (Functions: f)
 (Predicates: p)
 Axiom. A
 Axiom. B
 Theorem. U
 Proof....
 (New constant: c)
 Definition. $c = t$
 Axiom. E
 Theorem. $\exists xP(x)$
 Proof....
 (New constant: d)

Definition. $P(d)$

Axiom. F

Theorem. $\forall x \forall y \forall z ((G(x, y) = G(x, z)) \rightarrow (y = z))$

Proof...

(New functional symbol: g)

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

Definition. I

Theorem. V

Proof...

(New predicate: q)

Definition. $\forall x (q(x) \leftrightarrow Q(x))$

Theorem. W

(New constant: e)

Proof...

etc. (the list can continue indefinitely)

Here:

A could be a specific axiom

B could be a general axiom

t is a term written using only symbols introduced before introducing c ,

P is a formula written using only symbols introduced before introducing d ,

G is a formula written using only symbols introduced before introducing g ,

Q is a formula written using only symbols introduced before introducing q ,

etc.

The definition of g is permitted only if the Theorem before is proved. The texts between parentheses are usually not included in the text of the proof.

We proceed by giving our examples of proofs. In these examples we assume we are dealing with a theory (which is not necessarily the one considered before) in which we have axioms A, B, \dots . 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 5.4. (Direct proof). A direct proof could look as follows.

Theorem. $H \rightarrow C$.

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

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

are accepted sentences.

EXAMPLE 5.5. (Proof by contradiction). Another method of proving sentences such as $H \rightarrow C$ is by contradiction and may look as follows.

Theorem. $H \rightarrow C$.

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 $H \rightarrow C$. \square

The above counts as a proof if

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

are accepted sentences.

In fact the above proof can be thought of as an abbreviation of the following direct proofs:

Theorem. $(H \wedge \neg C) \rightarrow (Y \wedge \neg Y)$

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. $H \rightarrow C$

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

The above counts as a proof because

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

is a tautology and hence it is an accepted sentence.

EXAMPLE 5.6. Direct proofs and proofs by contradiction can be given to sentences which are not necessarily of the form $H \rightarrow C$. 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

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

are accepted sentences.

EXAMPLE 5.7. (Case by case proof) Say we want to prove a theorem of the form:

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

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

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

are accepted sentences.

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

Theorem. $H \rightarrow C$

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

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

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 can be reduced, again, to direct proofs 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. $(H \wedge W) \rightarrow C$

Then give a direct proof of:

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

Then finally prove:

Theorem. $H \rightarrow C$

using the fact that

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

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 5.8. Here is an example that combines proof by contradiction with “case by case” proof. Say we want to prove:

Theorem. $H \rightarrow C$.

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. \square

EXAMPLE 5.9. 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 $P \rightarrow Q$ then $Q \rightarrow R$ then $R \rightarrow P$.

EXAMPLE 5.10. In order to prove a theorem of the form $P \wedge Q$ one first proves P and proves Q .

EXAMPLE 5.11. 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 5.12. In order to prove a theorem U of the form $P \leftrightarrow Q$ one first proves $P \rightarrow Q$ and then one proves $Q \rightarrow P$.

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 5.13. Here is an example of proof by contradiction that involves witnesses.

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

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. \square

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 5.14. 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 5.15. Prove the following:

- 1) $(\neg(\forall x P(x))) \leftrightarrow (\exists x(\neg P(x)))$
- 2) $(\neg(\forall x \forall y \exists z P(x, y, z))) \leftrightarrow (\exists x \exists y \forall z \neg P(x, y, z))$

EXAMPLE 5.16. 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 $\exists xP(x)$ by using Rule 5.

METADefinition 5.17. 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 5.18. Let P, Q be atomic sentences. Let S be the sentence

$$P \wedge Q$$

Let S', S'' be the sentences

$$S' = \neg(P \wedge Q)$$

$$S'' = \neg P \vee \neg Q$$

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 5.19. Let $P(x, y), Q(x, y)$ be atomic formulae. Let S be the sentence

$$S = \forall x \exists y (P(x, y) \rightarrow Q(x, y))$$

Let S', S'', S''' be the sentences

$$S' = \neg(\forall x \exists y (P(x, y) \rightarrow Q(x, y)))$$

$$S'' = \exists x (\neg(\exists y (P(x, y) \rightarrow Q(x, y))))$$

$$S''' = \exists x \forall y ((P(x, y) \wedge \neg Q(x, y)))$$

Then S' is a negation of S because $\neg S \leftrightarrow S'$ is a theorem, but S' is not in reduced form because \neg appears in S' in front of a non-atomic formula. Also S'' is a negation of S because $\neg S \leftrightarrow S''$ is a theorem, but, again, S'' is not in reduced form because \neg appears in S'' in front of a non-atomic formula. Finally S''' is a negation of S in reduced form. It is a negation because

$$\neg(A \rightarrow B) \leftrightarrow (A \wedge \neg B)$$

is a tautology (hence a theorem) and because

$$\neg(\forall x A(x)) \leftrightarrow \exists x (\neg A(x))$$

and

$$\neg(\exists x A(x)) \leftrightarrow \forall x (\neg A(x))$$

are theorems. S''' is in reduced form because it does not contain \rightarrow and \neg only appears in it in front of atomic formulae.

EXERCISE 5.20. Find negations in reduced form of the following sentences. 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 5.21. 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 $| \cdot |$ are unary functions and $<$ is a binary predicate. The sentence above plays a role in analysis.)

We next introduce the predicate “is inconsistent” into metalanguage.

METAAXIOM 5.22. 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 5.23. One is tempted to make a metadefinition among 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 these notes); 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 5.24. *Confusing an implication with its converse.* Say we want to prove that $H \rightarrow C$. A typical mistaken proof would be: Assume C ; then by ... we get that ... hence H . The error consists of having proved $Q \rightarrow P$ rather than $P \rightarrow Q$.

EXAMPLE 5.25. *Proving a universal sentence by example.* Say we want to prove $\forall xP(x)$. A typical mistaken proof would be: By ... there exists c such that ... hence ... hence $P(c)$. The error consists in having proved $\exists xP(x)$ rather than $\forall xP(x)$.

EXAMPLE 5.26. *Defining a constant twice.* Say we want to prove $\neg(\exists xP(x))$ by contradiction. A mistaken proof would be: Assume there exists c such that $P(c)$. Since we know that $\exists xQ(x)$ 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 5.27. 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 5.28. Later, when we discuss induction we will discuss another typical fallacy; cf. Example 13.7.

CHAPTER 6

Theories

We analyze in what follows a few toy examples of theories and proofs of theorems in these theories. Later we will present the main example of theory in this course which is Set Theory (identified with mathematics itself).

EXAMPLE 6.1. The first example is that of *group theory*. The language L of the theory has a constant e , variables x, y, \dots , and a binary functional symbol \star . We introduce the following definition in L :

DEFINITION 6.2. z is a neutral element if

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

So we added “*is a neutral element*” as a new unary predicate.

We also introduce the following

DEFINITION 6.3. y is an inverse of x if $x \star y = y \star x = e$.

So we added “*is an inverse of*” as a new binary predicate.

The axioms of the theory are:

AXIOM 6.4. For all x, y, z we have $x \star (y \star z) = (x \star y) \star z$.

AXIOM 6.5. e is a neutral element.

AXIOM 6.6. For all x there is y such that y is an inverse of x .

We prove the following:

THEOREM 6.7. *For all z if z is a neutral element then $z = e$.*

The sentence that needs to be proved is of the form: “If H then C .” Recall that in general for such a sentence H is called hypothesis and C is called conclusion. Here is a direct proof:

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 $e = e \star f$. So by one of the general axioms, $e = f$. \square

Here is a proof by contradiction of the same theorem:

Proof. Assume $e \neq f$ where f is a neutral element and seek a contradiction. So either $e \neq e \star f$ or $e \star f \neq f$. Since $\forall x(e \star x = x)$ it follows that $e \star f = f$. Since $\forall x(x \star f = x)$ we get $e \star f = e$. So we get $e = f$, a contradiction. \square

Here is another

THEOREM 6.8. *For all x, y, z if y and z are inverses of x then $y = z$.*

Proof Let a, b, c be such that b and c are inverses of a . We want to show that $b = c$. By Axioms 6.4, 6.5 and the definition of inverse we have

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

□

Group theory can be developed independently of what later we will call *Set Theory* (which is identified with mathematics); a better approach however is to replace the group theory above with a series of theorems in Set Theory.

The next example is related to the “Pascal wager.”

EXAMPLE 6.9. 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:

A1) If God exists and a person does not believe sincerely in its existence then that person will not be saved.

A2) If God does not exist then nobody will be saved.

A3) 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 6.10. *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

A1) $\forall y((\exists xg(x)) \wedge (\neg s(y)) \rightarrow \neg r(y))$.

A2) $\forall y((\neg(\exists xg(x))) \rightarrow (\neg r(y)))$.

A3) $\forall y(w(y) \rightarrow (\neg(s(y))))$.

In this language Theorem 6.10 is the translation of the following:

THEOREM 6.11. *If $w(p)$ then $\neg r(p)$.*

So to prove Theorem 6.10 in L' it is enough to prove Theorem 6.11 in L . We will do this by using a combination of direct proof and case by case proof.

Proof of Theorem 6.11. Assume $w(p)$. There are two cases: the first case is $\exists xg(x)$; the second case is $\neg(\exists xg(x))$. Assume first that $\exists xg(x)$. Since $w(p)$, by

axiom A3 it follows that $\neg s(p)$. By axiom A1 $(\exists xg(x)) \wedge (\neg s(p)) \rightarrow \neg r(p)$. Hence $\neg r(p)$. Assume now $\neg(\exists xg(x))$. By axiom A2 we then get again $\neg r(p)$. So in either case we get $\neg r(p)$ which ends the proof. \square

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).

EXAMPLE 6.12. 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:

A1) There exists a thing belonging to mind that has all the positive qualities and no negative quality. (Call it God.)

A2) “Being real” is a positive quality.

A3) Two things belonging to mind that have exactly the same qualities are identical.

(Axiom A3 is Leibniz’s famous principle of *identity of indiscernibles*. It implies that God is unique.) Then one can prove the following:

THEOREM 6.13. *God belongs to reality.*

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 6.13 in L whose translation is Theorem 6.13. Assume L contains among its constants the constants r, m, p and a binary predicate E . We consider a translation of L into L' such that

r is translated as “reality”;

m is translated as “mind”;

p is translated as “the positive qualities”

xEy is translated as “ x belongs to y ”.

The specific axioms are:

A1) $\exists x((xE m) \wedge (\forall z((zE p) \leftrightarrow (xE z))))$.

A2) $rE p$.

A3) $\forall x \forall y(((xE m) \wedge (yE m)) \rightarrow ((\forall z(xE z \leftrightarrow yE z)) \rightarrow (x = y)))$.

Note that later, in Set Theory, we will have a predicate \in which, like E , 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.

By A1 we may introduce a new constant g and we can make the following:

DEFINITION 6.14. $(gEm) \wedge (\forall z((zEp) \leftrightarrow (gEz)))$.

We will translate g as “God”. We have the following theorem expressing the uniqueness of God:

THEOREM 6.15. $\forall x(((xEm) \wedge (\forall z((zEp) \leftrightarrow (xEz)))) \rightarrow (x = g))$.

The translation of the above in English is: “*If something in my mind has all the positive qualities and no negative quality then that thing is God.*”

Proof. A trivial exercise using axiom A3 only. □

EXERCISE 6.16. Prove Theorem 6.15.

On the other hand, and more importantly, we have the following Theorem whose translation in L' is “*God belongs to reality*”:

THEOREM 6.17. gEr .

Proof. By axiom A1 we have gEm and

$$\forall z((zEp) \leftrightarrow (gEz)).$$

Hence we have, in particular,

$$(rEp) \leftrightarrow (gEr).$$

By axiom A2, rEp . Hence gEr . □

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 notes, 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 or 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.

EXAMPLE 6.18. 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):

A1) 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.

A2) 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 6.19. *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, \dots , variables x, y, \dots , 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

A1) $\forall x(p(x) \rightarrow (f(x) \wedge m(x)))$.

A2) $\forall x(w(x) \rightarrow (\neg f(x) \wedge \neg m(x)))$.

Here we tacitly assume that the number of maxima cannot be 1. Theorem 6.19 above is the translation of the following theorem in L :

THEOREM 6.20. $\forall x((f(x) \wedge (\neg m(x))) \rightarrow ((\neg p(x)) \wedge (\neg w(x))))$.

So it is enough to prove Theorem 6.20. 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 A1 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 A2 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 6.21. Consider the specific axioms A1 and A2 above and also the specific axioms:

A3) $\exists x(f(x) \wedge (\neg m(x)))$.

A4) $\forall x(p(x) \vee w(x))$.

Metaprove that the theory with specific axioms A1, A2, A3, A4 is inconsistent. A3 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 A4 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 6.22. 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, \dots , variables x, y, z, \dots , 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:

A1) For all x, y, z if x owns y and y owns z then x owns z .

A2) For all y there exists x such that x owns y .

A3) For all x, y, z if y owns x and z owns x then $y = z$.

We will prove the following

THEOREM 6.23. *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 A1, i owns i , a contradiction. Assume now d does not own i . By axiom A2 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 A1, j owns d . But i also owns d . By axiom A3, $i = j$, a contradiction. \square

EXAMPLE 6.24. 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, \dots , 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:

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

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

$$\begin{aligned} &\forall x (x : x = 1). \\ &\forall x \forall y \forall z \forall u ((x : y = z : u) \leftrightarrow (x \times u = z \times y)). \end{aligned}$$

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

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

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:

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

To this one adds the following “obvious” axioms

$$\begin{aligned} \text{O1) } &\forall x (c(x) \rightarrow d(x, E) = R), \\ \text{O2) } &M \circ E, \\ \text{R) } &c(r), \\ \text{K1) } &\forall x (p(x) \rightarrow (x \circ S)), \end{aligned}$$

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!):

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

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

$$\text{F1) } \forall x(c(x) \rightarrow f(x, E))$$

$$\text{F2) } f(M, E)$$

$$\text{F3) } \forall x(p(x) \rightarrow f(x, S))$$

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 definition

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

and the following sentence:

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

The main results are the following theorems in T :

THEOREM 6.25. $N \rightarrow X$.

Proof. See the exercise after this example.

THEOREM 6.26. $N \rightarrow G$.

Proof. See the exercise after this example.

THEOREM 6.27. $N \rightarrow K3$.

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:

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

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

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

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

$T(M, E)$ = 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 6.28. 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 6.29.

- 1) Justify Axiom A above using calculus or even Euclidean geometry plus the definition of acceleration in an introductory physics course.
- 2) Prove Theorems 6.25, 6.26, and 6.27.
- 3) Verify that with the numerical data for $g, \pi, R, d(M, E), T(M, E)$ available from astronomy (find the numbers in astronomy books) the sentence X becomes a theorem. This is Newton's fundamental computation.

Part 2

Mathematics

CHAPTER 7

ZFC

Mathematics is identified with a particular 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 included in the theory from the beginning; they have to be added one by one on a need to use basis to the various successive extensions of the theory.

METADEFINITION 7.1. The language L_{set} of Set Theory is the language with variables x, y, z, \dots , 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

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

and new predicates to L_{set} together with definitions for each of these new symbols. The constants 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 7.2.

$$\begin{aligned} \forall x \forall y ((x \neq y) &\leftrightarrow (\neg(x = y))). \\ \forall x \forall y ((x \notin y) &\leftrightarrow (\neg(x \in y))). \\ \forall x \forall y ((x \subset y) &\leftrightarrow (\forall z ((z \in x) \rightarrow (z \in y)))). \\ \forall x \forall y ((x \not\subset y) &\leftrightarrow (\neg(x \subset y))). \end{aligned}$$

DEFINITION 7.3. x is a subset of y if and only if $x \subset y$.

REMARK 7.4. 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 7.5. 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 7.6. There is a standard translation of the language L_{set} of Set Theory into the English language as follows:

- a, b, \dots are translated as “the set a ,” “the set b ,” ...
- \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*” 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.

REMARK 7.7. 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 7.8. 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, \dots are translated as “*crocodile a*,” “*crocodile b*,” ...
- \in is translated as “*is dreamt by the crocodile*”
- $=$ is translated as “*has the same taste*”
- \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 7.9. Note that mathematics uses other symbols as well such as

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

These symbols will originally be all sets (hence constants) and will be introduced through appropriate definitions (like the earlier definition of an elephant); they will all be defined through the predicate \in . In particular in the language of sets, $+$ or \times are NOT originally functions; and \leq is NOT originally a predicate. However we will later tacitly enlarge the language of Set Theory by adding predicates (usually still denoted by) \leq , etc., or functions \leq , etc., via appropriate definitions.

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

AXIOM 7.10. (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 7.11. (Unordered pair axiom)

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

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

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

AXIOM 7.12. (Separation axiom for P)

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

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 7.13.

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

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

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 7.14. 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 in the Exercise below. This contradiction is called the “Russell paradox.” So the separation axiom cannot be replaced the “modified separation axiom” above unless one drastically changes the syntax of languages by declaring that $x \notin x$ is NOT a well formed formula. 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, \dots ; type two X, Y, \dots ; type three $\mathcal{X}, \mathcal{Y}, \dots$; 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.

EXERCISE 7.15. Show that if one introduces the “modified separation axiom”

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

one can derive a contradiction. This is called the Russell paradox. Hint: Assume $\exists z \forall x ((x \in z) \leftrightarrow (x \notin x))$ and 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 $A \leftrightarrow \neg A$.)

EXERCISE 7.16. Try to derive a contradiction from the separation axiom

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

by an argument similar to the one used in the previous Exercise and show where this similar argument breaks down. This is an indication (but not a proof) that Russell's paradox should disappear if one considers the correct (i.e., unmodified) separation axiom.

AXIOM 7.17. (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 7.18. (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 .

AXIOM 7.19. (Empty set axiom)

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

The translation in Argot is that there exists a set that has no elements.

AXIOM 7.20. (Power set axiom)

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

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 7.21. Two sets are disjoint if they have no element in common. The elements of a set are pairwise disjoint if any two elements are disjoint.

AXIOM 7.22. (Axiom of choice) For every set w whose elements are pairwise disjoint sets there is a set that has exactly one element in common with each of the sets in w .

AXIOM 7.23. (Axiom of infinity) There exists a set x such that x contains some element u and such that for every $y \in x$ there exists $z \in x$ with the property that y is the only element of z . Intuitively this axiom guarantees the existence of "infinite" sets.

AXIOM 7.24. (Axiom of foundation) For every 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 7.25. (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 $P(x, y, z)$). Here x, z may be tuples of variables.

EXERCISE 7.26. Write the axioms of choice, infinity, foundation, and replacement in the language of sets.

METADefinition 7.27. 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 7.28. Note the important fact that the axioms did not involve constants. In the next chapter we will show how to introduce constants, i.e., the sets.

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 8

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 in this theory.

METADefinition 8.1. A set is a constant in the language of Set Theory.

Sets will be denoted by $a, b, \dots, A, B, \dots, \mathcal{A}, \mathcal{B}, \dots, \alpha, \beta, \gamma, \dots$

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

We start by defining a new constant \emptyset (called the empty set) as being a witness for the axiom $\exists x \forall y (y \notin x)$ in other words we add \emptyset to the language and we introduce:

Definition 8.2. $\forall y (y \notin \emptyset)$.

In Argot we say that \emptyset is a set that contains no element. Note that “there is only one empty set” in the sense that we have:

Theorem 8.3. *Every set that contains no element is equal to \emptyset .*

In L_{set} the above translates as:

$$\forall x ((\forall y (y \notin x)) \rightarrow (x = \emptyset)).$$

Proof. Let c be arbitrary. We want to show that if $\forall y (y \notin c)$ then $c = \emptyset$. Assume $\forall y (y \notin c)$. We want to show that $c = \emptyset$. By the extensionality axiom we have

$$(c = \emptyset) \leftrightarrow \forall x ((x \in c) \leftrightarrow (x \in \emptyset)).$$

So it is enough to show that

$$\forall x ((x \in c) \leftrightarrow (x \in \emptyset)).$$

Let d be arbitrary. We want to show that

$$(d \in c) \leftrightarrow (d \in \emptyset).$$

Since $\forall y (y \notin c)$ it follows that $d \notin c$. Since $\forall y (y \notin \emptyset)$ it follows that $d \notin \emptyset$. Since $d \notin c$ and $d \notin \emptyset$ it follows

$$(d \notin c) \leftrightarrow (d \notin \emptyset).$$

Hence

$$(d \in c) \leftrightarrow (d \in \emptyset).$$

□

Next for every set (i.e., constant) a we have:

Theorem 8.4. $\exists y ((a \in y) \wedge (\forall z ((z \in y) \rightarrow (z = a))))$.

Proof. By the singleton axiom we directly get the statement of the Theorem. \square

So we may introduce, at any point in the theory where a is a constant, a new constant $\{a\}$ defined to be a witness for the sentence in the above Theorem. In other words we may introduce the constant $\{a\}$ to the language and we add to the theory the following:

DEFINITION 8.5. $(a \in \{a\}) \wedge (\forall z((z \in \{a\}) \rightarrow (z = a)))$

We can say (and we will usually say, by abuse of terminology) that $\{a\}$ is “the unique” set containing a only among its elements (where uniqueness can be proved again using the extensionality axiom); 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, for every two sets a, b denote by $\{a, b\}$ the set that only has a and b as elements; the set $\{a, b\}$ is a witness for a theorem that follows from the unordered pair axiom.

EXERCISE 8.6. If $a = b$ then $\{a\} = \{a, b\}$.

Next, for every set A and every formula $P(x)$ in the language of sets, having one free variable x we denote by $A(P)$ or $\{a \in A; P(a)\}$ or $\{x \in A; P(x)\}$ the set whose elements are the elements $a \in A$ such that $P(a)$; so the set $A(P)$ equals by definition a witness for the separation axiom that corresponds to A and P . More precisely we have:

DEFINITION 8.7. $\forall x((x \in A(P)) \leftrightarrow ((x \in A) \wedge P(x)))$

The set $A(P)$ is called the *extension* of P . This definition corresponds to what in philosophical terminology is called the *extension of a concept*.

LEMMA 8.8. 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 8.9. Prove that:

- 1) If $\{a\} = \{b\}$ then $a = b$.
- 2) $\{a, b\} = \{b, a\}$.
- 3) If $a \neq b$ then $\{a\} = \{x \in \{a, b\}; x \neq b\}$.
- 4) There is a set b whose only elements are $\{a\}$ and $\{a, \{a\}\}$; so

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

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

DEFINITION 8.10. For every two sets A and B we define the set $A \cup B$ (called the union of A and B) as the set such that for all c , $c \in A \cup B$ if and only if $c \in A$ or $c \in B$; the set $A \cup B$ is a witness for the union axiom.

EXERCISE 8.11. Explain in detail the definition of $A \cup B$.

DEFINITION 8.12. The difference between the set A and the set B is the set

$$A \setminus B = \{c \in A; c \notin B\}.$$

DEFINITION 8.13. The intersection of the sets A and B is the set

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

EXERCISE 8.14. Prove that

- 1) $A(P \wedge Q) = A(P) \cap A(Q)$;
- 2) $A(P \vee Q) = A(P) \cup A(Q)$;
- 3) $A(\neg P) = A \setminus A(P)$.

EXERCISE 8.15. 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.

Also, inside any proof, we may denote by $\{a, b, c, \dots\}$ any set that contains a, b, c and possibly other sets. In other words we may introduce the constant $\{a, b, c, \dots\}$ and the

DEFINITION 8.16.

$$(a \in \{a, b, c, \dots\}) \wedge (b \in \{a, b, c, \dots\}) \wedge (c \in \{a, b, c, \dots\}).$$

Such a constant should not be used after the end of the proof where it is used; different notations should be used for such a constant in different proofs because the above definition does not pin down the “meaning” of $\{a, b, c, \dots\}$ completely.

EXERCISE 8.17.

- 1) Prove that $\{\emptyset\} \neq \emptyset$.
- 2) Prove that $\{\{\emptyset\}\} \neq \{\emptyset\}$.

EXERCISE 8.18. Prove that $A = B$ if and only if $A \subset B$ and $B \subset A$.

EXERCISE 8.19. 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 8.20. 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) $A \cap B = \{c\}$, $A \setminus B = \{a, b\}$.

EXERCISE 8.21. 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) $A \setminus B$,
- 4) $B \setminus A$,
- 5) $(A \setminus B) \cup (B \setminus A)$.

EXERCISE 8.22. 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) $(A \setminus B) \cap (B \setminus A) = \emptyset$.

DEFINITION 8.23. For every set A we define the set $\mathcal{P}(A)$ as the set whose elements are the subsets of A ; we call $\mathcal{P}(A)$ the power set of A ; $\mathcal{P}(A)$ is a witness for (a theorem obtained from) the power set axiom.

EXERCISE 8.24. Explain in detail the definition of $\mathcal{P}(A)$.

EXAMPLE 8.25. 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 8.26. Let $A = \{a, b, c, d\}$. Write down the set $\mathcal{P}(A)$.

EXERCISE 8.27. Let $A = \{a, b\}$. Write down the set $\mathcal{P}(\mathcal{P}(A))$.

DEFINITION 8.28. (Ordered pairs) Let A and B be sets and let $a \in A$, $b \in B$. The ordered pair (a, b) is the set $\{\{a\}, \{a, b\}\}$. We sometimes say “pair” instead of “ordered pair.” Note that $(a, b) \in \mathcal{P}(\mathcal{P}(A \cup B))$.

Note that if $a = b$ then $(a, b) = \{\{a\}\}$.

Imitating the definition above one can clearly define a ternary predicate “ z is an ordered pair with first element x and second element y ” which we denote by $ordpair(x, y, z)$.

DEFINITION 8.29. For every 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,

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

PROPOSITION 8.30. $(a, b) = (c, d)$ 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. \square

EXERCISE 8.31. If $A = \{a, b, c\}$ and $B = \{c, d\}$ with a, b, c, d distinct then

$$A \times B = \{(a, c), (a, d), (b, c), (b, d), (c, c), (c, d)\}$$

with all listed pairs are distinct.

EXERCISE 8.32. Prove that

- 1) $(A \cap B) \times C = (A \times C) \cap (B \times C)$,
- 2) $(A \cup B) \times C = (A \times C) \cup (B \times C)$.

EXERCISE 8.33. 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. To derive a contradiction consider the set

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

There are two cases. First case is $S \notin S$. Since $S \in T$, we get that $S \in S$, a contradiction. The second case is $S \in S$. Since $S \in T$, we get that $S \notin S$, which is again a contradiction.

REMARK 8.34. 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 clear.

CHAPTER 9

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:

DEFINITION 9.1. A map (or function) from a set A to a set B is a subset $F \subset A \times B$ such that for every $a \in A$ there is a unique $b \in B$ with $(a, b) \in F$. If $(a, b) \in F$ we write $F(a) = b$ or $a \mapsto b$ or $a \mapsto F(a)$. We also write $F : A \rightarrow B$ or $A \xrightarrow{F} B$.

REMARK 9.2. The above defines a new (ternary) predicate *fun* which in Argot reads “...is a map from ... to ...” So *fun* is introduced by the definition

$$\forall x \forall y \forall z (fun(x, y, z) \leftrightarrow ((x \subset y \times z) \wedge (\forall u ((u \in y) \rightarrow \exists v ((u, v) \in x) \wedge \dots))))$$

Also for each F, A, B we may introduce a new functional symbol \widehat{F} by

$$\forall x \forall y ((fun(F, A, B) \wedge (x \in A) \wedge (y \in B)) \rightarrow ((\widehat{F}(x) = y) \leftrightarrow ((x, y) \in F)))$$

We will usually drop the $\widehat{}$ (or think of the Argot translation as dropping the hat). Also note that what we call a map $F \subset A \times B$ corresponds to what in elementary mathematics is called the graph of a map.

EXAMPLE 9.3. If a, b, c are distinct the set

$$(9.1) \quad F = \{(a, a), (b, c)\} \subset \{a, b\} \times \{a, b, c\}$$

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 9.4. For every A the identity map $I : A \rightarrow A$ is defined as $I(a) = a$, i.e.,

$$I = I_A = \{(a, a); a \in A\} \subset A \times A.$$

DEFINITION 9.5. A map $F : A \rightarrow B$ is injective (or an injection, or one-to-one) if $F(a) = F(c)$ implies $a = c$.

DEFINITION 9.6. A map $F : A \rightarrow B$ is surjective (or a surjection, or onto) if for every $b \in B$ there exists an $a \in A$ such that $F(a) = b$.

EXAMPLE 9.7. The map (9.1) is injective and not surjective.

EXERCISE 9.8. Give an example of a map which is surjective and not injective.

EXERCISE 9.9. Let $A \subset B$. Prove that there is an injective map $i : A \rightarrow B$ such that $i(a) = a$ for all $a \in A$. We call i the inclusion map; we sometimes say $A \subset B$ is the inclusion map.

EXERCISE 9.10. (Composition) Prove that if $F : A \rightarrow B$ and $G : B \rightarrow C$ are two maps then there exists a unique map $H : A \rightarrow C$ such that $H(a) = G(F(a))$ for all a . We write $H = G \circ F$ and call the latter the composition of G with F . Hint: We let $(a, c) \in H$ if and only if there exists $b \in B$ with $(a, b) \in F$, $(b, c) \in G$.

DEFINITION 9.11. (Restriction) If $F : A \rightarrow B$ 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 $F|_{A'} : A' \rightarrow B$.

DEFINITION 9.12. (Commutative diagram) By a commutative diagram of sets

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

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

EXERCISE 9.13. 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 9.14. Prove that the composition of two injective maps is injective and the composition of two surjective maps is surjective.

DEFINITION 9.15. A map is bijective (or a bijection) if it is injective and surjective.

Here is a fundamental theorem in Set Theory:

THEOREM 9.16. (*Bernstein's Theorem*) If A and B are sets and if there exist injective maps $F : A \rightarrow B$ and $G : B \rightarrow A$ then there exists a bijective map $U : A \rightarrow B$.

A proof will be given in the chapter on Sequences.

EXERCISE 9.17. Prove that if $F : A \rightarrow B$ is bijective then there exists a unique bijective map denoted by $F^{-1} : B \rightarrow A$ such that $F \circ F^{-1} = I_B$ and $F^{-1} \circ F = I_A$. F^{-1} is called the inverse of F .

EXERCISE 9.18. Let a, b, c, d, e be distinct. Let $F : \{a, b, c\} \rightarrow \{c, d, e\}$, $F(a) = d$, $F(b) = c$, $F(c) = e$. Prove that F has an inverse and compute F^{-1} .

EXERCISE 9.19. Prove that if A and B are sets then there exist maps $F : A \times B \rightarrow A$ and $G : A \times B \rightarrow B$ 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 9.20. Prove that $(A \times B) \times C \rightarrow A \times (B \times C)$, $((a, b), c) \mapsto (a, (b, c))$ is a bijection.

DEFINITION 9.21. 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 9.22. *If A is a set then there is no bijection between A and $\mathcal{P}(A)$*

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

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

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 9.23. Note the similarity between the above argument and the argument showing that there is no set having all sets as elements (the ‘‘Russell paradox’’).

DEFINITION 9.24. Let S be a set of sets and I a set. A family of sets in S indexed by I is a map $I \rightarrow S$, $i \mapsto A_i$. 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 a set (denoted by $\bigcap_{i \in I} A_i$, called their intersection) 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 $I \rightarrow \bigcup_{i \in I} A_i$, $i \mapsto a_i$, 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 9.25. 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 9.26. Let $F : A \rightarrow B$ be a map and $X \subset A$. Define the image of X as the set

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

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

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

For $y \in B$ define

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

(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 9.27. Let $F : \{a, b, c, d, e, f, g\} \rightarrow \{c, d, e, h\}$, $F(a) = d$, $F(b) = c$, $F(c) = e$, $F(d) = c$, $F(e) = d$, $F(f) = c$, $F(g) = c$. 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 9.28. Prove that if $F : A \rightarrow B$ is a map and $X \subset X' \subset A$ are subsets then $F(X) \subset F(X')$.

EXERCISE 9.29. Prove that if $F : A \rightarrow B$ is a map and $(X_i)_{i \in I}$ is a family of subsets of A then

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

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

EXERCISE 9.31. Prove that if $F : A \rightarrow B$ is a map and $(Y_i)_{i \in I}$ is a family of subsets of B then

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

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

DEFINITION 9.32. If A and B are sets we denote by $B^A \subset \mathcal{P}(A \times B)$ the set of all maps $F : A \rightarrow B$; sometimes one writes $\text{Map}(A, B) = B^A$.

EXERCISE 9.33. Let $0, 1$ be two elements. Prove that the map $\{0, 1\}^A \rightarrow \mathcal{P}(A)$ sending $F : A \rightarrow \{0, 1\}$ into $F^{-1}(1) \in \mathcal{P}(A)$ is a bijection.

EXERCISE 9.34. Find a bijection between $(C^B)^A$ and $C^{A \times B}$. Hint: Send $F \in (C^B)^A$, $F : A \rightarrow C^B$, into the set (map)

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

CHAPTER 10

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 10.1. If A is a set then a relation on A is a subset $R \subset A \times A$. If $(a, b) \in R$ we write aRb .

REMARK 10.2. Exactly as in Remark 9.2 the above defines a new (binary) predicate "... is a relation on ..." and we may introduce a corresponding new predicate (still denoted by R).

DEFINITION 10.3. A relation R is called an order if (writing $a \leq b$ instead of aRb) we have, for all $a, b, c \in A$, that

- 1) $a \leq a$ (reflexivity),
- 2) $a \leq b$ and $b \leq c$ imply $a \leq c$ (transitivity),
- 3) $a \leq b$ and $b \leq a$ imply $a = b$ (antisymmetry).

DEFINITION 10.4. One writes $a < b$ if $a \leq b$ and $a \neq b$.

DEFINITION 10.5. An order relation is called a total order if for every $a, b \in A$ either $a \leq b$ or $b \leq a$. Alternatively we say A is totally ordered (by \leq).

EXAMPLE 10.6. 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 10.7. 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 10.8. 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 10.9. 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 10.10. An ordered set is a pair (A, \leq) where A is a set and \leq is an order relation on A .

DEFINITION 10.11. Let (A, \leq) and (A', \leq') be ordered sets. A map $F : A \rightarrow A'$ is called increasing if for every $a, b \in A$ with $a \leq b$ we have $F(a) \leq' F(b)$.

EXERCISE 10.12. Prove that if (A, \leq) , (A', \leq') , (A'', \leq'') are ordered sets and $G : A \rightarrow A'$, $F : A' \rightarrow A''$ are increasing then $F \circ G : A \rightarrow A''$ is increasing.

DEFINITION 10.13. Let A be a set with an order \leq . We say $\alpha \in A$ is a minimal element of A if for all $a \in A$ such that $a \leq \alpha$ we must have $a = \alpha$.

DEFINITION 10.14. Let A be a set with an order \leq . We say $\beta \in A$ is a maximal element of A if for all $b \in A$ such that $\beta \leq b$ we must have $\beta = b$.

DEFINITION 10.15. Let A be a set with an order \leq and let $B \subset A$. We say $m \in B$ is a minimum element of B 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 $\min B$. Note that if $\min B$ exists then, by definition, $\min B$ belongs to B .

DEFINITION 10.16. Let A be a set with an order \leq and $B \subset A$. We say $M \in B$ is a maximum element of B if for all $b \in B$ we have $b \leq M$. If a maximum element exists it is unique and we denote it by $\max B$. Again, if $\max B$ exists then by definition it belongs to B .

DEFINITION 10.17. Let A be a set with an order \leq and let $B \subset A$. An element $u \in A$ is called an upper bound for B 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 if $l \leq b$ for all $b \in B$; we also say B is bounded from below by l . Let $U(B)$ be the set of upper bounds of B ; if $U(B)$ has a minimum element we call it the supremum of B and we denote it by $\sup B$. Let $L(B)$ be the set of lower bounds of B ; if $L(B)$ has a maximum element we call it the infimum of B and we denote it by $\inf B$. (Note that if one of $\sup B$ and $\inf B$ exists that element is by definition in A but does not necessarily belong to B .) We say B is bounded if it has both an upper bound and a lower bound.

EXERCISE 10.18. Consider the set A and the order \leq defined by the relation R in Exercise 10.8. 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 10.19. 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 10.20. A well ordered set is an ordered set (A, \leq) such that every non-empty subset $B \subset A$ has a minimum element.

EXERCISE 10.21. Prove that every well ordered set is totally ordered.

REMARK 10.22. Later, when we will have introduced the ordered set of integers and the ordered set of rational numbers we will see that 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 10.23. (*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 10.24. (*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 10.25. 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 10.26. 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 10.27. A relation R is called an equivalence relation if (writing $a \sim b$ instead of aRb) we have, for all $a, b, c \in A$, that

- 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 10.28. 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 10.29. Given an equivalence relation \sim as above for every $a \in A$ we may consider the set

$$\hat{a} = \{c \in A; c \sim a\}$$

called the equivalence class of a .

DEFINITION 10.30. Sometimes, instead of \hat{a} , one writes \bar{a} or $[a]$.

EXERCISE 10.31. Prove that $\hat{a} = \hat{b}$ if and only if $a \sim b$.

EXERCISE 10.32. 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 10.33. 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 10.34. 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 10.35. Let A be a set and $(A_i)_{i \in I}$ a partition of A . Define a relation R on A as follows:

$$R = \{(a, b) \in A \times A; \exists i((i \in I) \wedge (a \in A_i) \wedge (b \in A_i))\}.$$

Prove that R is an equivalence relation.

EXERCISE 10.36. 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 10.37. The set of equivalence classes

$$\{\alpha \in \mathcal{P}(A); \exists a((a \in A) \wedge (\alpha = \widehat{a}))\}$$

is denoted by A/\sim and is called the quotient of A by the relation \sim .

EXAMPLE 10.38. 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 10.39. 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 10.40. Let S be a set. For every sets $X, Y \in S$ write $X \sim Y$ if and only if there exists a bijection $F: X \rightarrow Y$. This defines a relation on S . Prove that this is an equivalence relation.

EXERCISE 10.41. 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 10.42. 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 10.43. 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 10.44. Prove that every two distinct non-parallel lines intersect in exactly one point.

EXERCISE 10.45. 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 10.46. 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!)

DEFINITION 10.47. A projective plane is a pair $(\bar{A}, \bar{\mathcal{L}})$ where \bar{A} is a set and $\bar{\mathcal{L}} \subset \mathcal{P}(\bar{A})$ is a set of subsets of \bar{A} satisfying a series of axioms which we now explain. Again it is convenient to introduce some terminology as follows. \bar{A} is called the projective plane. The elements of \bar{A} are called points, P . The elements \bar{L} of $\bar{\mathcal{L}}$ are called lines; so each such $\bar{L} \subset \bar{A}$. We say a point P lies on a line \bar{L} if $P \in \bar{L}$; we also say that \bar{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 10.48. One can attach to every affine plane (A, \mathcal{L}) a projective plane $(\bar{A}, \bar{\mathcal{L}})$ 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 (check!). Denote by \hat{L} the equivalence class of L . Then we consider the set of equivalence classes, $\bar{L}_\infty = \mathcal{L} / \parallel$; call this set the line at infinity. There exists a set \bar{A} such that $\bar{A} = A \cup \bar{L}_\infty$ and $A \cap \bar{L}_\infty = \emptyset$. Define a line in \bar{A} to be either \bar{L}_∞ or set of the form $\bar{L} = L \cup \{\hat{L}\}$. Finally define $\bar{\mathcal{L}}$ to be the set of all lines in \bar{A} .

EXERCISE 10.49. Explain why \bar{A} exists. Check that $(\bar{A}, \bar{\mathcal{L}})$ is a projective plane.

EXERCISE 10.50. Describe the projective plane attached to the affine plane in Exercise 10.45; how many points does it have? How many lines?

CHAPTER 11

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 11.1. A binary operation \star on a set A is a map $\star : A \times A \rightarrow A$, $(a, b) \mapsto \star(a, b)$. 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 ,

REMARK 11.2. Exactly as in Remark 9.2 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:

$$\forall x \forall y (\text{binop}(x, y) \leftrightarrow (\text{fun}(x, y \times y)))$$

DEFINITION 11.3. A unary operation $'$ on a set A is a map $' : A \rightarrow A$, $a \mapsto '(a)$. We usually write a' or $'a$ instead of $'(a)$. Instead of $'$ we sometimes use notation like $-$, i ,

EXAMPLE 11.4. 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:

$$\begin{aligned} 0 \wedge 0 &= 0, & 0 \wedge 1 &= 0, & 1 \wedge 0 &= 0, & 1 \wedge 1 &= 1; \\ 0 \vee 0 &= 0, & 0 \vee 1 &= 1, & 1 \vee 0 &= 1, & 1 \vee 1 &= 1; \\ 0 + 0 &= 0, & 0 + 1 &= 1, & 1 + 0 &= 1, & 1 + 1 &= 0. \end{aligned}$$

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

$$\neg 1 = 0, \quad \neg 0 = 1.$$

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 11.5. Compute $((0 \wedge 1) \vee 1) + (1 \wedge (0 \vee (1 + 1)))$.

DEFINITION 11.6. 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 11.7. 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

$$\forall x \forall y \forall z \forall u \forall v \forall w (ring(x, y, z, u, v, w) \leftrightarrow (binop(y, x) \wedge \dots))$$

DEFINITION 11.8. We write $a + b + c$ instead of $(a + b) + c$ and abc for $(ab)c$. We write $a - b$ instead of $a + (-b)$.

DEFINITION 11.9. An element a of a ring R is invertible 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} . A ring R is called a field if $0 \neq 1$ and every non-zero element is invertible.

DEFINITION 11.10. A Boolean ring is a commutative unital ring such that $1 \neq 0$ and for all $a \in A$ we have $a^2 = a$.

EXERCISE 11.11. Prove that in a Boolean ring A we have $a + a = 0$ for all $a \in A$.

EXERCISE 11.12. 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).

EXERCISE 11.13. Prove that if a Boolean ring A is a field then $A = \{0, 1\}$.

DEFINITION 11.14. Let A be a set and let $S = \mathcal{P}(A)$ be the power set of A . Define the following operations on S :

$$\begin{aligned} X \wedge Y &= X \cap Y \\ X \vee Y &= X \cup Y \\ X \Delta Y &= (X \cup Y) \setminus (X \cap Y) \\ \neg X &= \mathcal{C}X = A \setminus X. \end{aligned}$$

EXERCISE 11.15. 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).

Hint: For every $a \in A$ one can define a map $\psi_a : \mathcal{P}(A) \rightarrow \{0, 1\}$ by setting $\psi_a(X) = 1$ if and only if $a \in X$. Note that

- 1) $\psi_a(X \wedge Y) = \psi_a(X) \wedge \psi_a(Y)$,
- 2) $\psi_a(X \vee Y) = \psi_a(X) \vee \psi_a(Y)$,
- 3) $\psi_a(X \Delta Y) = \psi_a(X) + \psi_a(Y)$,
- 4) $\psi_a(\neg X) = \neg \psi_a(X)$.

Next note that $X = Y$ if and only if $\psi_a(X) = \psi_a(Y)$ for all $a \in A$. Use these functions to reduce the present exercise to Exercise 11.12.

DEFINITION 11.16. Given a subset $X \subset A$ one can define the characteristic function $\chi_X : A \rightarrow \{0, 1\}$ by letting $\chi_X(a) = 1$ if and only if $a \in X$; in other words $\chi_X(a) = \psi_a(X)$.

EXERCISE 11.17. Prove that

- 1) $\chi_{X \vee Y}(a) = \chi_X(a) \vee \chi_Y(a)$,
- 2) $\chi_{X \wedge Y}(a) = \chi_X(a) \wedge \chi_Y(a)$,
- 3) $\chi_{X \Delta Y}(a) = \chi_X(a) + \chi_Y(a)$,
- 4) $\chi_{\neg X}(a) = \neg \chi_X(a)$.

DEFINITION 11.18. An algebraic structure is a tuple $(A, \star, \bullet, \dots, \neg, -, \dots, 0, 1, \dots)$ where A is a set, \star, \bullet, \dots are binary operations, $\neg, -, \dots$ are unary operations, and $0, 1, \dots$ 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

$$(A, \star, \bullet, \dots, \neg, -, \dots, 0, 1, \dots) \quad \text{and} \quad (A', \star', \bullet', \dots, \neg', -', \dots, 0', 1', \dots)$$

(with the same number of corresponding operations). A map $F : A \rightarrow A'$ 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 11.19. A map $F : A \rightarrow A'$ 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 11.20. Prove that if $F : A \rightarrow A'$ is a homomorphism of algebraic structures and F is bijective then its inverse $F^{-1} : A' \rightarrow A$ is a homomorphism. Such an F will be called an isomorphism.

DEFINITION 11.21. A subset $\mathcal{A} \subset \mathcal{P}(A)$ is called a Boolean algebra of sets if the following hold:

- 1) $\emptyset \in \mathcal{A}, A \in \mathcal{A}$;
- 2) If $X, Y \in \mathcal{A}$ then $X \cap Y \in \mathcal{A}, X \cup Y \in \mathcal{A}, \mathcal{C}X \in \mathcal{A}$.

(Hence $(\mathcal{A}, \vee, \wedge, \mathcal{C}, \emptyset, A)$ is a Boolean algebra.)

EXERCISE 11.22. Prove that if \mathcal{A} is a Boolean algebra of sets then for every $X, Y \in \mathcal{A}$ we have $X \Delta Y \in \mathcal{A}$. Prove that $(\mathcal{A}, \Delta, \cap, I, \emptyset, A)$ is a Boolean ring.

DEFINITION 11.23. A subset $\mathcal{B} \subset \mathcal{P}(A)$ is called a Boolean ring of sets if the following properties hold:

- 1) $\emptyset \in \mathcal{B}, A \in \mathcal{B}$;
- 2) If $X, Y \in \mathcal{B}$ then $X \cap Y \in \mathcal{B}, X \Delta Y \in \mathcal{B}$.

(Hence $(\mathcal{A}, \Delta, \vee, I, \emptyset, A)$ is a Boolean ring.)

EXERCISE 11.24. Prove that every Boolean ring of sets \mathcal{B} is a Boolean algebra of sets.

DEFINITION 11.25. 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 11.26. 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 11.27. We cannot give examples yet of ordered rings. Later we will see that the rings of integers, rationals, and reals have natural structures of ordered rings.

DEFINITION 11.28. Let R be an ordered ring and let $R_+ = \{a \in R; a \geq 0\}$. A finite measure space is a triple (A, \mathcal{A}, μ) where A is a set, $\mathcal{A} \subset \mathcal{P}(A)$ is a Boolean algebra of sets, and $\mu : \mathcal{A} \rightarrow R_+$ is a map satisfying the property that for every $X, Y \in \mathcal{A}$ with $X \cap Y = \emptyset$ we have

$$\mu(X \cup Y) = \mu(X) + \mu(Y).$$

If in addition $\mu(A) = 1$ we say (A, \mathcal{A}, μ) is a finite probability measure. We say that $X, Y \in \mathcal{A}$ are independent if $\mu(X \cap Y) = \mu(X) \cdot \mu(Y)$.

EXERCISE 11.29. Prove that in a finite measure space $\mu(\emptyset) = 0$ and for every $X, Y \in \mathcal{A}$ we have

$$\mu(X \cup Y) = \mu(X) + \mu(Y) - \mu(X \cap Y).$$

EXERCISE 11.30. 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 11.31. 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 11.32. Let X be a set and $(R, +, \cdot, -, 0, 1)$ a commutative unital ring. Let R^X be the set of all functions $X \rightarrow R$. For $F, G \in R^X$ we define $F + G, F \cdot G, -F, 0, 1 \in R^X$ 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

$$(R^X, +, \cdot, -, 0, 1)$$

is a commutative unital ring.

EXERCISE 11.33. 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 $1 + (-1) = 0$

CHAPTER 12

Integers

In this Chapter we introduce the ring \mathbb{Z} of integers and we prove some easy theorems about this concept.

DEFINITION 12.1. A well ordered ring is an ordered ring $(R, +, -, \times, 0, 1, \leq)$ with $1 \neq 0$ having the property that every non-empty subset of R which is bounded from below has a minimum element.

As usual the above defines a 7-ary predicate *ordring*.

REMARK 12.2. If $(R, +, -, \times, 0, 1, \leq)$ is a well ordered ring then (R, \leq) is not a priori a well ordered set. But if $R_{>0} = \{a \in R; a > 0\}$ then $(R_{>0}, \leq)$ is a well ordered set.

We have the following remarkable theorem in Set Theory T_{set} :

THEOREM 12.3. *There exists a well ordered ring for which $1 \neq 0$.*

REMARK 12.4. The above theorem is formulated, as usual, in Argot; but it should be understood as being a sentence Z in L_{set} of the form

$$\exists x \exists y \dots (\text{ordring}(x, y, \dots))$$

REMARK 12.5. We are going to sketch the proof of Theorem 12.3 in an exercise below. The proof is involved. A cheap way to avoid the proof of this theorem is as follows: add this theorem to the ZFC axioms and let ZFC' be the resulting enriched system of axioms. Then replace T_{set} by the theory T'_{set} with axioms ZFC'. This is what all working mathematicians essentially do anyway.

DEFINITION 12.6. Let $(\mathbb{Z}, +, \times, 0, 1, \leq)$ be a well ordered ring; we call \mathbb{Z} the ring of integers. In particular the conditions in the definition of rings (associativity, commutativity, etc.) and order (transitivity, etc.) become theorems for \mathbb{Z} . We also set $\mathbb{N} = \{a \in \mathbb{Z}; a > 0\}$ and we call \mathbb{N} the set of natural numbers. Later we will prove the "essential uniqueness" of \mathbb{Z} .

The above definition should be read as saying that $(\mathbb{Z}, +, \times, 0, 1, \leq)$ is an existential witness for the sentence Z .

REMARK 12.7. The only predicate in the language L_{set} of sets is \in and 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 12.8. Prove that $1 \in \mathbb{N}$. Hint: Use $(-1) \times (-1) = 1$.

EXERCISE 12.9. Prove that if $a, b \in \mathbb{Z}$ and $ab = 0$ then either $a = 0$ or $b = 0$. Give an example of a ring where this is not true. (For the latter consider the Boolean ring $\mathcal{P}(A)$.)

EXERCISE 12.10. Prove that if $a \in \mathbb{Z}$ then the set $\{x \in \mathbb{Z}; a - 1 < x < a\}$ is empty. Hint: It is enough to show that $S = \{x \in \mathbb{Z}; 0 < x < 1\}$ is empty. Assume S is non-empty and let $m = \min S$. Then $0 < m^2 < m$, hence $0 < m^2 < 1$ and $m^2 < m$, a contradiction.

EXERCISE 12.11. 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.

In what follows we sketch the main idea behind the proof of Theorem 12.3. We begin with the following:

DEFINITION 12.12. A Peano triple is a triple $(N, 1, \sigma)$ where N is a set, $1 \in N$, and $\sigma : N \rightarrow N$ is a map such that

- 1) σ is injective;
 - 2) $\sigma(N) = N \setminus \{1\}$;
 - 3) for every subset $S \subset N$ if $1 \in S$ and $\sigma(S) \subset S$ then $S = N$.
- The conditions 1, 2, 3 are called ‘‘Peano’s axioms.’’

REMARK 12.13. Given a well ordered ring one can easily prove there exists a Peano triple; cf. Exercise 12.14 below. Conversely given a Peano triple one can prove there exists a well ordered ring; this is tedious and will be addressed in Exercise 12.15. The idea of proof of Theorem 12.3 is to prove in Set Theory (i.e., ZFC) that there exists a Peano triple; here the axiom of infinity is crucial. Then the existence of a well ordered ring follows.

EXERCISE 12.14. Prove that if Z is a well ordered ring, $N = \{x \in Z; x > 0\}$, and $\sigma : N \rightarrow N$ is $\sigma(x) = x + 1$ then $(N, 1, \sigma)$ is a Peano triple.

EXERCISE 12.15. This exercise gives some steps towards showing how to construct a well ordered ring from a given Peano triple. Assume $(N, 1, \sigma)$ is a Peano triple. For $y \in N$ let

$$A_y = \{\tau \in N^N; \tau(1) = \sigma(y), \forall x(\tau(\sigma(x)) = \sigma(\tau(x)))\}.$$

1) Prove that A_y has at most one element. Hint: If $\tau, \eta \in A_y$ and $S = \{x; \tau(x) = \eta(x)\}$ then $1 \in S$ and $\sigma(S) \subset S$; so $S = N$.

2) Prove that for every y , $A_y \neq \emptyset$. Hint: If $T = \{y \in N; A_y \neq \emptyset\}$ 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) Prove that $x + y = y + x$ and $(x + y) + z = x + (y + z)$ on N .

5) Prove 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 $N' = \{-\} \times N$, $Z = N' \cup \{0\} \cup N$ where 0 and $-$ are two sets. Naturally extend $+$ to Z .

7) Define \times on N and then on Z in the same style as for $+$.

8) Define \leq on N and prove (N, \leq) is well ordered. Extend this to Z .

9) Prove that $(Z, +, \times, 0, 1, \leq)$ is a well ordered ring.

From now on we accept Theorem 12.3 (either as a theorem whose proof we summarily sketched or as an additional axiom for Set Theory).

DEFINITION 12.16. Define the natural numbers 2, 3, ..., 9 by

$$\begin{aligned} 2 &= 1 + 1 \\ 3 &= 2 + 1 \\ &\dots \\ 9 &= 8 + 1. \end{aligned}$$

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 12.17. Prove that $12 = 9 + 3$. Hint: We have:

$$\begin{aligned} 12 &= 10 + 2 \\ &= 2 \times 5 + 2 \\ &= (1 + 1) \times 5 + 2 \\ &= 1 \times 5 + 1 \times 5 + 2 = 5 + 5 + 2 \\ &= 5 + 5 + 1 + 1 = 5 + 6 + 1 = 5 + 7 = 4 + 1 + 7 \\ &= 4 + 8 = 3 + 1 + 8 = 3 + 9 = 9 + 3. \end{aligned}$$

EXERCISE 12.18. Prove that $18 + 17 = 35$. Prove that $17 \times 3 = 51$.

REMARK 12.19. 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 12.20. Prove that $7 \leq 20$.

DEFINITION 12.21. For every integers $a, b \in \mathbb{Z}$ the set $\{x \in \mathbb{Z}; a \leq x \leq b\}$ will be denoted, for simplicity, by $\{a, \dots, b\}$. 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 $\{0, 1, \dots, n\}$ means $\{0, \dots, n\}$ whereas $\{2, 4, 6, \dots, 2n\}$ will mean $\{2x; 1 \leq x \leq n\}$, etc. A similar convention applies if there are no numbers after the dots.

EXAMPLE 12.22. $\{-2, \dots, 11\} = \{-2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$.

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 12.23. Prove that \mathbb{N} is not bounded.

EXERCISE 12.24. 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 $\{b - x; x \in A\}$.

DEFINITION 12.25. 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 12.26. Prove that if a is odd then $a - 1$ is even. Hint: Consider the set $\{b \in \mathbb{N}; 2b \geq a\}$, 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 12.27. 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 12.28. 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 $f : \mathbb{N} \rightarrow \mathbb{Z}$. Define $f(x) = x$. Then f is not surjective so it is not a bijection.

EXERCISE 12.29. Prove that there is a bijection between \mathbb{N} and \mathbb{Z} .

CHAPTER 13

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.) Let $P(x)$ be a formula in the language L_{set} of sets, with free variable x . For each such $P(x)$ we have:

PROPOSITION 13.1. (*Induction Principle*) Assume

1) $P(1)$.

2) For all $n \in \mathbb{N}$ if $n \neq 1$ and $P(n - 1)$ then $P(n)$.

Then for all $n \in \mathbb{N}$ we have $P(n)$.

The above is expressed, as usual, in Argot. The same expressed as a sentence in L_{set} reads:

$$(P(1) \wedge (\forall x((x \in \mathbb{N}) \wedge (x \neq 0) \wedge P(x - 1)) \rightarrow P(x))) \rightarrow (\forall x((x \in \mathbb{N}) \rightarrow P(x))).$$

We refer to the above as *induction on n* . For each explicit $P(x)$ this is a genuine theorem. Note that the above Proposition does not say “for all P something happens”; that would not be a sentence in the language of sets.

Proof. Let $S = \{n \in \mathbb{N}; \neg P(n)\}$. We want to show that $S = \emptyset$. Assume $S \neq \emptyset$ and seek a contradiction. Let m be the minimum of S . By 1) $m \neq 1$. By Exercise 12.11 $m - 1 \in \mathbb{N}$. By minimality of m , we have $P(m - 1)$. By 2) we get $P(m)$, a contradiction. \square

EXERCISE 13.2. Define $n^2 = n \times n$ and $n^3 = n^2 \times n$ for every integer n . Prove that for every natural n there exists an integer m such that $n^3 - n = 3m$. (Later we will say that 3 divides $n^3 - n$.) Hint: Proceed by induction on n as follows. Let $P(n)$ be the sentence: for all natural n there exists an integer m such that $n^3 - n = 3m$. $P(1)$ is true because $1^3 - 1 = 3 \times 0$. Assume now that $P(n - 1)$ is true i.e., $(n - 1)^3 - (n - 1) = 3q$ for some integer q and let us check that $P(n)$ is true i.e., that $n^3 - n = 3m$ for some integer m . The equality $(n - 1)^3 - (n - 1) = 3q$ reads $n^3 - 3n^2 + 3n - 1 - n + 1 = 3q$. Hence $n^3 - n = 3(n^2 - n)$ and we are done.

EXERCISE 13.3. Define $n^5 = n^3 \times n^2$. Prove that for every natural n there exists an integer m such that $n^5 - n = 5m$.

PROPOSITION 13.4. *If there exists a bijection $\{1, \dots, n\} \rightarrow \{1, \dots, m\}$ then $n = m$.*

Proof. We proceed by induction on n . Let $P(n)$ be the statement of the Proposition. 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 consider a bijection $F : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$; we want to prove that $n = m$. Let $i = F(n)$ and define the map $G : \{1, \dots, n - 1\} \rightarrow \{1, \dots, m\} \setminus \{i\}$ by $G(j) = F(j)$ for all $1 \leq j \leq n - 1$. Then

clearly G is a bijection. Now consider the map $H : \{1, \dots, m\} \setminus \{i\} \rightarrow \{1, \dots, m-1\}$ defined by $H(j) = j$ for $1 \leq j \leq i-1$ and $H(j) = j-1$ for $i+1 \leq j \leq m$. (The definition is correct because for every $j \in \{1, \dots, m\} \setminus \{i\}$ either $j \leq i-1$ or $j \geq i+1$; cf. Exercise 12.10.) Clearly H is a bijection. We get a bijection

$$H \circ G : \{1, \dots, n-1\} \rightarrow \{1, \dots, m-1\}.$$

Since $P(n-1)$ is true we get $n-1 = m-1$. Hence $n = m$ and we are done. \square

EXERCISE 13.5. Check that $P(1)$ is true in the above Proposition.

REMARK 13.6. 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 (called 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 13.7. Here is a wrong argument for the induction step in the proof of Proposition 13.4.

“Proof.” Let $G : \{1, \dots, n-1\} \rightarrow \{1, \dots, m-1\}$ be any bijection and let $F : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ 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 covers bijections $F : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ constructed from bijections $G : \{1, \dots, n-1\} \rightarrow \{1, \dots, m-1\}$ in the special way described above. In other words an arbitrary bijection $F : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ 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 12.28): 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 13.8. A set A is finite if there exists an integer $n \geq 0$ and a bijection $F : \{1, \dots, n\} \rightarrow A$. (Note that n is then unique by Proposition 13.4.) 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 $A = \{a_1, \dots, a_n\}$. A set is infinite if it is not finite.

EXERCISE 13.9. Prove that $|\{2, 4, -6, 9, -100\}| = 5$.

EXERCISE 13.10. 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: Reduce to the case $A \cap B = \emptyset$. Then if $F : \{1, \dots, a\} \rightarrow A$ and $G : \{1, \dots, b\} \rightarrow B$ are bijections prove that $H : \{1, \dots, a + b\} \rightarrow A \cup B$ 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.

EXERCISE 13.11. For every finite sets A and B we have that $A \times B$ is finite and

$$|A \times B| = |A| \times |B|.$$

Hint: Induction on $|A|$.

EXERCISE 13.12. Let $F : \{1, \dots, n\} \rightarrow R$ 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 $G : \{1, \dots, n\} \rightarrow R$ 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 13.13. 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 $a_1 + \dots + a_n$ for this sum. If $a_1 = \dots = a_n = a$ the sum $a_1 + \dots + a_n$ is written as $a + \dots + a$ (n times).

EXERCISE 13.14. Prove that for every $a, b \in \mathbb{N}$ we have

$$a \times b = a + \dots + a \text{ (} b \text{ times)} = b + \dots + b \text{ (} a \text{ times)}.$$

EXERCISE 13.15. Define in a similar way the (finite) product $\prod_{i=1}^n a_i$ (which is also denoted by $a_1 \dots a_n = a_1 \times \dots \times a_n$). Prove the analogues of associativity and distributivity for sums and products of families of numbers. Define a^b for $a, b \in \mathbb{N}$ and prove that $a^{b+c} = a^b \times a^c$ and $(a^b)^c = a^{bc}$.

EXERCISE 13.16. Prove that if a is an integer and n is a natural number then

$$a^n - 1 = (a - 1)(a^{n-1} + a^{n-2} + \dots + a + 1).$$

Hint: Induction on n .

EXERCISE 13.17. Prove that if a is an integer and n is an integer then

$$a^{2n+1} + 1 = (a + 1)(a^{2n} - a^{2n-1} + a^{2n-2} - \dots - a + 1).$$

Hint: Set $a = -b$.

EXERCISE 13.18. 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 $F : A \setminus \{b\} \rightarrow \{1, \dots, m\}$ and one constructs a bijection $G : A \rightarrow \{1, \dots, m + 1\}$ 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 $F : \{1, \dots, n\} \rightarrow A$ be a bijection, $a_i = F(i)$. Then $\{a_1, \dots, a_{n-1}\}$ is bounded from above by some b and conclude that A is bounded from above by either b or a_n .

EXERCISE 13.19. Prove that every subset of a finite set is finite. Hint: Use the previous exercise.

DEFINITION 13.20. Let A be a set and $n \in \mathbb{N}$. Define the set A^n to be the set $A^{\{1, \dots, n\}}$ of all maps $\{1, \dots, n\} \rightarrow A$. Call

$$A^* = \bigcup_{n=1} A^n$$

the set of words with letters in A .

DEFINITION 13.21. If $f : \{1, \dots, n\} \rightarrow A$ and $f(i) = a_i$ we write f as a “tuple” (a_1, \dots, a_n) and sometimes as a “word” $a_1 \dots a_n$; in other words we add to the definitions of Set Theory the following definitions

$$f = (a_1, \dots, a_n) = a_1 \dots a_n.$$

EXERCISE 13.22. Show that the maps $A^n \times A^m \rightarrow A^{n+m}$,

$$((a_1, \dots, a_n), (b_1, \dots, b_m)) \mapsto (a_1, \dots, a_n, b_1, \dots, b_m)$$

(called concatenations), are bijections. They induce a non-injective binary operation $A^* \times A^* \rightarrow A^*$, $(u, v) \rightarrow uv$. Prove that $u(vw) = (uv)w$.

CHAPTER 14

Rationals

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.

DEFINITION 14.1. For every $a, b \in \mathbb{Z}$ with $b \neq 0$ define the fraction $\frac{a}{b}$ to be the set of all pairs (c, d) with $c, d \in \mathbb{Z}$, $d \neq 0$ such that $ad = bc$. Denote by \mathbb{Q} the set of all fractions. So

$$\frac{a}{b} = \{(c, d) \in \mathbb{Z} \times \mathbb{Z}; d \neq 0, ad = bc\},$$

$$\mathbb{Q} = \left\{ \frac{a}{b}; a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

EXAMPLE 14.2.

$$\frac{6}{10} = \{(6, 10), (-3, -5), (9, 15), \dots\} \in \mathbb{Q}.$$

EXERCISE 14.3. 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 14.4. On the set $A = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ 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 $\mathbb{Q} = A / \sim$ is the quotient of A by the relation \sim .

EXERCISE 14.5. Prove that the map $\mathbb{Z} \rightarrow \mathbb{Q}$, $a \mapsto \frac{a}{1}$ is injective.

DEFINITION 14.6. By abuse we identify $a \in \mathbb{Z}$ with $\frac{a}{1} \in \mathbb{Q}$ and write $\frac{a}{1} = a$; this identifies \mathbb{Z} with a subset of \mathbb{Q} . Such identifications are very common and will be done later in similar contexts.

DEFINITION 14.7. Define $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$.

EXERCISE 14.8. Show that the above definition is correct (i.e., 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 similarly for the product).

EXERCISE 14.9. Prove that \mathbb{Q} (with the operations $+$ and \times defined above and with the elements $0, 1$) is a field.

DEFINITION 14.10. For $\frac{a}{b}, \frac{c}{d}$ 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 14.11. Prove that \mathbb{Q} equipped with \leq is an ordered ring but it is not a well ordered ring.

EXERCISE 14.12. Let A be a non-empty finite set and define $\mu : \mathcal{P}(A) \rightarrow \mathbb{Q}$ by

$$\mu(X) = \frac{|X|}{|A|}.$$

- 1) $(A, \mathcal{P}(A), \mu)$ is a finite probability measure space.
- 2) Prove that if $X = Y \neq A, \emptyset$ then X and Y are not independent.
- 3) Prove that if $X \cap Y = \emptyset$ and $X \neq \emptyset, Y \neq \emptyset$ then X and Y are not independent.
- 4) Prove that if $A = B \times C, X = B' \times C, Y = B \times C', B' \subset B, C' \subset C$, then X and Y are independent.

EXERCISE 14.13. Prove by induction the following equalities:

$$1 + 2 + \dots + n = \frac{n(n+1)}{2}.$$

$$1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

$$1^3 + 2^3 + \dots + n^3 = \frac{n^2(n+1)^2}{4}.$$

CHAPTER 15

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 15.1. For $n \in \mathbb{N}$ define the factorial of n (read n factorial) by

$$n! = 1 \times 2 \times \dots \times n \in \mathbb{N}.$$

Also set $0! = 1$.

DEFINITION 15.2. For $0 \leq k \leq n$ in \mathbb{N} define the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \in \mathbb{Q}.$$

One also reads this “ n choose k .”

EXERCISE 15.3. Prove that

$$\binom{n}{k} = \binom{n}{n-k}$$

and

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n.$$

EXERCISE 15.4. Prove that

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

Hint: Direct computation with the definition.

EXERCISE 15.5. Prove that

$$\binom{n}{k} \in \mathbb{Z}.$$

Hint: Proceed by induction on n ; use Exercise 15.4.

EXERCISE 15.6. 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 $mc = c + \dots + c$ (m times). Hint: Induction on n and use Exercise 15.4.

EXERCISE 15.7. (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 $A = \{a_1, \dots, a_{n+1}\}$ use

$$\mathcal{P}(A) = \{B \in \mathcal{P}(A); a_{n+1} \in B\} \cup \{B \in \mathcal{P}(A); a_{n+1} \notin B\}.$$

EXERCISE 15.8. (Combinations) Let A be a set with $|A| = n$, let $0 \leq k \leq n$, and set

$$\text{Comb}(k, A) = \{B \in \mathcal{P}(A); |B| = k\}.$$

Prove that

$$|\text{Comb}(k, A)| = \binom{n}{k}.$$

In other words a set of n elements has exactly $\binom{n}{k}$ 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 $A = \{a_1, \dots, a_{n+1}\}$ use Exercise 15.4 plus the fact that $\text{Comb}(k, A)$ can be written as

$$\{B \in \mathcal{P}(A); |B| = k, a_{n+1} \in B\} \cup \{B \in \mathcal{P}(A); |B| = k, a_{n+1} \notin B\}.$$

EXERCISE 15.9. (Permutations) For a set A let $\text{Perm}(A) \subset A^A$ be the set of all bijections $F : A \rightarrow A$. A bijection $F : A \rightarrow A$ is also called a permutation. Prove that if $|A| = n$ then

$$|\text{Perm}(A)| = n!.$$

So the exercise says that a set of n elements has $n!$ permutations. Hint: Let $|A| = |B| = n$ and let $\text{Bij}(A, B)$ be the set of all bijections $F : A \rightarrow B$; it is enough to show that $|\text{Bij}(A, B)| = n!$. Proceed by induction on n ; if $A = \{a_1, \dots, a_{n+1}\}$, $B = \{b_1, \dots, b_{n+1}\}$ then use the fact that

$$\text{Bij}(A, B) = \bigcup_{k=1}^{n+1} \{F \in \text{Bij}(A, B); F(a_1) = b_k\}.$$

For $d \in \mathbb{N}$ and X a set let X^d be the set of all maps $\{1, \dots, d\} \rightarrow X$. We identify a map $i \mapsto a_i$ with the tuple (a_1, \dots, a_d) .

EXERCISE 15.10. (Combinations with repetition) Let

$$\text{Combrep}(n, d) = \{(x_1, \dots, x_d) \in \mathbb{Z}^d; x_i \geq 0, x_1 + \dots + x_d = n\}.$$

Prove that

$$|\text{Combrep}(n, d)| = \binom{n+d-1}{d-1}.$$

Hint: Let $A = \{1, \dots, n+d-1\}$. Prove that there is a bijection

$$\text{Comb}(d-1, A) \rightarrow \text{Combrep}(n, d).$$

The bijection is given by attaching to every subset

$$\{i_1, \dots, i_{d-1}\} \subset \{1, \dots, n+d-1\}$$

(where $i_1 < \dots < i_{d-1}$) the tuple (x_1, \dots, x_{d-1}) where

- 1) $x_1 = |\{i \in \mathbb{Z}; 1 \leq i < i_1\}|$,
- 2) $x_k = |\{i \in \mathbb{Z}; i_k < i < i_{k+1}\}|$, for $2 \leq k \leq d-1$, and
- 3) $x_d = |\{i \in \mathbb{Z}; i_{d-1} < i \leq n+d-1\}|$.

CHAPTER 16

Sequences

DEFINITION 16.1. A sequence in a set A is a map $F : \mathbb{N} \rightarrow A$. If we write $F(n) = a_n$ we also say that a_1, a_2, \dots is a sequence in A or that (a_n) is a sequence in A .

THEOREM 16.2. (*Recursion theorem*) Let A be a set, $a \in A$ an element, and let F_1, F_2, \dots be a sequence of maps $A \rightarrow A$. Then there is a unique map $G : \mathbb{N} \rightarrow A$ 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 unique map

$$G_n : \{1, \dots, n\} \rightarrow A$$

such that for every $k < n$

$$G_n(k+1) = F_k(G_n(k)).$$

Next by uniqueness one gets that for every n and every $k \leq n$ we have

$$G_n(k) = G_{n+1}(k).$$

Finally, seeing G_n as a subset of $\{1, \dots, n\} \times A$ one defines

$$G := \bigcup G_n \subset \mathbb{N} \times A$$

and one proves G is a map with the desired properties. \square

Here are some applications of recursion. First we are ready to give a proof of:

THEOREM 16.3. (*Bernstein's Theorem*) If there exist injections $F : A \rightarrow B$ and $G : B \rightarrow A$ then there exists a bijection $U : A \rightarrow B$

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 $H : A \rightarrow C$ is an injection. We are reduced to proving the Lemma below. \square

LEMMA 16.4. Let A be a set and $C \subset A$ a subset. Assume there is an injection $H : A \rightarrow C$. Then there is a bijection $U : A \rightarrow C$.

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 $U : A \rightarrow C$ by $U(x) = H(x)$ if $x \in A_n \setminus C_n$ for some $n \in \mathbb{N}$ and $U(x) = x$ otherwise. We claim that $U : A \rightarrow C$ is a bijection. Indeed note that we have

$$A_\infty \subset \dots \subset C_n \subset A_n \subset \dots \subset C_2 \subset A_2 \subset C_1 \subset A_1$$

where

$$A_\infty := \bigcap_n A_n = \bigcap_n C_n.$$

Also note that the sets

$$A_1 \setminus C_1, C_1 \setminus A_2, A_2 \setminus C_2, C_2 \setminus A_3, \dots, A_\infty$$

define a partition of A and the sets

$$C_1 \setminus A_2, A_2 \setminus C_2, C_2 \setminus A_3, \dots, A_\infty$$

define a partition of C . We are done by noting that U induces bijections

$$A_1 \setminus C_1 \rightarrow A_2 \setminus C_2, A_2 \setminus C_2 \rightarrow A_3 \setminus C_3, \dots$$

and is the identity on

$$C_1 \setminus A_2, C_2 \setminus A_3, \dots, A_\infty$$

hence U is bijection. □

Here is another application of recursion:

PROPOSITION 16.5. *Let (A, \leq) be an ordered set that has no maximal element. Then there is a sequence $F : \mathbb{N} \rightarrow A$ such that for all $n \in \mathbb{N}$ we have $F(n) < F(n+1)$.*

Proof. Let $B = \{(a, b) \in A \times A; a < b\}$. By hypothesis the first projection $F : B \rightarrow A$, $(a, b) \mapsto a$ is surjective. By the axiom of choice there exists $G : A \rightarrow B$ such that $F \circ G = I_A$. Then $G(a) > a$ for all a . By the recursion theorem there exists $F : \mathbb{N} \rightarrow A$ such that $F(n+1) = G(F(n))$ for all n and we are done. □

EXERCISE 16.6. (Uniqueness of the ring of integers) Let Z and Z' be two well ordered rings with identities 1 and $1'$. Prove that there exists a unique ring homomorphism $F : Z \rightarrow Z'$; prove that this F is bijective and increasing. Hint: Let Z_+ be the set of all elements in Z which are > 0 and similarly for Z' . By recursion (which is, of course, valid in every well ordered ring) there is a unique $F : Z_+ \rightarrow Z'_+$ satisfying $F(1) = 1'$ and $F(n+1) = F(n) + 1'$. Define F on Z by $F(-n) = -F(n)$ for $-n \in Z_+$.

DEFINITION 16.7. A set A is countable if there exists a bijection $F : \mathbb{N} \rightarrow A$. A set is at most countable if it is either finite or countable.

EXAMPLE 16.8. The set of all squares $S = \{n^2; n \in \mathbb{N}\}$ is countable; indeed $F : \mathbb{Z} \rightarrow S$, $F(n) = n^2$ is a bijection.

EXERCISE 16.9. Any infinite subset of a countable set is countable. Hint: It is enough to show that every subset $A \subset \mathbb{N}$ is countable. Let $F \subset \mathbb{N} \times \mathbb{N}$ be the set

$$F = \{(x, y) \in \mathbb{N} \times \mathbb{N}; y = \min(A \cap \{z \in \mathbb{N}; z > x\})\}$$

which is of course a map. By the recursion theorem there exists $G : \mathbb{N} \rightarrow \mathbb{N}$ such that $G(n+1) = F(G(n))$. One checks that G is injective and its image is A .

EXERCISE 16.10. Prove that $\mathbb{N} \times \mathbb{N}$ is countable. Hint: One can find injections $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$; e.g., $(m, n) \mapsto (m+n)^2 + m$.

EXERCISE 16.11. Prove that \mathbb{Q} is countable.

EXERCISE 16.12. Prove that $\mathcal{P}(\mathbb{N})$ is not countable.

Hint. Indeed this is a consequence of the more general theorem we proved that there is no bijection between a set A and its power set $\mathcal{P}(A)$. However it is interesting to give a reformulation of the argument in this case (Cantor's diagonal

argument). Assume $\mathcal{P}(\mathbb{N})$ is countable and seek a contradiction. Since $\mathcal{P}(\mathbb{N})$ is in bijection with $\{0, 1\}^{\mathbb{N}}$ we get that there is a bijection $F : \mathbb{N} \rightarrow \{0, 1\}^{\mathbb{N}}$. Denote $F(n)$ by $F_n : \mathbb{N} \rightarrow \{0, 1\}$. Construct a map $G : \mathbb{N} \rightarrow \{0, 1\}$ by the formula

$$G(n) = \neg(F_n(n))$$

where $\neg : \{0, 1\} \rightarrow \{0, 1\}$, $\neg 0 = 1$, $\neg 1 = 0$. (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.

EXERCISE 16.13. 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 16.14. Let us call \mathcal{F} the set of all functions $f : \mathbb{N} \rightarrow \mathbb{N} \cup \{\infty\}$ where $\infty \notin \mathbb{N}$. If $f(n) = \infty$ we say that f with input n does not terminate (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 $g(n) = 2$.

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.

(This is Turing’s proof that there is no “program” that decides if a given “program” with a given input terminates. Indeed Turing defined the notion of “program” as being a special type of function in \mathcal{F} . One proves there are countably many programs. Let (f_n) be the sequence of all programs. One proves that if the oracle g of this sequence is a program then the Turing function attached to (f_n) is a program. One concludes that g cannot be a program.)

REMARK 16.15. Consider the following sentence called the continuum hypothesis:

For every set A if there exists an injection $A \rightarrow \mathcal{P}(\mathbb{N})$ then either there exists an injection $A \rightarrow \mathbb{N}$ or there exists a bijection $A \rightarrow \mathcal{P}(\mathbb{N})$.

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 these notes.

CHAPTER 17

Reals

Real numbers have been implicitly around throughout the history of mathematics as an expression of the idea of continuity of magnitudes. What amounts to an axiomatic introduction of the reals can be found in Euclid (and is attributed to Eudoxus). The first construction of the reals from the “discrete” (i.e., from the rationals) is due to Dedekind.

DEFINITION 17.1. (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.

EXAMPLE 17.2. Any rational number $x \in \mathbb{Q}$ can be identified with the real number

$$u_x = \{y \in \mathbb{Q}; x < y\}.$$

It is clear that $u_x = u_{x'}$ for $x, x' \in \mathbb{Q}$ implies $x = x'$. We identify any rational number x with u_x . So we may view $\mathbb{Q} \subset \mathbb{R}$.

DEFINITION 17.3. A real number $u \in \mathbb{R}$ is called irrational if $u \notin \mathbb{Q}$.

DEFINITION 17.4. 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

$$\begin{aligned} u + v &= \{x + y; x \in u, y \in v\} \\ u \times v &= \{xy; x \in u, y \in v\}. \end{aligned}$$

Note that this extends addition and multiplication on the non-negative rationals.

EXERCISE 17.5.

- 1) Prove that \leq is a total order on \mathbb{R} .
- 3) Prove that $u + v$ and $u \times v$ are real numbers.
- 4) Naturally extend the definition of addition $+$ and multiplication \times of real numbers to the case when the numbers are not necessarily ≥ 0 . Prove that $(\mathbb{R}, +, \times, -, 0, 1)$ is a field. 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 17.6. 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).

EXERCISE 17.7. Let $r > 0$ be a rational number. Prove that for every n natural there exist rational numbers t_n and s_n such that $t_n^2 \leq r \leq s_n^2$ and

$$s_n - t_n \leq (s_1 - t_1)/2^{n-1}.$$

(Hint: For $n = 1$ take $t_1 = 0$ and s_1 such that $s_1^2 > r$. Assuming the above is true for n one sets $t_{n+1} = t_n$ and $s_{n+1} = (t_n + s_n)/2$ if $t_n^2 \leq r \leq ((t_n + s_n)/2)^2$ and one sets $t_{n+1} = (t_n + s_n)/2$ and $s_{n+1} = s_n$ if $((t_n + s_n)/2)^2 \leq r < s_n^2$.)

EXERCISE 17.8. Prove that for $r \in \mathbb{Q}$, $r \geq 0$, the set

$$\sqrt{r} := \{x \in \mathbb{Q}; x \geq 0, x^2 > r\}$$

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 $x \geq 0$, $y \geq 0$, $z = xy$, $x^2 > r$, $y^2 > r$. It is enough to show that there exists a rational number ρ with $z > \rho^2 > r$ because then we can write $z = xy$ with $x = \rho$ and $y = z/\rho > \rho$. To show this use the previous exercise.)

EXERCISE 17.9. Prove that for every $r \in \mathbb{R}$ with $r > 0$ there exists a unique number $\sqrt{r} \in \mathbb{R}$ such that $\sqrt{r} > 0$ and $(\sqrt{r})^2 = r$.

EXERCISE 17.10. Prove that $\sqrt{2}$ is irrational i.e., $\sqrt{2} \notin \mathbb{Q}$. Hint: Assume there exists a rational number x 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 17.11. 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 irrationality of $\sqrt{2}$ 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.

EXERCISE 17.12. Prove that the set

$$\{r \in \mathbb{Q}; r > 0, r^2 < 2\}$$

has no supremum in \mathbb{Q} .

REMARK 17.13. Later we will prove that \mathbb{R} is not countable.

DEFINITION 17.14. For every $a \in \mathbb{R}$ we let $|a|$ be a or $-a$ according as $a \geq 0$ or $a \leq 0$, respectively.

EXERCISE 17.15. Prove the so-called triangle inequality:

$$|a + b| \leq |a| + |b|$$

for all $a, b \in \mathbb{R}$.

DEFINITION 17.16. For $a < b$ in \mathbb{R} we define the open interval

$$(a, b) = \{c \in \mathbb{R}; a < c < b\} \subset \mathbb{R}.$$

(Not to be confused with the pair $(a, b) \in \mathbb{R} \times \mathbb{R}$ which is denoted by the same symbol.)

Topology

Topology is about geometric properties that are invariant under *continuous* transformations. An early topological result is the formula of Descartes-Euler relating the number of vertices, edges, and faces of a convex polyhedron. (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 analysis and nowadays topological ideas are to be found everywhere in mathematics, including number theory. Here we will restrict ourselves to explaining the basic idea of continuity.

DEFINITION 18.1. 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 $\bigcup_{i \in I} U_i \in \mathcal{T}$.

A subset $U \subset X$ is called open if $U \in \mathcal{T}$. A subset $Z \subset X$ is called closed if $X \setminus Z$ is open. Elements of X are called points of X .

EXAMPLE 18.2. $\mathcal{T} = \mathcal{P}(X)$ is a topology on X .

EXAMPLE 18.3. $\mathcal{T} = \{\emptyset, X\} \subset \mathcal{P}(X)$ is a topology on X .

EXAMPLE 18.4. 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.

EXERCISE 18.5. Prove that \mathcal{T} in Example 18.4 is a topology.

EXERCISE 18.6. Prove that if $U, V \subset X$ then

$$\mathcal{T} = \{\emptyset, U, V, U \cup V, U \cap V, X\}$$

is a topology. Find the closed sets of X .

EXERCISE 18.7. 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 18.8. 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 18.9. 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 18.10. 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 18.11. Let X and X' be two topological spaces. A map $F : X \rightarrow X'$ is continuous if for all open $V \subset X'$ the set $F^{-1}(V) \subset X$ is open.

EXERCISE 18.12. 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 $F : X \rightarrow X'$ is continuous.

EXERCISE 18.13. If \mathcal{T} is the topology $\mathcal{T} = \mathcal{P}(X)$ on X and \mathcal{T}' is any topology on X' then every map $F : X \rightarrow X'$ is continuous.

EXERCISE 18.14. Prove that if X, X', X'' are topological spaces and $G : X \rightarrow X', F : X' \rightarrow X''$ are continuous maps then the composition $F \circ G : X \rightarrow X''$ is continuous.

EXERCISE 18.15. Give an example of two topological spaces X, X' and of a bijection $F : X \rightarrow X'$ 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 11.20.)

Motivated by the above phenomenon, one gives the following

DEFINITION 18.16. A homeomorphism between two topological spaces is a continuous bijection whose inverse is also continuous.

DEFINITION 18.17. 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.

EXERCISE 18.18. 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 18.19. Let X be a topological space and let $A \subset X$ be a subset. We say that A is connected 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 18.20. Prove that if $F : X \rightarrow X'$ is continuous and $A \subset X$ is connected then $F(A) \subset X'$ is connected.

DEFINITION 18.21. 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 for every open set U in X containing x the set $U \setminus \{x\}$ contains a point of A .

EXERCISE 18.22. 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 18.23. Let X be a topological space and $A \subset X$. We say A is compact 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 18.24. Prove that if X is a topological space and X is a finite set then it is compact.

EXERCISE 18.25. 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 18.26. Prove that no open interval (a, b) in \mathbb{R} is compact ($a < b$).

EXERCISE 18.27. Prove that if $F : X \rightarrow X'$ is a continuous map of topological spaces and $A \subset X$ is compact then $F(A) \subset X'$ is compact.

DEFINITION 18.28. A topological space X is a Hausdorff space 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 18.29. Prove the \mathbb{R} with the Euclidean topology is a Hausdorff space.

EXERCISE 18.30. Prove that if X is a Hausdorff space, $A \subset X$ is compact, and $x \in X \setminus A$ 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, $B = \{b_1, \dots, b_n\}$. Then let

$$\begin{aligned} U &= U_{b_1} \cap \dots \cap U_{b_n} \\ V &= V_{b_1} \cup \dots \cup V_{b_n}. \end{aligned}$$

DEFINITION 18.31. 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 $X_1 \times \dots \times X_n$ of n topological spaces.

EXERCISE 18.32. Prove that for every $r \in \mathbb{R}$ with $r > 0$, the set

$$D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 < r^2\}$$

is open in the product topology of \mathbb{R}^2 .

DEFINITION 18.33. 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 $F : U \rightarrow V$ 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 18.34. 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 $X \rightarrow X'$. 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} will be examined in the chapter on limits.

CHAPTER 19

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 19.1. (Hamilton) A complex number is a pair (a, b) where $a, b \in \mathbb{R}$. 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

$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d) \\ (a, b) \times (c, d) &= (ac - bd, ad + bc).\end{aligned}$$

REMARK 19.2. 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 19.3. 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 19.4. We set $i = (0, 1)$.

REMARK 19.5. $i^2 = -1$. Indeed

$$i^2 = (0, 1) \times (0, 1) = (0 \times 0 - 1 \times 1, 0 \times 1 + 1 \times 0) = (-1, 0) = -1.$$

REMARK 19.6. For every complex number $(a, b) = a + bi$. Indeed

$$(a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0)(0, 1) = a + bi.$$

DEFINITION 19.7. For every complex number $z = a + bi$ we define its absolute value

$$|z| = \sqrt{a^2 + b^2}.$$

EXERCISE 19.8. Prove the so-called triangle inequality:

$$|a + b| \leq |a| + |b|$$

for all $a, b \in \mathbb{C}$.

DEFINITION 19.9. For every complex number $z = a + bi$ we define its conjugate

$$\bar{z} = a - bi.$$

(The upper bar is not to be confused with the notation used in the chapter on residues.)

EXERCISE 19.10. Prove that for every $z, w \in \mathbb{C}$ we have

- 1) $\overline{z + w} = \bar{z} + \bar{w}$;
- 2) $\overline{z \cdot w} = \bar{z} \cdot \bar{w}$;
- 3) $\overline{z^{-1}} = \bar{z}^{-1}$ for $z \neq 0$;
- 4) $z \cdot \bar{z} = |z|^2$.

DEFINITION 19.11. For every complex number $z = a + bi \in \mathbb{C}$ and every real number $r > 0$ we define the open disk with center z and radius r ,

$$D(z, r) = \{w \in \mathbb{C}; |w - z| < r\} \subset \mathbb{C}.$$

A subset $U \subset \mathbb{C}$ is called open if for every $z \in U$ there exists an open disk centered at z and contained in U . Let $\mathcal{T} \subset \mathcal{P}(\mathbb{C})$ be the set of all open sets of \mathbb{C} .

EXERCISE 19.12. Prove that \mathcal{T} is a topology on \mathbb{C} ; we call this the Euclidean topology.

EXERCISE 19.13. Prove that \mathbb{C} cannot be given the structure of an ordered ring.

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. The arithmetic of \mathbb{Z} to be discussed below already appears in Euclid. Congruences and residue classes were introduced by Gauss.

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 . If a does not divide b we write $a \nmid b$.

EXAMPLE 20.2. $4|20$; $-4|20$; $6 \nmid 20$.

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 $0 \leq r < b$.

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. \square

DEFINITION 20.5. For $a \in \mathbb{Z}$ denote $\langle a \rangle$ the set $\{na; n \in \mathbb{Z}\}$ of integers divisible by a . For $a, b \in \mathbb{Z}$ denote by $\langle a, b \rangle$ the set $\{ma + nb; m, n \in \mathbb{Z}\}$ 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 $\langle a, b \rangle = \langle c \rangle$.

Proof. If $a = b = 0$ we can take $c = 0$. Assume a, b are not both 0. Then the set $S = \langle a, b \rangle \cap \mathbb{N}$ is non-empty. Let c be the minimum of S . Clearly $\langle c \rangle \subset \langle a, b \rangle$. Let us prove that $\langle a, b \rangle \subset \langle c \rangle$. Let $u = ma + nb$ and let us prove that $u \in \langle c \rangle$. 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

$$r = u - cq = (ma + nb) - (m'a + n'b)q = (m - m'q)a + (n - n'q)b \in \langle a, b \rangle.$$

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 $\langle a, b \rangle = \langle c \rangle$ for some $c \geq 1$. In particular $c|a$ and $c|b$. The hypothesis implies $c = 1$ hence $1 \in \langle a, b \rangle$. \square

One of the main definitions of number theory is

DEFINITION 20.8. An integer p is prime 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 $p \nmid a$ 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 $p|ab$, $p \nmid a$, $p \nmid b$, 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, \dots, p_s , where $s \geq 1$, such that*

$$n = p_1 p_2 \dots p_s.$$

Moreover any such representation is unique in the following sense: if

$$p_1 p_2 \dots p_s = q_1 q_2 \dots q_t$$

with p_i and q_j prime and $p_1 \leq p_2 \leq \dots$, $q_1 \leq q_2 \leq \dots$ then $s = t$ and $p_1 = q_1$, $p_2 = q_2$, \dots

Proof. Uniqueness follows from Euclid's Lemma 20.10. 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. \square

EXERCISE 20.12. Prove the uniqueness part in the above theorem.

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 \bar{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 $\bar{3} = \bar{10} = \{\dots, -4, 3, 10, 17, \dots\}$.

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

$$\mathbb{Z}/m\mathbb{Z} = \{\bar{0}, \bar{1}, \dots, \overline{m-1}\}.$$

So the residue classes mod m are: $\bar{0}, \bar{1}, \dots, \overline{m-1}$. 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

$$\begin{aligned}\bar{a} + \bar{b} &= \overline{a+b} \\ \bar{a} \times \bar{b} &= \overline{ab} \\ -\bar{a} &= \overline{-a}.\end{aligned}$$

EXERCISE 20.21. Check that the above definitions are correct, in other words that if $\bar{a} = \overline{a'}$ and $\bar{b} = \overline{b'}$ then

$$\begin{aligned}\overline{a+b} &= \overline{a'+b'} \\ \overline{ab} &= \overline{a'b'} \\ \overline{-a} &= \overline{-a'}.\end{aligned}$$

Furthermore check that $(\mathbb{Z}/m\mathbb{Z}, +, \times, -, \bar{0}, \bar{1})$ 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.

CHAPTER 21

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).

DEFINITION 21.1. A group is a tuple $(G, \star, ', e)$ consisting of a set G , a binary operation \star on G , a unary operation $'$ on G (write $'(x) = x'$), and an element $e \in G$ (called the identity element) such that for every $x, y, z \in G$ the following axioms are satisfied:

- 1) $x \star (y \star z) = (x \star y) \star z$;
- 2) $x \star e = e \star x = x$;
- 3) $x \star x' = x' \star x = e$.

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 21.2. For every group G , every element $g \in G$, and every $n \in \mathbb{Z}$ one defines $g^n \in G$ exactly as in Exercise 13.15.

EXERCISE 21.3. Check the above.

DEFINITION 21.4. Sometimes one writes $e = 1$, $x \star y = xy$, $x' = x^{-1}$, $x \star \dots \star x = x^n$ ($n \geq 1$ times). In the Abelian case one sometimes writes $e = 0$, $x \star y = x + y$, $x' = -x$, $x \star \dots \star x = nx$ ($n \geq 1$ times). These notations depend on the context and are justified by the following examples.

EXAMPLE 21.5. If R is a ring then R is an Abelian group with $e = 0$, $x \star y = x + y$, $x' = -x$. 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 21.6. If R is a field then $R^\times = R \setminus \{0\}$ 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 21.7. The set $\text{Perm}(X)$ of bijections $\sigma : X \rightarrow X$ from a set X into itself is a group with $e = 1_X$ (the identity map), $\sigma * \tau = \sigma \circ \tau$ (composition), σ^{-1} = inverse map. If $X = \{1, \dots, n\}$ then one writes $S_n = \text{Perm}(X)$ and calls this group the symmetric group. If $\sigma(1) = i_1, \dots, \sigma(n) = i_n$ one usually writes

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}.$$

EXERCISE 21.8. 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 21.9. A 2×2 matrix with coefficients in a field R is a map

$$A : \{1, 2\} \times \{1, 2\} \rightarrow R.$$

If the map is given by

$$A(1, 1) = a$$

$$A(1, 2) = b$$

$$A(2, 1) = c$$

$$A(2, 2) = d$$

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 $r \in R$ 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 $M_2(R)$ be the set of all matrices and $GL_2(R)$ be the set of all invertible matrices. Then the following are true:

- 1) $M_2(R)$ is a group with respect to addition of matrices;
- 2) $GL_2(R)$ 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 21.10. Prove 1), 2), 3), 4), 5) above.

EXAMPLE 21.11. 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 $F : G \rightarrow G'$ 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.

DEFINITION 21.12. 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 21.13. Show that if H is a subgroup of G then H , with the natural operation induced from G , is a group.

EXERCISE 21.14.

- 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 R is a field then the set

$$SL_2(R) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix}; a, b \in R, ad - bc = 1 \right\}$$

is a subgroup of $GL_2(R)$; it is called the special linear group.

- 5) If R is a field then the set

$$SO_2(R) = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix}; a, b \in R, a^2 + b^2 = 1 \right\}$$

is a subgroup of $SL_2(R)$; it is called the special orthogonal group.

DEFINITION 21.15. If $F : G \rightarrow G'$ is a group homomorphism define the kernel of F ,

$$\text{Ker } F = \{a \in G; F(a) = e'\}$$

and the image of F :

$$\text{Im } F = \{b \in G'; \exists a \in G, F(a) = b\}.$$

EXERCISE 21.16. Prove that $\text{Ker } F$ is a subgroup of G and $\text{Im } F$ 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 phrase *order relations*.)

DEFINITION 21.17. 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 21.18. Prove that $\langle g \rangle$ is a subgroup of G . We call $\langle g \rangle$ the subgroup generated by g .

DEFINITION 21.19. We say that a group G is cyclic if there exists $g \in G$ such that $G = \langle g \rangle$; g is called a generator of G .

EXAMPLE 21.20. \mathbb{Z} is cyclic. 1 is a generator of \mathbb{Z} ; -1 is also a generator of \mathbb{Z} .

EXERCISE 21.21. Prove that \mathbb{Q} is not cyclic.

DEFINITION 21.22. 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 21.23. The order of a finite group G is the cardinality $|G|$ of G .

EXERCISE 21.24. The order $o(g)$ of g equals the order $|\langle g \rangle|$ of $\langle g \rangle$.

EXERCISE 21.25. 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 21.26. 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 21.25, $k|Nl$. 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 21.27. Prove that if $o(a) = kl$ then $o(a^k) = l$.

THEOREM 21.28. (*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 $\pi : G \rightarrow \mathcal{P}(G)$ be the map $\pi(g) = gH \in \mathcal{P}(G)$. Let $\mathcal{X} = \pi(G)$ and let $\sigma : \mathcal{X} \rightarrow G$ 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

$$(21.1) \quad \mathcal{X} \times H \rightarrow G, \quad (X, h) \mapsto \sigma(X)h, \quad X \in \mathcal{X}, \quad h \in H$$

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 21.1 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 21.29. Check the injectivity of 21.1.

THEOREM 21.30. (*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 $p \nmid a$ let \bar{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 $\bar{a}^{p-1} = \bar{1}$. Hence $a^{p-1} \equiv 1 \pmod{p}$. So $a^p \equiv a \pmod{p}$. \square

CHAPTER 22

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 22.1. Let R be a field. A vector space is an Abelian group $(V, +, -, 0)$ together with a map $R \times V \rightarrow V$, $(a, v) \mapsto av$ 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) $1v = v$.

The elements of V are called vectors.

EXAMPLE 22.2. R^n is a vector space over R viewed with the operations

$$\begin{aligned}(a_1, \dots, a_n) + (b_1, \dots, b_n) &= (a_1 + b_1, \dots, a_n + b_n), \\ -(a_1, \dots, a_n) &= (-a_1, \dots, -a_n), \\ c(a_1, \dots, a_n) &= (ca_1, \dots, ca_n).\end{aligned}$$

DEFINITION 22.3. The elements $u_1, \dots, u_n \in V$ are linearly independent if whenever $a_1, \dots, a_n \in R$ satisfies $(a_1, \dots, a_n) \neq (0, \dots, 0)$ it follows that $a_1u_1 + \dots + a_nu_n \neq 0$.

DEFINITION 22.4. The elements $u_1, \dots, u_n \in V$ generate V if for every $u \in V$ there exist $a_1, \dots, a_n \in R$ such that $u = a_1u_1 + \dots + a_nu_n$. (We also say that u is a linear combination of u_1, \dots, u_n .)

DEFINITION 22.5. The elements $u_1, \dots, u_n \in V$ are a basis of V if they are linearly independent and generate V .

EXERCISE 22.6.

- 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 22.7. If V has a basis u_1, \dots, u_n then the map $R^n \rightarrow V$, $(a_1, \dots, a_n) \mapsto a_1u_1 + \dots + a_nu_n$ is bijective. Hint: Directly from definitions.

EXERCISE 22.8. If V is generated by u_1, \dots, u_n then V has a basis consisting of at most n elements. Hint: Considering a subset of $\{u_1, \dots, u_n\}$ minimal with the property that it generates V we may assume that every subset obtained from $\{u_1, \dots, u_n\}$

does not generate V . We claim that u_1, \dots, u_n are linearly independent. Assume not. Hence there exists $(a_1, \dots, a_n) \neq (0, \dots, 0)$ such that $a_1u_1 + \dots + a_nu_n = 0$. We may assume $a_1 = 1$. Then one checks that u_2, \dots, u_n generate V , contradicting minimality.

EXERCISE 22.9. Assume $R = \mathbb{F}_p$ and V has a basis with n elements. Then $|V| = p^n$.

THEOREM 22.10. *If V has a basis u_1, \dots, u_n and a basis v_1, \dots, 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, \dots, u_n generate V we may write $v_1 = a_1u_1 + \dots + a_nu_n$ with not all a_1, \dots, a_n zero. Renumbering u_1, \dots, u_n we may assume $a_1 \neq 0$. Hence v_1, u_2, \dots, u_n generates V . Hence $v_2 = b_1v_1 + b_2u_2 + \dots + b_nu_n$. But not all b_2, \dots, b_n can be zero because v_1, v_2 are linearly independent. So renumbering u_2, \dots, u_n we may assume $b_2 \neq 0$. So $v_1, v_2, u_3, \dots, u_n$ generates V . Continuing (one needs induction) we get that v_1, v_2, \dots, v_n generates V . So $v_{n+1} = d_1v_n + \dots + d_nv_n$. But this contradicts the fact that v_1, \dots, v_m are linearly independent. \square

EXERCISE 22.11. Give a quick proof of the above theorem in case $R = \mathbb{F}_p$. Hint: We have $p^n = p^m$ hence $n = m$.

DEFINITION 22.12. We say V is finite dimensional (or that it has a finite basis) if there exists a basis u_1, \dots, 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 22.10.)

DEFINITION 22.13. If V and W are vector spaces a map $F : V \rightarrow W$ is called linear if for all $a \in K$, $u, v \in V$ we have:

- 1) $F(au) = aF(u)$,
- 2) $F(u + v) = F(u) + F(v)$.

EXAMPLE 22.14. If $a, b, c, d, e, f \in R$ then the map $F : R^3 \rightarrow R^2$ given by

$$F(u, v, w) = (au + bv + cw, du + ev + fw)$$

is a linear map.

EXERCISE 22.15. Prove that if $F : V \rightarrow W$ is a linear map of vector spaces then $V' = F^{-1}(0)$ and $V'' = F(V)$ are vector spaces (with respect to the obvious operations). If in addition V and W are finite dimensional then V' and V'' are finite dimensional and

$$\dim V = \dim V' + \dim V''.$$

Hint: Construct corresponding bases.

EXERCISE 22.16. Give an example of a vector space that is not finite dimensional.

CHAPTER 23

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 23.1. Let $m, n \in \mathbb{N}$. An $m \times n$ matrix with coefficients in a field R is a map

$$A : \{1, \dots, m\} \times \{1, \dots, n\} \rightarrow R.$$

If $A(i, j) = a_{ij}$ for $1 \leq i \leq m$, $1 \leq j \leq n$ then we write

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

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

$$(a_1, \dots, a_n)$$

and are called row matrices. Similarly the elements of $R^{m \times 1}$ are of the form

$$\begin{pmatrix} a_1 \\ \dots \\ \dots \\ a_m \end{pmatrix}$$

and are called column matrices. If $A = (a_{ij}) \in R^{m \times n}$ then

$$u^1 = \begin{pmatrix} a_{11} \\ \dots \\ \dots \\ a_{m1} \end{pmatrix}, \dots, u^n = \begin{pmatrix} a_{1n} \\ \dots \\ \dots \\ a_{mn} \end{pmatrix}$$

are called the columns of A and we also write

$$A = (u^1, \dots, u^n).$$

Similarly

$$(a_{11}, \dots, a_{1n}), \dots, (a_{m1}, \dots, a_{mn})$$

are called the rows of A .

DEFINITION 23.2. 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 $I \in R^{m \times n}$ 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 23.3. 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

$$C = (c_{ij}), \quad c_{ij} = \sum_{s=1}^k a_{is} b_{sj}.$$

EXERCISE 23.4. 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 $A \in R^{n \times n}$.

EXERCISE 23.5. If $A \in R^{m \times k}, B \in R^{k \times n}$, and the columns of B are $b^1, \dots, b^n \in R^{k \times 1}$ then the columns of AB are $Ab^1, \dots, Ab^n \in R^{m \times 1}$ (where Ab^i is the product of the matrices A and b^i). In other words

$$B = (b^1, \dots, b^n) \Rightarrow AB = (Ab^1, \dots, Ab^n).$$

DEFINITION 23.6.

$$\left(\begin{array}{c} 1 \\ 0 \\ \dots \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \\ \dots \\ 0 \end{array} \right), \dots, \left(\begin{array}{c} 0 \\ 0 \\ \dots \\ 1 \end{array} \right)$$

is called the standard basis of $R^{m \times 1}$

EXERCISE 23.7. Prove that the above is indeed a basis of $R^{m \times 1}$.

Here is the link between linear maps and matrices:

DEFINITION 23.8. If $F: V \rightarrow W$ is a linear map of vector spaces and v_1, \dots, v_n and w_1, \dots, w_m are bases of V and W , respectively, then for $j = 1, \dots, n$ 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, \dots, v_n and w_1, \dots, w_m .

EXERCISE 23.9. Consider a matrix $A = (a_{ij}) \in R^{m \times n}$ and consider the map

$$F : R^{n \times 1} \rightarrow R^{m \times 1}, \quad F(u) = Au \quad (\text{product of matrices}).$$

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, \dots, e^n be the standard basis of $R^{n \times 1}$ and let f^1, \dots, f^m be the standard basis of $R^{m \times 1}$. Then

$$F(e^1) = Ae^1 = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{pmatrix} = a_{11}f^1 + \cdots + a_{m1}f^m.$$

A similar computation can be done for e^2, \dots, e^n .

EXERCISE 23.10. Let $F : R^{n \times 1} \rightarrow R^{m \times 1}$ 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 23.11. Let $G : R^{n \times 1} \rightarrow R^{k \times 1}$ and let $F : R^{k \times 1} \rightarrow R^{m \times 1}$ 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 23.12. 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 23.13.

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}^t = \begin{pmatrix} a & d \\ b & e \\ c & f \end{pmatrix}.$$

EXERCISE 23.14. Prove that:

- 1) $(A + B)^t = A^t + B^t$;
- 2) $(AB)^t = B^t A^t$;
- 3) $I^t = I$.

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 as follows.

DEFINITION 24.1. Let V and W be vector spaces over a field R and let

$$f : V^n = V \times \dots \times V \rightarrow W$$

be a map. We say f is multilinear if for every $v_1, \dots, v_n \in V$ and every $i \in \{1, \dots, n\}$ we have:

1) If $v_i = v'_i + v''_i$ then

$$f(v_1, \dots, v_n) = f(v_1, \dots, v'_i, \dots, v_n) + f(v_1, \dots, v''_i, \dots, v_n).$$

2) If $v_i = cv'_i$ then

$$f(v_1, \dots, v_n) = cf(v_1, \dots, v'_i, \dots, v_n).$$

EXAMPLE 24.2. $f : R^{3 \times 1} \times R^{3 \times 1} \rightarrow R$ 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 multilinear.

DEFINITION 24.3. A multilinear map $f : V^n = V \times \dots \times V \rightarrow W$ is called alternating if whenever $v_1, \dots, v_n \in V$ and there exist indices $i \neq j$ such that $v_i = v_j$ we have $f(v_1, \dots, v_n) = 0$.

EXAMPLE 24.4. f in Example 24.2 is not alternating. On the other hand $g : R^{2 \times 1} \times R^{2 \times 1} \rightarrow R$ defined by

$$g \left(\left(\begin{array}{c} a \\ c \end{array} \right), \left(\begin{array}{c} b \\ d \end{array} \right) \right) = 2ad - 2bc = 2 \det \left(\begin{array}{cc} a & b \\ c & d \end{array} \right)$$

is alternating.

LEMMA 24.5. If $f : V^n \rightarrow W$ is multilinear alternating and $v_1, \dots, v_n \in V$ then for every indices $i < j$ we have

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -f(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

Here $v_1, \dots, v_j, \dots, v_i, \dots, v_n$ is obtained from $v_1, \dots, v_i, \dots, v_j, \dots, v_n$ by replacing v_i with v_j and v_j with v_i while leaving all the other v s unchanged.

Proof. We have

$$\begin{aligned}
f(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_n) &= f(v_1, \dots, v_i, \dots, v_i, \dots, v_n) \\
&\quad + f(v_1, \dots, v_i, \dots, v_j, \dots, v_n) \\
&\quad + f(v_1, \dots, v_j, \dots, v_i, \dots, v_n) \\
&\quad + f(v_1, \dots, v_j, \dots, v_j, \dots, v_n).
\end{aligned}$$

Hence

$$0 = f(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + f(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

□

EXERCISE 24.6. Let $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ be a bijection. Then there exists $\epsilon(\sigma) \in \{-1, 1\}$ with the following property. Let $f : V^n \rightarrow W$ be any multilinear alternating map and $v_1, \dots, v_n \in V$. Then

$$f(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \epsilon(\sigma) \cdot f(v_1, \dots, v_n).$$

Hint: Induction on n . For the induction step distinguish two cases: $\sigma(1) = 1$ and $\sigma(1) \neq 1$. In the first case one concludes directly by the induction hypothesis. The second case can be reduced to the first case via Lemma 24.5.

We identify $(R^{n \times 1})^n$ with $R^{n \times n}$ by identifying a tuple of columns (b^1, \dots, b^n) with the $n \times n$ matrix whose columns are b^1, \dots, b^n . We denote $I = I_n$ the identity $n \times n$ matrix.

LEMMA 24.7. *There exists a multilinear alternating map*

$$f : R^{n \times n} \rightarrow R$$

such that $f(I) = 1$.

Proof. We proceed by induction on n . For n we take $f(a) = a$. Assume we constructed a multilinear alternating map

$$f_{n-1} : R^{(n-1) \times (n-1)} \rightarrow R$$

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 24.8. Check the last sentence in the proof above.

LEMMA 24.9. *If f and g are multilinear alternating maps $R^{n \times n} \rightarrow R$ and $f(I) \neq 0$ then there exists $c \in R$ such that $g(A) = cf(A)$ for all A .*

Proof. Let $A = (a_{ij})$. Let e^1, \dots, e^n be the standard basis of $R^{n \times 1}$. Then

$$g(A) = g\left(\sum_{i_1} a_{i_1 1} e^{i_1}, \dots, \sum_{i_n} a_{i_n n} e^{i_n}\right) = \sum_{i_1} \dots \sum_{i_n} a_{i_1 1} \dots a_{i_n n} g(e^{i_1}, \dots, e^{i_n}).$$

The terms for which i_1, \dots, i_n are not distinct are zero. The terms for which i_1, \dots, i_n are distinct are indexed by permutations σ . By Exercise 24.6 we get

$$g(A) = \left(\sum_{\sigma} \epsilon(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} \right) g(I).$$

A similar formula holds for $f(A)$ and the Lemma follows. \square

By Lemmas 24.7 and 24.9 we get:

THEOREM 24.10. *There exists a unique multilinear alternating map (called determinant)*

$$\det : R^{n \times n} \rightarrow R$$

such that $\det(I) = 1$.

EXERCISE 24.11. Using the notation in the proof of Lemma 24.7 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 24.9.

We also have:

THEOREM 24.12. *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 $f : R^{n \times n} \rightarrow R$ defined by

$$f(u^1, \dots, u^n) = \det(Au^1, \dots, Au^n)$$

for $u^1, \dots, u^n \in R^{n \times 1}$. By Lemma 24.9 there exists $c \in R$ such that

$$f(u^1, \dots, u^n) = c \cdot \det(u^1, \dots, u^n).$$

Hence

$$\det(Au^1, \dots, Au^n) = c \cdot \det(u^1, \dots, u^n).$$

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 24.13. Prove that $\epsilon(\sigma\tau) = \epsilon(\sigma)\epsilon(\tau)$ for every permutations $\sigma, \tau \in S_n$; in other words $\epsilon : S_n \rightarrow \{1, -1\}$ is a group homomorphism.

EXERCISE 24.14. Prove that if $A \in R^{n \times n}$ is a matrix such that $\det(A) \neq 0$ 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 24.7). Prove that $AB = BA = I$ using Exercise 24.11.

EXERCISE 24.15. Prove that if $A \in R^{n \times n}$ is a matrix then $\det(A) = \det(A^t)$.

Hint. Use Lemma 24.9.

EXERCISE 24.16. Let R be a field.

1) Prove that the set $GL_n(R) = \{A \in R^{n \times n}; \det(A) \neq 0\}$ is a group with respect to multiplication; $GL_n(R)$ is called the general linear group.

2) Prove that the set $SL_n(R) = \{A \in R^{n \times n}; \det(A) = 1\}$ is a subgroup of $GL_n(R)$; $SL_n(R)$ is called the special linear group.

3) Prove that the set $SO_n(R) = \{A \in R^{n \times n}; \det(A) = 1, AA^t = I\}$ is a subgroup of $SL_n(R)$; $SO_n(R)$ is called the special orthogonal group.

Check that for $n = 2$ the above correspond to the previously defined groups $GL_2(R), SL_2(R), SO_2(R)$.

EXERCISE 24.17.

1) Prove that if a linear map $F : V \rightarrow W$ is bijective then its inverse $F^{-1} : W \rightarrow V$ is also linear. Such a map will be called an isomorphism (of vector spaces).

2) Prove that the set $GL(V)$ of all isomorphisms $V \rightarrow V$ is a group under composition.

3) Assume V has a basis v_1, \dots, v_n and consider the map $GL(V) \rightarrow GL_n(R)$, $F \mapsto A_F$ where A_F is the matrix of F with respect to v_1, \dots, v_n . Prove that $GL(V) \rightarrow GL_n(R)$ is an isomorphism of groups.

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 25.1. 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 $f : \mathbb{N} \cup \{0\} \rightarrow R$, $i \mapsto a_i$ 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

$$f = (a_0, a_1, a_2, a_3, \dots).$$

We define $0, 1 \in R[x]$ by

$$0 = (0, 0, 0, 0, \dots),$$

$$1 = (1, 0, 0, 0, \dots).$$

If f is as above and $g = (b_0, b_1, b_2, b_3, \dots)$ then addition and multiplication are defined by

$$\begin{aligned} f + g &= (a_0 + b_0, a_1 + b_1, a_2 + b_2, a_3 + b_3, \dots), \\ fg &= (a_0b_0, a_0b_1 + a_1b_0, a_0b_2 + a_1b_1 + a_2b_0, a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0, \dots). \end{aligned}$$

We define the degree of $f = (a_0, a_1, a_2, a_3, \dots)$ as

$$\deg(f) = \min\{i; a_i \neq 0\}$$

if $f \neq 0$ and $\deg(0) = 0$. We also define

$$x = (0, 1, 0, 0, \dots)$$

and we write

$$a = (a, 0, 0, 0, \dots)$$

for every $a \in R$.

EXERCISE 25.2.

- 1) Prove that $R[x]$ with the operations above is a ring.
- 2) Prove that the map $R \rightarrow R[x]$, $a \mapsto a = (a, 0, 0, 0, \dots)$ is a ring homomorphism.
- 3) Prove that $x^2 = (0, 0, 1, 0, \dots)$, $x^3 = (0, 0, 0, 1, 0, \dots)$, etc.
- 4) Prove that if $f = (a_0, a_1, a_2, a_3, \dots)$ then

$$f = a_n x^n + \dots + a_1 x + a_0$$

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 25.3. If $R = \mathbb{Z}$ then

$$\begin{aligned} & (3x^2 + 5x + 1)(8x^3 + 7x^2 - 2x - 1) = \\ & = (3 \times 8)x^5 + (3 \times 7 + 5 \times 8)x^4 + (3 \times (-2) + 5 \times 7 + 1 \times 8)x^3 + \dots \end{aligned}$$

DEFINITION 25.4. For every $b \in R$ we define an element $f(b) \in R$ by

$$f(b) = a_n b^n + \dots + a_1 b + a_0.$$

So for every polynomial $f \in R[x]$ we can define a map (called the polynomial map defined by the polynomial f):

$$R \rightarrow R, b \mapsto f(b).$$

(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 25.5. If $R = \mathbb{F}_2$ and we consider the polynomial $f(x) = x^2 + x \in R[x]$ then $f = (\bar{0}, \bar{1}, \bar{1}, \bar{0}, \dots) \neq (\bar{0}, \bar{0}, \bar{0}, \dots) = 0$ as an element of $R[x]$; but the polynomial map $R \rightarrow R$ defined by f sends $\bar{1} \mapsto \bar{1}^2 + \bar{1} = \bar{0}$ and $\bar{0} \mapsto \bar{0}^2 + \bar{0} = \bar{0}$ so this map is the constant map with value $\bar{0}$. This shows that different polynomials (in our case f and 0) can define the same polynomial map.

EXERCISE 25.6. 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 25.7. 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 25.8. 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 25.9. Leibniz mistakenly thought that the polynomial $x^4 + 1$ should have no root in \mathbb{C} .

EXERCISE 25.10. Let $R = \mathbb{F}_7$. Show that:

- 1) $x^2 + \bar{1}$ has no root in R .
- 2) $\bar{2}$ is a root of $x^3 - x + \bar{1}$ in R .

EXERCISE 25.11. Let $R = \mathbb{F}_5$. Show that:

- 1) $\bar{2}$ is a root of $x^2 + \bar{1}$ in R .
- 2) $x^5 - x + 1$ has no root in \mathbb{F}_5 .

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 25.12. (*Lagrange*) *If R is a field then every polynomial of degree $d \geq 1$ has at most d roots in R .*

THEOREM 25.13. (*Fundamental Theorem of Algebra, Gauss*) *If $R = \mathbb{C}$ is the complex field then every polynomial of degree ≥ 1 has at least one root in \mathbb{C} .*

In what follows we prove Theorem 25.12. (Theorem 25.13 is beyond the scope of this course.) We need a preparation.

DEFINITION 25.14. A polynomial $f(x) = a_n x^n + \dots + a_1 x + a_0$ of degree n is monic if $a_n = 1$.

PROPOSITION 25.15. (*Long division*) 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 $f(x) = a_n x^n + \dots + a_0$, $a_n \neq 0$. 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 25.12. Assume there exists a polynomial f of degree $d \geq 1$ that has $d + 1$ roots. Choose f such that d is minimal and seek a contradiction. Let $a_1, \dots, a_{d+1} \in R$ 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., $r(x) = c \in R$. 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 $k = 1, \dots, d$ it follows that $0 = (a_k - a_{d+1})g(a_k)$. Since R is a field and $a_k - a_{d+1} \neq 0$ for $k = 1, \dots, d$ it follows that $g(a_k) = 0$ for $k = 1, \dots, d$. But $\deg(g) = d - 1$ which contradicts the minimality of d . \square

DEFINITION 25.16. A number $\alpha \in \mathbb{C}$ is called algebraic if there exists a polynomial $f(x) = a_n x^n + \dots + a_1 x + a_0$ of degree $n \geq 1$ with coefficients in \mathbb{Q} such that $f(\alpha) = 0$. A number $\alpha \in \mathbb{C}$ is called transcendental if it is not algebraic.

EXAMPLE 25.17. $\sqrt{2}$ is algebraic because it is a root of $x^2 - 2$.

EXERCISE 25.18. Prove that $\sqrt{\sqrt{3} + 4} + 5\sqrt{7}$ is algebraic.

REMARK 25.19. It is not clear that transcendental numbers exist. We will check that later.

EXERCISE 25.20. Prove that the set of algebraic numbers in \mathbb{C} is countable.

REMARK 25.21. The main problems about roots are:

1) Find the number of roots; in case $R = \mathbb{F}_p$ 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.

DEFINITION 25.22. Let R be a ring. 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 $R[x]$ which we simply denote by $R[x, y]$ and refer to as the ring of polynomials in two variables x, y with coefficients in R . Again $R[x, y]$ is a ring so we can consider the ring of polynomials $R[x, y][z]$ in one variable z with coefficients in $R[x, y]$ which we denote by $R[x, y, z]$ and which we refer to as the ring of polynomials in 3 variables x, y, z with coefficients in R , etc.

EXAMPLE 25.23.

$$3x^7y^4z - x^8 + x^4y^9z^2 + 5xyz^2 = ((x^4)y^9 + (5x)y)z^2 - ((3x^7)y^4)z - (x^8) \in \mathbb{Z}[x, y, z].$$

CHAPTER 26

Congruences

We discuss here polynomial congruences which lie at the heart of number theory. The main results below are due to Fermat, Lagrange, Euler, and Gauss.

DEFINITION 26.1. 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 $f(x) \pmod p$ (or a solution to the congruence $f(x) \equiv 0 \pmod p$) if $f(c) \equiv 0 \pmod p$; in other words if $p|f(c)$. Let $\bar{f} \in \mathbb{F}_p[x]$ 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 $f \pmod p$ if and only if the image \bar{c} of c in \mathbb{F}_p is a root of \bar{f} . We denote by $N_p(f)$ the number of roots of $f(x) \pmod p$ contained in $\{0, 1, \dots, p-1\}$; equivalently $N_p(f)$ is the number of roots of \bar{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 \bar{f} in \mathbb{F}_p then of course $N_p(f) = |Z_p(f)|$.

EXERCISE 26.2.

- 1) 3 is a root of $x^3 + x - 13 \pmod{17}$.
- 2) Any integer a is a root of $x^p - x \pmod p$; 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 $p \nmid a$.
- 4) $N_p(x^2 - 1) = 2$ if $p \neq 2$.

PROPOSITION 26.3. 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 26.4. Consider the polynomials

$$f(x) = x^{p-1} - 1 \text{ and } g(x) = (x-1)(x-2)\dots(x-p+1) \in \mathbb{Z}[x].$$

Prove that all the coefficients of the polynomial $f(x) - g(x)$ are divisible by p . Conclude that p divides the sums

$$\sum_{a=1}^{p-1} a = 1 + 2 + 3 + \dots + (p-1)$$

and

$$\sum_{1 \leq a < b \leq p-1} ab = 1 \times 2 + 1 \times 3 + \dots + 1 \times (p-1) + 2 \times 3 + \dots + 2 \times (p-1) + \dots + (p-2) \times (p-1).$$

EXERCISE 26.5. Assume $p \geq 5$ is a prime. Prove that the numerator of any fraction that is equal to

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{p-1}$$

is divisible by p^2 .

REMARK 26.6. 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 $f(x) \equiv 0 \pmod{p}$ 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 $f(x) \equiv 0 \pmod{p}$ has k solutions in $\{0, 1, \dots, p-1\}$). 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 $f(x, y, z, \dots)$ 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 25.12:

COROLLARY 26.7. Assume $p \equiv 1 \pmod{d}$. 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

$$g(x) = x^{d(k-1)} + x^{d(k-2)} + \dots + x^d + 1.$$

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. \square

COROLLARY 26.8.

1) If $p \equiv 1 \pmod{4}$ 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 $p \equiv 3 \pmod{4}$ 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 26.7 if $p \equiv 1 \pmod{4}$ 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 $p \equiv 3 \pmod{4}$ so $p = 4k + 3$ and assume $N_p(x^2 = -1) > 0$ so there exists $c \in \mathbb{Z}$ such that $c^2 \equiv -1 \pmod{p}$; we want to derive a contradiction. We have (by Fermat's Little Theorem) that $c^p \equiv c \pmod{p}$. Since $p \nmid c$ we get $c^{p-1} \equiv 1 \pmod{p}$. But

$$c^{p-1} \equiv c^{4k+2} \equiv (c^2)^{2k+1} \equiv (-1)^{2k+1} \equiv -1 \pmod{p},$$

a contradiction. \square

EXERCISE 26.9. Prove that:

1) If $p \equiv 1 \pmod{3}$ 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 $p \equiv 2 \pmod{3}$ 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 26.10. Let a be an integer not divisible by a prime p . The order of $a \pmod{p}$ is the smallest positive integer k such that $a^k \equiv 1 \pmod{p}$. We write $k = o_p(a)$. Clearly $o_p(a)$ equals the order $o(\bar{a})$ of the image \bar{a} of a in \mathbb{F}_p .

DEFINITION 26.11. 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 \bar{g} of g in \mathbb{F}_p^\times is a generator of the group \mathbb{F}_p^\times .

EXERCISE 26.12. Prove that g is a primitive root mod p if and only if it is not divisible by p and

$$g^{(p-1)/q} \not\equiv 1 \pmod{p}$$

for all primes $q|p - 1$.

EXERCISE 26.13. 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 proved by Gauss:

THEOREM 26.14. (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, $p - 1 = p_1^{e_1} \dots p_s^{e_s}$ with p_1, \dots, p_s distinct primes and $e_1, \dots, e_s \geq 1$. Let $i \in \{1, \dots, s\}$. By Corollary 26.7 $N_p(x^{p_i^{e_i}} - 1) = p_i^{e_i}$ 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 $c_i \pmod{p}$ which is not a root mod p of $x^{p_i^{e_i-1}} - 1$. So

$$c_i^{p_i^{e_i}} \equiv 1 \pmod{p},$$

$$c_i^{p_i^{e_i-1}} \not\equiv 1 \pmod{p}.$$

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 21.26

$$o_p(c_1 \dots c_s) = p_1^{e_1} \dots p_s^{e_s} = p - 1$$

so $c_1 \dots c_s$ is a primitive root mod p . \square

Geometry

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 (or algebraic) one. Synthetic geometry is geometry without algebra. Analytic geometry is geometry through algebra. 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 analytic geometry. We start with lines. Later we will look at more complicated curves.

DEFINITION 27.1. Let R be a field. The affine plane $\mathbb{A}^2 = \mathbb{A}^2(R)$ over R is the set $R^2 = R \times R$. A point $P = (x, y)$ in the plane is an element of $R \times R$. A subset $L \subset R \times R$ is called a line if there exist $a, b, c \in R$ such that $(a, b) \neq (0, 0)$ and

$$L = \{(x, y) \in R^2; ax + by + c = 0\}.$$

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.

DEFINITION 27.2. We sometimes write $L = L(R)$ if we want to stress that coordinates are in R .

EXERCISE 27.3. 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 (called sometimes P_1P_2) passing through P_1 and P_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:

$$L = \{(x, y) \in R \times R; y - y_1 = m(x - x_1)\}.$$

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 $\mathbb{A}^2 = R^2 = R \times R$ together with the set \mathcal{L} of all lines (in the sense above) is an affine plane in the sense of Definition 10.42.

REMARK 27.4. Not all affine planes in the sense of Definition 10.42 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 (Parts I and II) and Pappus (stated below) hold. See below for the “only if direction.”

EXERCISE 27.5. Prove that every line in $\mathbb{F}_p \times \mathbb{F}_p$ has exactly p points.

EXERCISE 27.6. How many lines are there in the plane $\mathbb{F}_p \times \mathbb{F}_p$?

EXERCISE 27.7. (Desargues’ Theorem, Part I) 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 $A_i A_j$ and $A'_i A'_j$ are not parallel and let P_{ij} be their intersection. Assume the 3 lines $A_1 A'_1, A_2 A'_2, A_3 A'_3$ have a point in common. Then prove that the points L_{12}, L_{13}, L_{23} are collinear (i.e., lie on some line). Hint: Consider the “space” $R \times R \times R$ 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 $R \times R \times R \rightarrow R \times R, (x, y, z) \mapsto (x, y)$ and show that lines project onto lines. Next show that configuration of points $A_i, A'_i \in R \times R$ can be realized as the projection of a similar configuration of points $B_i, B'_i \in R \times R \times R$ not contained in a plane. (Identifying $R \times R$ 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 $B_1 B'_1, B_2 B'_2, B_3 B'_3$ have a point in common.) Then prove “Desargues’ Theorem in Space” (by noting that if Q_{ij} is the intersection of $B_i B_j$ with $B'_i B'_j$ 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 27.8. (Desargues’ Theorem, Part II) Let $A_1, A_2, A_3, A'_1, A'_2, A'_3$ be distinct points in the plane. Assume the 3 lines $A_1 A'_1, A_2 A'_2, A_3 A'_3$ have a point in common or they are parallel. Assume $A_1 A_2$ is parallel to $A'_1 A'_2$ and $A_1 A_3$ is parallel to $A'_1 A'_3$. Prove that $A_2 A_3$ is parallel to $A'_2 A'_3$. Hint: Compute coordinates. There is an alternative proof that reduces Part II to Part I by using the “projective plane over our field.”

EXERCISE 27.9. (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 $P_2 Q_3$ and $P_3 Q_2$ 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 $L = \{(x, 0); x \in R\}, M = \{(0, y); y \in R\}$ (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 27.10. One can identify the projective plane $(\overline{\mathbb{A}^2}, \overline{\mathcal{L}})$ attached to the affine plane $(\mathbb{A}^2, \mathcal{L})$ with the pair $(\mathbb{P}^2, \check{\mathbb{P}}^2)$ defined as follows. Let $\mathbb{P}^2 = R^3 / \sim$ where $(x, y, z) \sim (x', y', z')$ if and only if there exists $0 \neq \lambda \in R$ 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 $\mathbb{A}^2 = R^2 = R \times R$ with the point $(x : y : 1) \in \mathbb{P}^2$. Identify a point $(x_0 : y_0 : 0)$ in the complement $\mathbb{P}^2 \setminus \mathbb{A}^2$ 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 $\mathbb{P}^2 \setminus \mathbb{A}^2$ with the line at infinity L_∞ of \mathbb{A}^2 . Hence we get an identification of \mathbb{P}^2 with $\overline{\mathbb{A}^2}$. Finally define a line in \mathbb{P}^2 as a set of the form

$$\overline{L} = \{(x : y : z); ax + by + cz = 0\}.$$

So under the above identifications,

$$\bar{L} = \{(x : y : 1); ax + by + c = 0\} \cup \{(x : y : 0); ax + by = 0\} = L \cup \{\widehat{L}\}$$

where L is the line in R^2 defined by $ax + by + c = 0$. Then define $\widehat{\mathbb{P}^2}$ to be the set of all lines \bar{L} in \mathbb{P}^2 . We get an identification of $\widehat{\mathbb{P}^2}$ with $\bar{\mathcal{L}}$.

Some familiar concepts such as area and distance can be defined in the above context. Assume in what follows that R is a field such that $2 \neq 0$ and identify R^2 with $R^{2 \times 1}$.

DEFINITION 27.11. Let $P_1, P_2, P_3 \in R^{2 \times 1}$ be 3 points in the plane. Define

$$\text{area}(P_1, P_2, P_3) = \frac{1}{2} \det(P_2 - P_1, P_3 - P_1) \in R.$$

EXERCISE 27.12. 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 $F : R^2 \rightarrow F^2$ be an isomorphism of vector spaces and A its matrix with respect to the canonical basis. Then $A \in SL_2(R)$ if and only if “ F preserves areas” in the sense that for all $P_1, P_2, P_3 \in R^2$ we have

$$\text{area}(F(P_1), F(P_2), F(P_3)) = \text{area}(P_1, P_2, P_3).$$

- 4) For every $P_0 \in R^{2 \times 1}$ area is “invariant under translation by P_0 ” in the sense that

$$\text{area}(P_1 + P_0, P_2 + P_0, P_3 + P_0) = \text{area}(P_1, P_2, P_3).$$

DEFINITION 27.13. Let $P_1, P_2 \in R^{2 \times 1}$ be 2 points in the plane. Define the distance squared between these points as

$$\text{dist}^2(P_1, P_2) = (P_2 - P_1)^t(P_2 - P_1) \in R.$$

EXERCISE 27.14. Prove that

- 1) $\text{dist}^2(P_1, P_2) = \text{dist}^2(P_2, P_1)$.
- 2) If $R = \mathbb{R}$ then $\text{dist}^2(P_1, P_2) = 0$ if and only if $P_1 = P_2$. (Show that this may fail for other fields.)
- 3) Let $F : R^2 \rightarrow F^2$ be an isomorphism of vector spaces and A its matrix with respect to the canonical basis. Then $A \in SO_2(R)$ if and only if “ F preserves areas” and also “preserves distances” in the sense that for all $P_1, P_2 \in R^{2 \times 1}$ we have

$$\text{dist}^2(F(P_1), F(P_2)) = \text{dist}^2(P_1, P_2).$$

- 4) For every $P_0 \in R^{2 \times 1}$, dist^2 is “invariant under translation by P_0 ” in the sense that

$$\text{dist}^2(P_1 + P_0, P_2 + P_0) = \text{dist}^2(P_1, P_2).$$

So far we were concerned with lines in the plane. Let us discuss now “higher degree curves.” We start with conics. Assume R is a field with $2 = 1 + 1 \neq 0$; equivalently R does not contain the field \mathbb{F}_2 .

DEFINITION 27.15. The circle of center $(a, b) \in R \times R$ and radius r is the set

$$C(R) = \{(x, y) \in R \times R; (x - a)^2 + (y - b)^2 = r^2\}.$$

DEFINITION 27.16. A line is tangent to a circle if it meets it in exactly one point. (We say that the line is tangent to the circle at that point.) Two circles are tangent if they meet in exactly one point.

EXERCISE 27.17. Prove that for every circle and every point on it there is exactly one line tangent to the circle at that point.

EXERCISE 27.18. Prove that:

1) A circle and a line meet in at most 2 points.

2) Two circles meet in at most 2 points.

EXERCISE 27.19. How many points does a circle of radius 1 have if $R = \mathbb{F}_{13}$? Same problem for \mathbb{F}_{11} .

EXERCISE 27.20. Prove that the circle $C(R)$ 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

$$C(R) \rightarrow SO_2(R), \quad (a, b) \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

is a bijective group homomorphism. (Cf. Exercise 21.14 for $SO_2(R)$.)

EXERCISE 27.21. 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 27.22. A conic is a subset $Q \subset R \times R$ of the form

$$Q = Q(R) = \{(x, y) \in R \times R; ax^2 + bxy + cy^2 + dx + ey + f = 0\}$$

for some $(a, b, c, d, e, f) \in R \times \dots \times R$, where $(a, b, c) \neq (0, 0, 0)$.

We refer to (a, b, c, d, e) as the equation of the conic and if the corresponding conic passes through a point we say that the equation of the conic passes through the point. We sometimes say “conic” instead of “equation of the conic.”

EXERCISE 27.23. Prove that if 5 points are given in the plane such that no 4 of them are collinear then there exists a unique conic passing through these given 5 points. Hint: Consider the vector space of all (equations of) conics that pass through a given set S of points. Next note that if one adds a point to S the dimension of this space of conics either stays the same or drops by one. Since the space of all conics has dimension 6 it is enough to show that for $r \leq 5$ the conics passing through r points are fewer than those passing through $r - 1$ of the r points. For $r = 4$, for instance, this is done by taking a conic that is a union of 2 lines.

We next investigate cubics.

DEFINITION 27.24. Let R be a field in which $2 = 1 + 1 \neq 0$, $3 = 1 + 1 + 1 \neq 0$. Equivalently R does not contain \mathbb{F}_2 or \mathbb{F}_3 . A subset $Z = Z(R) \subset R \times R$ is called an affine elliptic curve if there exist $a, b \in R$ with $4a^3 + 27b^2 \neq 0$ such that

$$Z(R) = \{(x, y) \in R \times R; y^2 = x^3 + ax + b\}.$$

We call $Z(R)$ the elliptic curve over R defined by the equation $y^2 = x^3 + ax + b$. Next we define the projective elliptic curve defined by the equation $y^2 = x^3 + ax + b$ as the set

$$E(R) = Z(R) \cup \{\infty\}$$

where ∞ is an element not belonging to $Z(R)$. (We usually drop the word “projective” and we call ∞ the point at infinity on $E(R)$.) If $(x, y) \in E(R)$ define $(x, y)' = (x, -y)$. Also define $\infty' = \infty$. Next we define a binary operation \star on $E(R)$ called the chord-tangent operation; we will see that $E(R)$ 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 $(x_1, y_1), (x_2, y_2) \in E(R)$ 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

$$L_{12} = \{(x, y) \in R \times R; y - y_1 = m(x - x_1)\}$$

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 $Z(R)$ 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 $E(R)$ with L_{12} ”; more precisely (x_3, y_3) is defined by solving the system consisting of the equations defining $E(R)$ 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

$$x^3 - m^2x^2 + \dots = 0.$$

x_1, x_2 are known to be roots of this equation. We define x_3 to be the third root which is then

$$x_3 = m^2 - x_1 - x_2;$$

so we define

$$y_3 = y_1 + m(x_3 - x_1).$$

Summarizing, the definition of (x_3, y_3) is

$$(x_3, y_3) = ((y_2 - y_1)^2(x_2 - x_1)^{-2} - x_1 - x_2, y_1 + (y_2 - y_1)(x_2 - x_1)^{-1}(x_3 - x_1))$$

if $(x_1, y_1) \neq (x_2, y_2)$, $(x_1, y_1) \neq (x_2, -y_2)$ and

$$(x_3, y_3) = ((3x_1^2 + a)^2(2y_1)^{-2} - x_1 - x_2, y_1 + (3x_1^2 + a)(2y_1)^{-1}(x_3 - x_1))$$

if $(x_1, y_1) = (x_2, y_2)$, $y_1 \neq 0$.

Then $E(R)$ with the above definitions is an Abelian group.

EXERCISE 27.25. Check the last statement. (N.B. Checking associativity is a very laborious exercise.)

EXERCISE 27.26. Consider the group $E(\mathbb{F}_{13})$ defined by the equation $y^2 = x^3 + 8$. 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 examples of cubics:

DEFINITION 27.27. A cubic is a subset $X = X(R) \subset R \times R$ of the form $X(R) = \{(x, y) \in R \times R; ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0\}$ where $(a, b, c, \dots, j) \in R \times \dots \times R$, $(a, b, c, d) \neq (0, \dots, 0)$.

As usual we refer to the tuple (a, b, c, \dots, j) as the equation of a cubic (or, by abuse, simply a cubic).

EXERCISE 27.28. (Three Cubics Theorem) Prove that if two cubics meet in exactly 9 points and if a third cubic passes through 8 of the 9 points then the third cubic must pass through the 9th point. Hint: First show that if $r \leq 8$ and r points are given then the set of cubics passing through them is strictly larger than the set of cubics passing through $r - 1$ of the r points. (To show this show first that no 4 of the 9 points are on a line. Then in order to find, for instance, a cubic passing through P_1, \dots, P_7 but not through P_8 one considers the cubics $C_i = Q_{1234i} + L_{jk}$, $\{i, j, k\} = \{5, 6, 7\}$, 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. From here one immediately gets a contradiction.) Once this is proved let P_1, \dots, P_9 be the points of intersection of the cubics with equations F and G . We know that the space of cubics passing through P_1, \dots, P_8 has dimension 2 and contains F and G . So every cubic in this space is a linear combination of F and G , hence will pass through P_9 .

EXERCISE 27.29. (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 P_2Q_3 with 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

$$Q_1P_2 \cup Q_2P_3 \cup Q_3P_1 \quad \text{and} \quad P_1Q_2 \cup P_2Q_3 \cup P_3Q_1$$

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 $C \cup A_2A_3$ passes through all these points except possibly A_1 . Then by the Three Cubics Theorem $C \cup A_2A_3$ passes through A_1 . Hence A_2A_3 passes through A_1 .

EXERCISE 27.30. Show how Pascal's Theorem implies Pappus' Theorem.

REMARK 27.31. An extended version of the Three Cubics Theorem implies the associativity of the chord-tangent operation on a cubic. (The extended version, to be used below, follows from the usual version by passing to the "projective plane"; we will not explain this proof here.) The rough idea is as follows. Let E be the elliptic curve and Q, P, R points on it. Let

$$\begin{aligned} PQ \cup E &= \{P, Q, U\} \\ \infty U \cup E &= \{\infty, U, V\} \\ VR \cup E &= \{V, R, W\} \\ PR \cap E &= \{P, R, X\} \end{aligned}$$

$$\infty X \cap E = \{\infty, X, Y\}.$$

Here ∞A is the vertical passing through a point A . 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

$$E \quad \text{and} \quad PQ \cup WR \cup YX$$

both pass through the 9 points

$$P, Q, R, U, V, W, X, Y, \infty.$$

On the other hand the cubic

$$\Gamma = UV \cup PR \cup QY$$

passes through all 9 points except W . By a generalization of the Three Cubics Theorem (covering the case when one of the points is ∞) we get that Γ passes through W hence QY passes through W . The above argument only applies to chords and not to tangents. When dealing with tangents one needs to repeat the argument and look at multiplicities. So making the argument rigorous becomes technical.

CHAPTER 28

Limits

We start discussing now some simple topics in analysis. Analysis is the study of “passing to the limit.” The key words are sequences, convergence, limits, and later differential 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.

DEFINITION 28.1. A sequence in \mathbb{R} is a map $F : \mathbb{N} \rightarrow \mathbb{R}$; if $F(n) = a_n$ we denote the sequence by a_1, a_2, a_3, \dots or by (a_n) . We let $F(\mathbb{N})$ be denoted by $\{a_n; n \geq 1\}$; the latter is a subset of \mathbb{R} .

DEFINITION 28.2. A subsequence of a sequence $F : \mathbb{N} \rightarrow \mathbb{R}$ is a sequence of the form $F \circ G$ where $G : \mathbb{N} \rightarrow \mathbb{N}$ is an increasing map. If a_1, a_2, a_3, \dots is F then $F \circ G$ is $a_{k_1}, a_{k_2}, a_{k_3}, \dots$ (or (a_{k_n})) where $G(n) = k_n$.

DEFINITION 28.3. A sequence (a_n) is convergent to $a_0 \in \mathbb{R}$ 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 $a_n \rightarrow a_0$ and we say a_0 is the limit of (a_n) . A sequence is called convergent if there exists $a \in \mathbb{R}$ such that the sequence converges to a . A sequence is called divergent if it is not convergent.

EXERCISE 28.4. 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

$$\left| \frac{1}{n} - 0 \right| < \epsilon;$$

it is enough to take N to be any integer such that $N > \frac{1}{\epsilon}$.

EXERCISE 28.5. Prove that $a_n = \frac{1}{\sqrt{n}}$ converges to 0.

EXERCISE 28.6. Prove that $a_n = \frac{1}{n^2}$ converges to 0.

EXERCISE 28.7. Prove that $a_n = n$ is divergent.

EXERCISE 28.8. Prove that $a_n = (-1)^n$ is divergent.

EXERCISE 28.9. Prove that if $a_n \rightarrow a_0$ and $b_n \rightarrow b_0$ then

1) $a_n + b_n \rightarrow a_0 + b_0$

2) $a_n b_n \rightarrow a_0 b_0$.

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) $\frac{a_n}{b_n} \rightarrow \frac{a_0}{b_0}$.

Hint for 1: Consider any $\epsilon > 0$. Since $a_n \rightarrow a_0$ there exists N_a such that for all $n \geq N_a$ we have $|a_n - a_0| < \frac{\epsilon}{2}$. Since $b_n \rightarrow b_0$ 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 28.10. Prove that if $a_n \rightarrow a$, $b_n \rightarrow b$, and $a_n \leq b_n$ for all $n \geq 1$ then $a \leq b$.

DEFINITION 28.11. A sequence F is bounded if the set $F(\mathbb{N}) \subset \mathbb{R}$ is bounded.

DEFINITION 28.12. A sequence F is increasing if F is increasing. A sequence F is decreasing if $-F$ is increasing.

DEFINITION 28.13. A sequence (a_n) is Cauchy 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 28.14. Prove that every convergent sequence is Cauchy.

EXERCISE 28.15. 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 $A_n = \{a_m; m \geq n\}$. If at least one of these sets has no maximal element we get an increasing subsequence by Proposition 16.5. 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. Statement 4 is easy. Statement 5 follows by combining the previous statements.

EXERCISE 28.16. Prove that every subset in \mathbb{R} which is bounded from below has an infimum; and every subset in \mathbb{R} which is bounded from above has a supremum.

DEFINITION 28.17. A function $F : \mathbb{R} \rightarrow \mathbb{R}$ is continuous at a point $a_0 \in \mathbb{R}$ if for every sequence (a_n) converging to a_0 we have that the sequence $(F(a_n))$ converges to $F(a_0)$.

EXERCISE 28.18. (ϵ and δ criterion). Prove that a function $F : \mathbb{R} \rightarrow \mathbb{R}$ is continuous at a_0 if and only if for every real $\epsilon > 0$ there exists a real $\delta > 0$ such that for every $a \in \mathbb{R}$ with $|a - a_0| < \delta$ we have $|F(a) - F(a_0)| < \epsilon$.

EXERCISE 28.19. Prove that a function $F : \mathbb{R} \rightarrow \mathbb{R}$ is continuous (for the Euclidean topology on both the source and the target) if and only if it is continuous at every point of \mathbb{R} .

EXERCISE 28.20. Prove that every polynomial function $f : \mathbb{R} \rightarrow \mathbb{R}$ (i.e., every function of the form $a \mapsto f(a)$ where f is a polynomial) is continuous.

EXERCISE 28.21. 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 $b_n - a_n \rightarrow 0$. (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 $a_n \rightarrow a_0$ and $b_n \rightarrow b_0$ 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.

DEFINITION 28.22. For $a, b \in \mathbb{R}$ the closed interval $[a, b] \subset \mathbb{R}$ is defined as

$$[a, b] = \{x \in \mathbb{R}; a \leq x \leq b\}.$$

EXERCISE 28.23. Prove that $[a, b]$ are closed in the Euclidean topology.

EXERCISE 28.24. Prove that the open intervals (a, b) and the closed intervals $[a, b]$ are connected in \mathbb{R} .

EXERCISE 28.25. (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 $b_n - a_n \rightarrow 0$, such that $[a_n, b_n]$ cannot be covered by finitely many U_i s. (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_i s; 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 $a_n \rightarrow a_0$ and $b_n \rightarrow b_0$ and $a_0 = b_0$. But $a_0 = b_0$ is in one of the U_i s; this U_i will completely contain one of the intervals $[a_n, b_n]$ which is a contradiction.

CHAPTER 29

Series

DEFINITION 29.1. 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 $\sum_{k=1}^{\infty} a_n$ is a convergent series and that this series converges to s ; we write

$$\sum_{k=1}^{\infty} a_n = s.$$

If the sequence (s_n) is divergent we say that $\sum_{k=1}^{\infty} a_n$ is a divergent series.

EXERCISE 29.2. 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 29.3. 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 29.2.

EXERCISE 29.4. Prove that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^k}$$

is convergent for $k \geq 3$.

EXERCISE 29.5. 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:

$$\frac{1}{2^n + 1} + \frac{1}{2^n + 2} + \frac{1}{2^n + 3} + \dots + \frac{1}{2^n + 2^n} > 2^n \times \frac{1}{2^n + 2^n} = \frac{1}{2}.$$

EXERCISE 29.6. Prove that $a^n \rightarrow 0$ if $|a| < 1$.

Hint: We may assume $0 < a < 1$. Note that (a^n) is decreasing. Since it is bounded it is convergent. Let α be its limit. Assume $\alpha \neq 0$ and get a contradiction by noting that

$$\frac{1}{a} = \frac{a^n}{a^{n+1}} \rightarrow \frac{\alpha}{\alpha} = 1.$$

EXERCISE 29.7. Prove that

$$\sum_{n=1}^{\infty} a^n = \frac{1}{1-a}$$

if $|a| < 1$.

EXERCISE 29.8. 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

$$\mathbb{R} \rightarrow \mathbb{R}, \quad a \mapsto \exp(a)$$

is called the exponential map. Prove that

$$\exp(a+b) = \exp(a)\exp(b).$$

EXERCISE 29.9. Prove that the function $\exp : \mathbb{R} \rightarrow \mathbb{R}$ is continuous.

Hint: Let $a \in \mathbb{R}$ and $\epsilon > 0$. It is enough to show that there exists $\delta > 0$ such that if $|b - a| < \delta$ then $|\exp(b) - \exp(a)| < \epsilon$. Show that there is a δ such that for every b with $|b - a| < \delta$ there exists an n such that for all $m \geq n$

$$\begin{aligned} \left| \sum_{k=0}^m \frac{a^k}{k!} - \sum_{k=0}^n \frac{a^k}{k!} \right| &< \frac{\epsilon}{3} \\ \left| \sum_{k=0}^m \frac{b^k}{k!} - \sum_{k=0}^n \frac{b^k}{k!} \right| &< \frac{\epsilon}{3} \\ \left| \sum_{k=0}^n \frac{b^k}{k!} - \sum_{k=0}^n \frac{a^k}{k!} \right| &< \frac{\epsilon}{3}. \end{aligned}$$

From the first two inequalities we get

$$\begin{aligned} \left| \exp(a) - \sum_{k=0}^n \frac{a^k}{k!} \right| &\leq \frac{\epsilon}{3} \\ \left| \exp(b) - \sum_{k=0}^n \frac{b^k}{k!} \right| &\leq \frac{\epsilon}{3}. \end{aligned}$$

Then

$$|\exp(b) - \exp(a)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

EXERCISE 29.10. Let $S \subset \{0, 1\}^{\mathbb{N}}$ 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

$$\{0, 1\}^{\mathbb{N}} \setminus S \rightarrow \mathbb{R}, \quad (a_n) \mapsto \sum_{n=1}^{\infty} \frac{a_n}{2^n}$$

is (well defined and) injective. Conclude that \mathbb{R} is not countable.

EXERCISE 29.11. Prove that there exist transcendental numbers in \mathbb{R} . Hint: \mathbb{R} is uncountable whereas the set of algebraic numbers is countable; cf. Exercise 25.20. 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:

DEFINITION 29.12. A sequence (z_n) in \mathbb{C} is convergent to $z_0 \in \mathbb{C}$ 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 $z_n \rightarrow z_0$ and we say z_0 is the limit of (z_n) . A sequence is called convergent if there exists $z \in \mathbb{C}$ such that the sequence converges to z . A sequence is called divergent if it is not convergent.

EXERCISE 29.13. Let (z_n) be a sequence in \mathbb{C} and let

$$z_n = a_n + b_n i,$$

$a_n, b_n \in \mathbb{R}$. Let $z_0 = a_0 + b_0 i$. Prove that $z_n \rightarrow z_0$ if and only if $a_n \rightarrow a_0$ and $b_n \rightarrow b_0$.

DEFINITION 29.14. A sequence (z_n) in \mathbb{C} is Cauchy 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 29.15. Prove that a sequence in \mathbb{C} is convergent if and only if it is Cauchy.

EXERCISE 29.16. Prove that:

1) The series

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

is convergent for all $z \in \mathbb{C}$; 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) $\overline{\exp(z)} = \exp(\bar{z})$ for all $z \in \mathbb{C}$.

4) $|\exp(it)| = 1$ for all $t \in \mathbb{R}$.

5) The map

$$\mathbb{C} \rightarrow \mathbb{C}, \quad z \mapsto \exp(z),$$

is continuous. This map is called the (complex) exponential map.

There is a version of the above theory in p -adic analysis (which is crucial to number theory). Recall the ring of p -adic numbers \mathbb{Z}_p whose elements are denoted by $[a_n]$.

DEFINITION 29.17. Say that p^e divides $\alpha = [a_n]$ 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 $p^{v+1} \nmid a_{v+1}$. Then define the norm of α by the formula $|\alpha| = p^{-v(\alpha)}$. We also set $|0| = 0$.

EXERCISE 29.18. Prove that if $\alpha = [a_n]$ and $\beta = [b_n]$ then

$$|\alpha + \beta| \leq \max\{|\alpha|, |\beta|\}.$$

DEFINITION 29.19. Consider a sequence $[a_{n1}], [a_{n2}], [a_{n3}], \dots$ of elements in \mathbb{Z}_p which for simplicity we denote by $\alpha_1, \alpha_2, \alpha_3, \dots$

1) $\alpha_1, \alpha_2, \alpha_3, \dots$ is called a Cauchy sequence 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 $\alpha_1, \alpha_2, \alpha_3, \dots$ converges to some $\alpha_0 \in \mathbb{Z}_p$ 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 $\alpha_n \rightarrow \alpha_0$.

EXERCISE 29.20. Prove that a sequence in \mathbb{Z}_p is convergent if and only if it is Cauchy.

The following is in deep contrast with the case of \mathbb{R} :

EXERCISE 29.21. Prove that if (α_n) is a sequence in \mathbb{Z}_p with $\alpha_n \rightarrow 0$ then the sequence $s_n = \sum_{k=1}^n \alpha_k$ is convergent in \mathbb{Z}_p ; the limit of the latter is denoted by $\sum_{n=1}^{\infty} \alpha_n$.

EXERCISE 29.22.

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 . Hint: 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.

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 $(n!) \cdot \frac{a^n}{n!} = a^n$. Hint: Use 2) above.

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 $\exp_p(a)$.

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 30.1. 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 30.2. 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 30.3. Prove that the map $f : \mathbb{R} \rightarrow SO_2(\mathbb{R})$ defined by

$$f(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

is a group homomorphism.

EXERCISE 30.4. 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

$$H = \{nT; n \in \mathbb{Z}\}.$$

Hint: One first shows that if T is the infimum of the set

$$\{a \in H; a > 0\}$$

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 $a_n \rightarrow 0$. Deduce from this and the fact that H is closed that $G = \mathbb{R}$, a contradiction. Finally one shows that $H = \{nT; n \in \mathbb{Z}\}$ using an argument similar to the one used to prove Proposition 20.6.

EXERCISE 30.5. Prove that the map

$$F : \mathbb{R} \rightarrow \mathbb{C}^\times, F(t) = \exp(it)$$

is non-constant and non-injective. Conclude that there exists a unique real number $\pi \in \mathbb{R}, \pi > 0$, such that

$$\text{Ker } F = \{2n\pi; n \in \mathbb{Z}\}.$$

(This is our definition of the number π . In particular, by this very definition one gets $\exp(\pi i) + 1 = 0$ which is a celebrated formula of Euler; for us this is a trivial consequence of our definition of π .)

CHAPTER 31

Calculus

Calculus was invented by Newton and Leibniz, motivated by problems in mechanics and analytic geometry. The main concept of calculus is that of derivative of a function which we briefly review here.

DEFINITION 31.1. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a map and $a_0 \in \mathbb{R}$. We say F is differentiable at a_0 if there exists a real number (denoted by) $F'(a_0) \in \mathbb{R}$ such that for every sequence $a_n \rightarrow a_0$ with $a_n \neq a_0$ we have

$$\frac{F(a_n) - F(a_0)}{a_n - a_0} \rightarrow F'(a_0).$$

EXERCISE 31.2. Prove that if F is differentiable at $a_0 \in \mathbb{R}$ then it is continuous at a_0 .

EXERCISE 31.3. Prove that if F is a constant function (i.e., $F(x) = F(y)$ for all $x, y \in \mathbb{R}$) then F is differentiable at every a and $F'(a) = 0$.

DEFINITION 31.4. We say $F : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable if F is differentiable at every $a \in \mathbb{R}$. If this is the case the map $a \mapsto F'(a)$ is called the derivative of F and is denoted by $F' = \frac{dF}{dx} : \mathbb{R} \rightarrow \mathbb{R}$. If F' is differentiable we say F is twice differentiable and F'' is called the second derivative of F . One similarly defines what it means for F to be n times differentiable ($n \in \mathbb{N}$). We say F is infinitely differentiable (or smooth) if it is n times differentiable for every $n \in \mathbb{N}$. One denotes by $C^\infty(\mathbb{R})$ the set of smooth functions. We denote by $D : C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$ the map $D(F) = F'$.

EXERCISE 31.5. Prove that for every $F, G \in C^\infty(\mathbb{R})$ we have $F + G, F \cdot G \in C^\infty(\mathbb{R})$ 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 $C^\infty(\mathbb{R})$ is a ring with respect to $+$ and \cdot ; 0 and 1 are the functions $0(x) = 0$ and $1(x) = 1$.

EXERCISE 31.6. Prove that every polynomial function $F : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$\frac{a_n^k - a_0^k}{a_n - a_0} = a_n^{k-1} + a_n^{k-2} a_0 + \dots + a_0^{k-1} \rightarrow k a_0^{k-1}.$$

EXERCISE 31.7. Prove that $F(x) = \exp(x)$ is differentiable and $F'(x) = \exp(x)$. Hence F is smooth.

EXERCISE 31.8. 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 31.9. (Chain rule) Prove that if $F, G \in C^\infty(\mathbb{R})$ then $F \circ G \in C^\infty(\mathbb{R})$ and

$$D(F \circ G) = (D(F) \circ G) \cdot D(G).$$

(Here, as usual, \circ denotes composition.)

More generally one can define derivatives of functions of several variables as follows:

DEFINITION 31.10. Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function. Let $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and define $F_i : \mathbb{R} \rightarrow \mathbb{R}$ by

$$F_i(x) = F(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$$

(with the obvious adjustment if $i = 1$ or $i = n$). We say that F is differentiable with respect to x_i at a if 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 it is differentiable with respect to x_i at every $a \in \mathbb{R}^n$. For such a function we have a well defined function $\frac{\partial F}{\partial x_i} : \mathbb{R}^n \rightarrow \mathbb{R}$ which is also denoted by $D_i F$. We say that F is infinitely differentiable (or smooth) if F is differentiable, each $D_i F$ is differentiable, each $D_i D_j F$ is differentiable, each $D_i D_j D_k F$ is differentiable, etc. We denote by $C^\infty(\mathbb{R}^n)$ the set of smooth functions; it is a ring with respect to pointwise addition and multiplication.

DEFINITION 31.11. Let $P \in C^\infty(\mathbb{R}^{r+2})$. An equation of the form

$$P\left(x, F(x), \frac{dF}{dx}(x), \frac{d^2F}{dx^2}(x), \dots, \frac{d^r F}{dx^r}(x)\right) = 0$$

is called a differential equation. Here $F \in C^\infty(\mathbb{R})$ is an unknown function and one defines $\frac{d^i F}{dx^i} = D^{i+1}(F) = D(D^i(F))$.

The study of differential equations has numerous applications within mathematics (e.g., geometry) as well as natural sciences (e.g., physics).

EXAMPLE 31.12. Here is a random example of a differential equation:

$$\exp\left(\left(\frac{d^2 F}{dx^2}\right)^5\right) - x^3 \left(\frac{dF}{dx}\right) \left(\frac{d^3 F}{dx^3}\right) - x^5 F^6 = 0.$$

The additivity and the Leibniz rule have an algebraic . This suggests the following:

DEFINITION 31.13. Let R be a commutative unital ring. A map $D : R \rightarrow R$ is called a derivation if

- 1) $D(a + b) = D(a) + D(b)$ (additivity);
- 2) $D(a \cdot b) = a \cdot D(b) + b \cdot D(a)$ (Leibniz rule).

EXAMPLE 31.14. $D : C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$ is a derivation.

EXERCISE 31.15. Prove that every derivation $D : \mathbb{Z} \rightarrow \mathbb{Z}$ 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 31.16. Exercise 31.15 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*, Mathematical Surveys 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 $D = \frac{d}{dx} : C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$ is the operator

$$\frac{d}{dp} : \mathbb{Z} \rightarrow \mathbb{Z}, \quad \frac{dx}{dp} = \frac{x - x^p}{p}$$

which is well defined by Fermat's Little Theorem. For example,

$$\frac{d4}{d5} = \frac{4 - 4^5}{5} = -204.$$

CHAPTER 32

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. We begin with the simpler concept of correspondence.

DEFINITION 32.1. A correspondence on a set $X^{(0)}$ is a triple $(X^{(1)}, \sigma, \tau)$ where $\sigma, \tau : X^{(1)} \rightarrow X^{(0)}$ are maps called source and target.

EXAMPLE 32.2. If $R \subset A \times A$ is a relation and $\sigma : R \rightarrow A$ and $\tau : R \rightarrow A$ are defined by $\sigma(a, b) = a$, $\tau(a, b) = b$ then (R, σ, τ) is a correspondence.

DEFINITION 32.3. Assume $(X^{(1)}, \sigma, \tau)$ is a correspondence on $X = X^{(0)}$. Then one can define sets

$$X^{(2)} = \{(a, b) \in X^2; \tau(b) = \sigma(a)\}$$

$$X^{(3)} = \{(a, b, c) \in X^3; \tau(c) = \sigma(b), \tau(b) = \sigma(a)\}, \text{ etc.}$$

We have natural maps $p_1, p_2 : X^{(1)} \rightarrow X^{(0)}$, $p_1(a, b) = a$, $p_2(a, b) = b$.

DEFINITION 32.4. A category is a tuple $\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$ where $X^{(0)}$ is a set, $(X^{(1)}, \sigma, \tau)$ is a correspondence on $X^{(0)}$, and $\mu : X^{(2)} \rightarrow X^{(1)}$, $\epsilon : X^{(0)} \rightarrow X^{(1)}$ are maps. We assume that $\sigma \circ \epsilon = \tau \circ \mu = I$, the identity of $X^{(0)}$. Also we assume that the following diagrams are commutative:

$$\begin{array}{ccccc} X^{(2)} & \xrightarrow{\mu} & X^{(1)} & & X^{(2)} & \xrightarrow{\mu} & X^{(1)} & & X^{(3)} & \xrightarrow{\mu \times 1} & X^{(2)} \\ p_1 \downarrow & & \downarrow \tau & , & p_2 \downarrow & & \downarrow \sigma & , & 1 \times \mu \downarrow & & \downarrow \mu \\ X^{(1)} & \xrightarrow{\tau} & X^{(0)} & & X^{(1)} & \xrightarrow{\sigma} & X^{(0)} & & X^{(2)} & \xrightarrow{\mu} & X^{(0)} \end{array} .$$

Finally we assume that the compositions

$$X^{(1)} \xrightarrow{1 \times (\epsilon \circ \sigma)} X^{(2)} \xrightarrow{\mu} X^{(1)}$$

and

$$X^{(1)} \xrightarrow{(\epsilon \circ \tau) \times 1} X^{(2)} \xrightarrow{\mu} X^{(1)}$$

are the identity of $X^{(1)}$, where

$$(1 \times (\epsilon \circ \sigma))(a) = (a, \epsilon(\sigma(a))),$$

$$((\epsilon \circ \tau) \times 1)(a) = (\epsilon(\tau(a)), a).$$

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 $a : x \rightarrow y$. 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.

What we called a category in the above definition is sometimes called a small category; since our presentation will involve universes (see below) we do not need to make any distinction between small categories and categories.

REMARK 32.5. Given a category

$$\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$$

one can define the opposite category

$$\mathcal{C}^\circ = (X^{(0)}, X^{(1)}, \tau, \sigma, \mu \circ S, \epsilon)$$

where $S(a, b) = (b, a)$.

REMARK 32.6. Given a category

$$\mathcal{C} = (X^{(0)}, X^{(1)}, \sigma, \tau, \mu, \epsilon)$$

and a subset $Y \subset X^{(0)}$ one can define the full subcategory associated to Y by

$$\mathcal{C}_Y = (Y^{(0)}, Y^{(1)}, \tau_Y, \sigma_Y, \mu_Y, \epsilon_Y)$$

where $Y^{(0)} = Y$, $Y^{(1)} = \{a \in X^{(1)}; \sigma(a) \in Y, \tau(a) \in Y\}$, and $\tau_Y, \sigma_Y, \mu_Y, \epsilon_Y$ defined as the restrictions of $\tau, \sigma, \mu, \epsilon$, 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 32.7. (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 32.8. Let \mathcal{U} be a universe. Define the sets and maps:

$$\begin{aligned} X^{(0)} &= \mathcal{U}, \\ X^{(1)} &= \{(A, B, F) \in \mathcal{U}^3; F \in B^A\}, \\ \sigma(A, B, F) &= A, \quad \tau(A, B, F) = B. \\ \epsilon(A) &= I_A, \text{ (identity of } A) \end{aligned}$$

We get a category; indeed by the axiom of the universes if $A, B \in \mathcal{U}$ we have $B^A \in \mathcal{U}$ (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 32.9. If in the example above we insist that all F s are bijections we get a category called

$$\{\text{sets} + \text{bijections}\}.$$

This category is a groupoid.

EXAMPLE 32.10. 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 $F : A \rightarrow A$, 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 32.11. 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 $F : A \rightarrow A'$ increasing, we take μ to be again, composition, and we take $\epsilon(A, \leq) = I_A$.

EXAMPLE 32.12. 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 32.13. 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 32.14. 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 $(A, \star, \dots, \neg, \dots, 1, \dots)$ of the given type with A in a given universe, $X^{(1)}$ is the set of all triples

$$((A, \star, \dots, \neg, \dots, 1, \dots), (A', \star', \dots, \neg', \dots, 1', \dots), F)$$

with F a homomorphism, σ and τ are the usual source and target, and ϵ is the usual identity.

EXAMPLE 32.15. 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 $F : A \rightarrow B$ is a ring homomorphism. Also the target, source, and identity are the obvious ones; the composition map is the usual composition.

EXAMPLE 32.16. 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 $F : X \rightarrow X'$ 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 32.17. 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 $F : X \rightarrow X'$ 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 32.18. 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 32.19. If in the above example we restrict ourselves to groups which are Abelian we get the category of Abelian groups

$$\{\text{Abelian groups}\}.$$

EXAMPLE 32.20. 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 32.21. 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 $F : V \rightarrow W$ a linear map; source, target, and identity are defined in the obvious way.

EXAMPLE 32.22. Define the category

{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 $f_1, \dots, f_m \in \mathbb{C}[x_1, \dots, x_n]$ such that

$$X = \{(a_1, \dots, a_n) \in \mathbb{C}^n; f_1(a_1, \dots, a_n) = \dots = f_m(a_1, \dots, a_n) = 0\}.$$

(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 $F : X \rightarrow X'$ such that there exist polynomials $F_1, \dots, F_{n'} \in \mathbb{C}[x_1, \dots, x_n]$ with the property that for every $(a_1, \dots, a_n) \in X$ we have

$$F(a_1, \dots, a_n) = (F_1(a_1, \dots, a_n), \dots, F_{n'}(a_1, \dots, a_n)).$$

Composition of morphisms is composition of maps.

EXAMPLE 32.23. Define the category

{ordinary differential equations}

as follows. The set of objects consists of all pairs (\mathbb{R}^n, V) where $V : C^\infty(\mathbb{R}^n) \rightarrow C^\infty(\mathbb{R}^n)$ is a derivation which is \mathbb{R} -linear. (Such a V is called a vector field on \mathbb{R}^n .) A morphism $(\mathbb{R}^n, V) \rightarrow (\mathbb{R}^m, W)$ is a map $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 32.24. Check that, in all examples above, the axioms 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 32.25. A functor $\Phi : \mathcal{C} \rightarrow \tilde{\mathcal{C}}$ 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 $(\Phi^{(0)}, \Phi^{(1)})$,

$$\Phi^{(0)} : X^{(0)} \rightarrow \tilde{X}^{(0)}, \quad \Phi^{(1)} : X^{(1)} \rightarrow \tilde{X}^{(1)}$$

such that the following diagrams are commutative:

$$\begin{array}{ccccc} X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} & & X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} & & X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} \\ \sigma \downarrow & & \downarrow \tilde{\sigma} & , & \tau \downarrow & & \downarrow \tilde{\tau} & , & \epsilon \uparrow & & \uparrow \tilde{\epsilon} \\ X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} & & X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} & & X^{(0)} & \xrightarrow{\Phi^{(0)}} & \tilde{X}^{(0)} \\ & & & & & & & & & & \\ & & & & X^{(2)} & \xrightarrow{\Phi^{(2)}} & \tilde{X}^{(2)} & & & & \\ & & & & \mu \downarrow & & \downarrow \tilde{\mu} & & & & \\ & & & & X^{(1)} & \xrightarrow{\Phi^{(1)}} & \tilde{X}^{(1)} & & & & \end{array}$$

where $\Phi^{(2)} : X^{(2)} \rightarrow \tilde{X}^{(2)}$ 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 32.26. Consider the “forgetful” functor

$$\Phi : \{\text{commutative unital rings}\} \rightarrow \{\text{Abelian groups}\}$$

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 32.27. Consider the “forgetful” functor

$$\Phi : \{\text{commutative unital rings}\} \rightarrow \{\text{Abelian groups}\}$$

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 32.28. Consider the functor

$$\Phi : \{\text{Abelian groups}\} \rightarrow \{\text{groups}\}$$

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 32.29. Consider the “forgetful” functor

$$\Phi : \{\text{groups}\} \rightarrow \{\text{sets}\}$$

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 32.30. 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

$$\Phi : \{\text{topological spaces}\} \rightarrow \{\text{commutative unital rings}\}^\circ,$$

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 $f : X \rightarrow \mathbb{R}$, the addition $+$ and multiplication \cdot are the pointwise operations, and $0(x) = 0$, $1(x) = 1$. For every continuous map $F : X \rightarrow X'$ we let $\Phi(F) = F^*$ where, for $f : X' \rightarrow \mathbb{R}$, $F^*(f) = f \circ F$.

EXERCISE 32.31. Prove that all Φ s in the examples above are functors. (This is tedious but straightforward.)

EXAMPLE 32.32. Consider the following functor that, again, is the prototype for some important functors in functional analysis and algebraic geometry. The functor is

$$\Phi : \{\text{commutative unital rings}\}^\circ \rightarrow \{\text{topological spaces}\},$$

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 $\text{Spec } R$ be the set of all prime ideals in R . For any ideal I in R we let $D(I) \subset \text{Spec } R$ be the set of all prime ideals P such that $I \not\subset P$. Then the collection of all subsets of the form $D(I) \in \mathcal{P}(\text{Spec } R)$ is a topology on $\text{Spec } R$ called the Zariski topology. With this topology $\text{Spec } R$ becomes a topological space and we define

$$\Phi(R) = \text{Spec } R.$$

If $F : R \rightarrow R'$ is a ring homomorphism we define $\Phi(F) = F^* : \text{Spec } R' \rightarrow \text{Spec } R$ by $F^*(P') = F^{-1}(P')$.

EXERCISE 32.33. Prove that the collection $\{D(I); I \text{ an ideal in } R\}$ is a topology on $\text{Spec } R$. Prove that F^* is continuous. Prove that Φ is a functor.

EXERCISE 32.34. 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 $\text{Spec } \mathbb{Z}$ is not a Hausdorff space.

EXERCISE 32.35. Consider the following functor that plays a key role in algebraic geometry. The functor is

$$\Phi : \{\text{complex affine algebraic varieties}\} \rightarrow \{\text{topological spaces}\},$$

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 $F : X \rightarrow X'$ we let $\Phi(F) = F$ (which is continuous for the Euclidean topologies).

EXERCISE 32.36. Check that if $F : X \rightarrow X'$ is a morphism of complex affine algebraic varieties then F is continuous for the Euclidean topologies.

EXAMPLE 32.37. Consider the following functor that plays a key role in an area of mathematics called algebraic topology. The functor is

$$\Phi : \{\text{pointed topological spaces}\} \rightarrow \{\text{groups}\}$$

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 $\gamma : [0, 1] \rightarrow X$ 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 $\gamma_0, \gamma_1 : [0, 1] \rightarrow X$ are called homotopic (and write $\gamma_0 \sim \gamma_1$) if there exists a continuous map $F : [0, 1] \times [0, 1] \rightarrow X$ 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”

$$\Pi(X, x) \times \Pi(X, x) \rightarrow \Pi(X, x), \quad (\gamma_1, \gamma_2) \mapsto \gamma_1 \star \gamma_2,$$

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

$$(32.1) \quad \gamma_1 \star (\gamma_2 \star \gamma_3) \sim (\gamma_1 \star \gamma_2) \star \gamma_3.$$

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 from Equation 32.1. For every continuous map $F : X \rightarrow X'$ we let $\Phi(F) = F_*$ where, for $\gamma \in \Pi(X, x)$, we let $F_*([\gamma]) = [F \circ \gamma]$.

EXERCISE 32.38. Prove that \sim is an equivalence relation on $\Pi(X, x)$. Prove the homotopy in Equation 32.1. 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 32.39. (Equations and solutions) Let \mathcal{C} be a category as above. Let us define an equation to be a morphism $b \in \text{Hom}(y, z)$ and let $a \in \text{Hom}(x, z)$. Let us define the set of solutions in a of the equation b as the set

$$\text{Sol}(a, b) = \{c \in \text{Hom}(x, y); a = b \star c\}.$$

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 $f \in A = \mathbb{Z}[x_1, \dots, x_n]$ be a polynomial in variables x_1, \dots, 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 $b \in \text{Hom}(A/(f), \mathbb{Z})$ be the morphism corresponding to the natural ring homomorphism $\mathbb{Z} \rightarrow A/(f)$ and let $a \in \text{Hom}(R, \mathbb{Z})$ be the morphism corresponding to the natural homomorphism $\mathbb{Z} \rightarrow R$ where R is a field. Then we claim that there is a natural bijection

$$\psi : \text{Sol}(a, b) \rightarrow \{(c_1, \dots, c_n) \in R^n; f(c_1, \dots, c_n) = 0\}$$

given as follows: for every solution $c \in \text{Hom}(R, A/(f))$ corresponding to a ring homomorphism $c : A/(f) \rightarrow R$ we can attach the tuple $\psi(c) = (c([x_1]), \dots, c([x_n]))$ 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

$$(32.2) \quad \frac{d^r F}{dx^r} = Q(x, F(x), \frac{dF}{dx}(x), \dots, \frac{d^{r-1}F}{dx^{r-1}}(x))$$

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 $f \in C^\infty(\mathbb{R}^{r+1})$, $f = f(x, y_0, \dots, y_{r-1})$, we set

$$(32.3) \quad V_Q(f) = \frac{\partial f}{\partial x} + y_1 \frac{\partial f}{\partial y_0} + y_2 \frac{\partial f}{\partial y_1} + \dots + y_{r-1} \frac{\partial f}{\partial y_{r-2}} + Q(x, y_0, \dots, y_{r-1}) \frac{\partial f}{\partial y_{r-1}}.$$

We let $a : X \rightarrow X = Z$ be the identity and $b : Y \rightarrow X$ be defined by the first projection $\mathbb{R}^{r+1} \rightarrow \mathbb{R}$, $(x, y_0, \dots, y_{r-1}) \mapsto x$. Then there is a natural bijection

$$\Psi : \text{Sol}(a, b) \rightarrow \{F \in C^\infty(\mathbb{R}); F \text{ is a solution to 32.2}\}$$

given as follows. For every solution $c \in \text{Sol}(a, b)$ given by a map $c : \mathbb{R} \rightarrow \mathbb{R}^{r+1}$, $c(x) = (x, c_0(x), c_1(x), \dots, c_{r-1}(x))$, 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 of the form 32.2.

EXERCISE 32.40. 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 32.41. In the notation above prove that Ψ is well defined and a bijection.

EXAMPLE 32.42. (Symmetries) Let \mathcal{C} be a category and $x \in \text{Ob}(\mathcal{C})$. We denote by $\text{Aut}(x)$ the set of isomorphisms in $\text{Hom}(x, x)$. Then $(\text{Aut}(x), \star, ()^{-1}, 1_x)$ 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 $Z = \{\alpha_1, \dots, \alpha_n\}$ be the set of all roots of f in \mathbb{C} . Let

$$K_f = \{P(\alpha_1, \dots, \alpha_n); P \in \mathbb{Q}[x_1, \dots, x_n]\} \in \text{Ob}(\mathcal{C}).$$

(This is actually a field, called the splitting field of f .) Then $\text{Aut}(K_f)$ is the group of all ring isomorphisms $g : K_f \rightarrow K_f$; 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 $\{1, \dots, n\}$ such that $g(\alpha_i) = \alpha_{\sigma_g(i)}$ for all i . We get an injective homomorphism $G_f \rightarrow S_n$, $g \mapsto \sigma_g$.

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 32.3, the group $\text{Aut}(\mathbb{R}^{r+1}, V_Q)$ is viewed as the group of symmetries of the Equation 32.2. For $V = 0$ the group $\text{Aut}(\mathbb{R}^n, V)$ is called the diffeomorphism group of \mathbb{R}^n and is usually denoted by $\text{Diff}(\mathbb{R}^n)$.

Here is another example. Let \mathcal{C} be the category of all vector spaces over a field R . Then there is a natural isomorphism $Aut(R^n) \rightarrow GL_n(R)$.

EXERCISE 32.43. Look at other examples of categories and analyze the Aut groups of their objects.

EXAMPLE 32.44. (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 $x \rightarrow y$. 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 R consisting of the finite dimensional vector spaces then there is a natural bijection

$$Ob(\mathcal{C})/\simeq \rightarrow \mathbb{N} \cup \{0\}, [V] \mapsto \dim V.$$

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

$$I : Ob(\mathcal{C}) \rightarrow S$$

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

$$\bar{I} : Ob(\mathcal{C})/\simeq \rightarrow S, \bar{I}([x]) = I(x).$$

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 R 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 $I' : Ob(\mathcal{C}) \rightarrow S$ and $I'' : Ob(\mathcal{C}) \rightarrow S$. As corollaries one sometimes obtains interesting equalities between (integer or real) numbers.

Finally note that functors can produce systems of invariants as follows. If $\Phi : \mathcal{C} \rightarrow \mathcal{C}'$ is a functor then we have a natural map

$$I : Ob(\mathcal{C})/\simeq \rightarrow Ob(\mathcal{C}')/\simeq'.$$

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 32.45. 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 33

Models

We briefly indicate here how one can create a “mirror” of (pre-mathematical) 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 the formalization). The mirror of (pre-mathematical) logic in mathematics is not entirely accurate: there is no one to one correspondence between (pre-mathematical) logic and mathematical logic. The main concept in this discussion 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 33.1. 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:

$$\begin{aligned} c &= 0 \\ f_i &= (1, i), \\ r_i &= (2, i), \\ x_i^* &= (3, i), \\ \wedge^* &= 1, \\ \vee^* &= 2, \\ \neg^* &= 3, \\ \rightarrow^* &= 4, \\ \leftrightarrow^* &= 5, \\ \forall^* &= 6, \\ \exists^* &= 7, \\ =^* &= 8, \\ (*, *)^* &= 9, \\)^* &= 10, \\ C &= 0, \\ T &= \{c\} \cup \{f_1, f_2, f_3, \dots\} \cup \{r_1, r_2, r_3, \dots\}, \\ V &= \{x_1^*, x_2^*, x_3^*, \dots\}, \\ W &= \{\vee^*, \wedge^*, \neg^*, \rightarrow^*, \leftrightarrow^*, \forall^*, \exists^*, =^*, (*, *)^*\}. \end{aligned}$$

V is called the set of variables; W is called the set of logical symbols. We sometimes write x^*, y^*, z^*, \dots instead of $x_1^*, x_2^*, x_3^*, \dots$. By a T -partitioned set we mean in what follows a set S together with a map $S \rightarrow T$. 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 Λ_S^* with letters in Λ_S . One defines (in an obvious way, imitating the metadefinitions in the chapters on “pre-mathematical” logic) what it means for an element $\varphi \in \Lambda_S^*$ 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 33.2. Note that symbols “ \wedge, \vee, \dots ” are, of course, connectives in L_{set} while “ \wedge^*, \vee^*, \dots ” 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, \dots ” are the variables in L_{set} while “ x^*, y^*, z^*, \dots ” are constants in L_{set} .

EXAMPLE 33.3. 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 $\{\dots\}^*$ in a text in L_{set} one needs to replace $\{\dots\}^*$ 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 $\square \in S_{f_2}$ and a, x, z are as above then the word

$$\varphi = \{\forall x \exists z (\square(z, a) = z)\}^*$$

is an S -formula.

REMARK 33.4. 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 $M^n \rightarrow M$. 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

$$\mu : V \rightarrow M.$$

For every set M and assignment μ one can prove (by recursion) that there exists a map $v_{M, \mu} : \Lambda_{S(M)}^f \rightarrow \{0, 1\}$ 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. One example is the condition

$$\forall x \forall y ((x \in M) \wedge (y \in \mathcal{P}(M)) \rightarrow Q)$$

where

$$Q = “((v_M(y(*x)^*) = 1)) \leftrightarrow (x \in y)”$$

Another example of such a condition is

$$\forall y (y \in \mathcal{P}(M) \rightarrow P)$$

where

$$P := “((v_{M,\mu}(\forall^* z^* (*y(*x^*, z^*)^*)^*) = 1) \leftrightarrow (\forall z ((z \in M) \rightarrow ((\mu(x), z) \in y))))”.$$

Here

$$y(*x)^* = (y, (*, x,)^*),$$

$$\forall^* z^* (*y(*x^*, z^*)^*)^* = (\forall^*, z^*, (*, y, (*, x^*, z^*),)^*),$$

and the definition of Λ_S^f is such that the following is a theorem:

$$\forall x \forall y ((x \in M) \wedge (y \in \mathcal{P}(M)) \rightarrow (y(*x)^* \in \Lambda_S^f))$$

(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; we will not go here into this.

Next we discuss “semantics” of formal languages. The word “semantics” here is taken in a metaphorical sense.

DEFINITION 33.5. By a translation of Λ_S into $\Lambda_{S'}$ we understand a map

$$\mathbf{m} : S \rightarrow S'$$

which is compatible with the partitions (in the sense that $\mathbf{m}(S_t) \subset S'_t$ for all $t \in T$). For every $\varphi \in \Lambda_S^f$ one can form, in an obvious way, a formula $\mathbf{m}(\varphi) \in \Lambda_{S'}^f$, obtained from φ by replacing the constants and relational and functional symbols by their images under \mathbf{m} , respectively. So we get a map

$$\mathbf{m} : \Lambda_S^f \rightarrow \Lambda_{S'}^f$$

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 $\mathcal{M} = (M, \mathbf{m})$ where M is a set and \mathbf{m} is a translation

$$\mathbf{m} : S \rightarrow S(M).$$

So we get a map

$$\mathbf{m} : \Lambda_S^f \rightarrow \Lambda_{S(M)}^f,$$

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

$$v_M : \Lambda_{S(M)}^f \rightarrow \{0, 1\}$$

which is again a homomorphism compatible with \forall, \exists . So we may consider the composition

$$v_{\mathcal{M}} = v_M \circ \mathbf{m} : \Lambda_S^f \rightarrow \{0, 1\}.$$

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 Λ_S^s / \approx is a Boolean algebra in a natural way. Moreover each structure \mathcal{M} defines a homomorphism $v_{\mathcal{M}} : \Lambda_S^s / \approx \rightarrow \{0, 1\}$ of Boolean algebras.

EXAMPLE 33.6. 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 33.7. 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

$$\Phi^{\models} = \{\varphi \in \Lambda_S^s; \Phi \models \varphi\};$$

Φ^{\models} is called the formal theory generated by the system of axioms (S, Φ) .

REMARK 33.8. One thinks of Φ^{\models} as semantically defined because its definition involves models and hence translations. One may define a “syntactic” version of this set namely the set

$$\Phi^{\vdash} = \{\varphi \in \Lambda_S^s; \Phi \vdash \varphi\};$$

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 “do not say anything **about** mathematics itself.”

Commentary

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 axioms that are being introduced. Dogmatism definitely helped keeping our introduction minimal. But the question of a priori justifying our (meta)definitions and 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 $\triangle, \square, \dots$, 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, \dots 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}, \dots . 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}, \dots of axioms (which are variants of “axioms for mathematics”). For each n and m we can then give various metadefinitions $M_{n1m1}, M_{n1m2}, \dots$ of the concept of *theory*. All of these lead to corresponding theories $T_{n1m1}, T_{n1m2}, \dots$ (which are variants of “mathematics”). Within each T_{n1mk} one can set up various systems of definitions $D_{n1mk1}, D_{n1mk2}, \dots$ 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 notes identify with one of our sets of syntactic rules, say with S_1 . The language L_{set} introduced in Part 2 of our notes 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. Traditional metaphysics (from Aristotle to Medieval to Leibniz) is concerned with problems such as: “being qua being” (existence in itself), “ontology” (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. Modern metaphysics (especially post Kant) has put more stress on problems of epistemology (sensibility, understanding, knowledge, the dichotomies of a priori/a posteriori and analytic/synthetic, the nature of space/time/causality), problems of consciousness (the mind/body problem, free will), problems of psychology, etc. Our presentation of logic, and in particular of mathematics, is as free as possible from both traditional and modern 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., whose variables are *substance, predicate*, etc., and whose predicates are *exists, predicates, is one, is infinite*, etc. This is doable but leads to a “mock metaphysics” which is a “copy” of metaphysics inside logic. One can iterate this move and create a copy 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 copy inside mathematics which is mathematical logic. Now recall Alain Badiou’s maxim that the metaphysics of existence qua existence *is nothing but* [Cantor] 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 notes, 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 notes 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 notes, of course, are 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.

Physics. In our notes the term is used in the modern (rather than Aristotelian) sense; it is physics as a scientific body of knowledge, as it emerged after Galileo and Newton. It is written in a language containing abstract terms (force, mass, energy) and it comes with translations into both mathematics (Set Theory) and the language of empirical data (statements about measurements).

Realism, conceptualism, nominalism. According to (Quine 1980) the three doctrines in the 20th century philosophy of mathematics (*logicism, intuitionism, and formalism*) correspond to the three “medieval points of view regarding universals designated by historians as *realism, conceptualism, and nominalism*.” Realism holds that “universals or abstract entities have being independently of the mind.” (Putnam calls this type of realism *metaphysical realism* or *externalism*.) Conceptualism holds that “there are universals but they are mind-made.” And nominalism objects to “admitting abstract entities at all.” 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 notes are 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', \dots considered in these notes may be accused of being non-nominalist entities (because, for instance, they are potentially infinite entities whose very consideration seems to imply an abstract conception). This accusation has some merit and the way these notes implicitly respond 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 these notes defer all matters related to consistency and completeness to mathematical logic and essentially ban 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 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 his *Logics of Worlds* (Bloomsbury, London, New York, 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.* What our approach amounts to is to 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 these notes 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 these notes 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.

Intension and extension. Classically the *intension* of a predicate in a language is its “dictionary” definition. So what we call *definition* of a new predicate in our notes (e.g. *an astrochicken is a thing which is both a chicken and a spaceship*) corresponds to the intension of the predicate *is an astrochicken*. 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 $A(P) = \{x \in A; P(x)\}$ attached to a formula P corresponds to the idea of “extension of P ”. 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.

Comment on Tarski’s scheme. In a realist model theory (that is based on the maximal ontology of Cantor sets or on any other similar realist ontology), one classically says that a sentence P in an object language is *true in M* if the sentence becomes *true* as a sentence “about the elements of M .” This classical definition of truth in a (realist) model (essentially due to Tarski) 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 ! (As a matter of fact the whole Tarski scheme can be reproduced 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.) In any case our ban on truth predicates puts Tarski’s theory of truth outside the scope of our notes.

Truth and reference in Wittgenstein’s Tractatus. Wittgenstein’s *Tractatus Logico-Philosophicus* seems to be based on the concept of “*is the case*”, as synonymous of “*is true*,” and is, therefore, entirely perpendicular to the viewpoint of these notes. The very first sentence of the Tractatus reads, “*The world is all that is the case.*” The Tractatus then offers a vision of reality as a tightly articulated logical structure in which all sentences are either the case or not the case. Nothing is more remote from the picture put forward in our notes where objective truth is treated as metaphysics and banned from discourse. By the way, in relation to reference, the Tractatus 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). This amounts to an attempt at a logical proof of the fact that reference is based on similarity of form.

Similarity and reference. Similarity theory of reference is, of course, as old as philosophy, and 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.

Cross-reference. Introducing new definitions in a language (be it object language or metalanguage) is a form of *cross-reference*; cross-reference acts as *synonymy internal to the language* and, as such, it is different from what, in our notes, we call *reference* (the latter being always reference to things external to the language such as another language, as in linguistic reference to the "word Napoleon", or even noumenal objects such as the "man Napoleon".) There are other forms of cross-reference; the very use of variables (or, from a grammatical viewpoint, pronouns) plays the role of cross-reference. On the other hand *synonymy between fragments of different languages* is covered by our concept of translation, and hence, by semantics.

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, The problem of meaning in linguistics, 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.

Translation and possible worlds. Note that among translations of an object language into other object languages one finds what we called earlier the *interpretations* in *possible worlds* (or equivalently the *realist models*). The latter are, in

today's philosophy of language, the point of departure of semantics (e.g. of concepts such as extension); so our working definition of meaning is consistent with (although more minimalistic, hence more vague than) the accepted definition of meaning.

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" have different meanings but, as was discovered at some point, they turn out to have the same reference (the same physical star); if meaning and reference were the same the previous sentence would be tautological, which it is not, because it accounts for a discovery, so meaning and reference are different. In our notes 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 notes, 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.

Philogeny 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.

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.

Dialectic First order logic, which is the basis for mathematics, fixes its concepts in terms of definitions and axioms. And these are viewed as "outside time" ("eternal"). One can wonder where dialectic (Socratic, Hegelian, etc.) fits into this picture. Dialectic, identified with the 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.

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 Socrates is a king."

The grammatical (as opposed to logical) categories here are:

- nouns: *Socrates, father, king*
- verbs: *is*
- determinators: *a, the*

The sentence (S) above is constructed from a noun phrase (NP) "*the father of Socrates*" followed by a verb phrase (VP) "*is a king.*" The noun phrase "*the father of Socrates*" is constructed from the noun phrase "*the father*" and the prepositional phrase (PP) "*of Socrates.*" 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 Socrates*" is constructed from a preposition (P) "*of*" and a noun phrase which itself consists of a noun, "*Socrates.*" 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:

<i>S</i>						
<i>NP</i>			<i>VP</i>			
<i>NP</i>	<i>PP</i>		<i>V</i>		<i>NP</i>	
<i>D</i>	<i>N</i>	<i>P</i>	<i>N</i>	<i>V</i>	<i>D</i>	<i>N</i>
the	father	of	Socrates	is	a	king

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:

$[S[NP[NP[Dthe]_D[Nfather]_N]NP[PP[Pof]_P[Nsocrates]_N]PP]NP[VP...]VP]_S$.

Grammatical sentence formations are obtained by applying substitution rules symbolically written, for instance, as

$S \rightarrow NP VP$
 $NP \rightarrow NP PP$
 $VP \rightarrow V NP$
 $NP \rightarrow D NP$
 $NP \rightarrow N$
 $PP \rightarrow P NP$
 $N \rightarrow \text{father}$
 $N \rightarrow \text{Socrates}$
 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 *Formalese* or other languages of interest to science and mathematics. Mathematics requires logical (rather than grammatical) analysis.

Bibliography

- [1] Cantor, G. 1955. *Contributions to the Founding of the Theory of Transfinite Numbers*. New York: Dover.
- [2] Cassirer, E. 1953. *Language and Myth*. New York: Dover.
- [3] Chomsky, N. 2006. *Language and Mind*. Cambridge University Press.
- [4] Cohen, P. 2008. *Set Theory and the Continuum Hypothesis*, Dover.
- [5] Gödel, K. 1992. *On formally undecidable propositions of Principia Mathematica and related systems*. New York: Dover.
- [6] Hegel, G.W.F. 1975. *Hegel's Logic: Being Part One of the Encyclopaedia of the Philosophical Sciences*. Oxford University Press.
- [7] Hilbert, D. 1994. *Foundations of Geometry*. La Salle: Open Court.
- [8] Hilbert, D., Bernays, P. 1934. *Grundlagen der Mathematik*. Berlin.
- [9] Kant, I. 1991. *Critique of Pure Reason*. London: J. M. Dent & Sons.
- [10] Manin, Yu. I. 2009. *A Course in Mathematical Logic for Mathematicians*. New York: Springer.
- [11] Putnam, H. 1971. *Philosophy of Logic*. New York, San Francisco, London: Harper and Row.
- [12] Putnam, H. 1981. *Reason, Truth, and History*, Cambridge Univ. Press.
- [13] Quine, W. V. 1964. *Word and Object*, MIT Press.
- [14] Quine, W. V. 1980. *From a Logical Point of View*, Harvard University Press.
- [15] Quine, W. V. 1986. *Philosophy of Logic*. Cambridge: Harvard University Press.
- [16] Russell, B. 1993. *Introduction to Mathematical Philosophy*. New York: Dover.
- [17] Tarski, A. 1995. *Introduction to Logic and to the Methodology of Deductive Sciences*, New York: Dover.
- [18] Wang, Hao. 1996. *A Logical Journey: from Gödel to Philosophy*, Cambridge, MIT Press.
- [19] Weyl, H. 1963. *Philosophy of Mathematics and Natural Sciences*. New York: Atheneum.
- [20] Wittgenstein, L. 2001. *Tractatus Logico-Philosophicus*. Routledge: Taylor and Francis.