Foundations of Logical Thought

Johan Noldus[1]

March 17, 2026

[1] Johan.Noldus@gmail.com.

Contents

Some words upfront.

This text has been written to satisfy the curiosity of young eager minds who are willing to dig deep into the subject and acquire an understanding from a foundational linguistic point of view. The reader can find a wealth of ideas and exercises here which should enable him or her to get as well an intuitive as formalist grasp upon the subject. All exercises are original and require a high level of abstraction; the more formalist your argument is, the deeper one can dig into improving the way to speak about the subject.

This book contains all material presented in high school as well as in a master education in mathematics at a standard level university. The presentation is original and emphasizes creative thought and formalization skills over mechanically solving standard exercises; the reader is supplied with a huge amount of ideas regarding the formalization of standard concepts in nature which are alpha numerical or geometrical in nature. Every chapter proceeds by first acquiring an intuitive understanding, followed by a formalization of the latter. Subsequently, the reader is invited to solve standard exercises followed by challenges of a more abstract nature requiring profound symbolic reflection. Ultimately, the goal is to familiarize the reader with a low abstraction language and provide him or her with the necessary manipulative skills. It is not mandatory to write down every proof in an optimal symbolic way such that a computer can verify its truth something which is impossible, especially in the field of topology.

Mathematics distinguishes itself from other languages regarding the concept of truth which is absolute given the presumption of complete knowledge. In our world, knowledge is incomplete and reflects itself into the inadequacy in determining the state of a system resulting in a feeling of discomfort when something strange happens. In chapter fifteen, we proceed upon this matter when developing the field of vague or incomplete logic. Another way to proceed is not to obfuscate logic but to make language more fuzzy. This is not an interesting path to follow as it blocks the objectification process of truth something which is contrary to the goal of the scientific endeavor. From a young age onwards, I have experienced that an appropriate balance between metaphysical thought and computational power is the best strategy to proceed in ones understanding of the world. It is not realistic to digest more than five pages a day in this book for a very smart student. Hence, a minimal time span of a month needs to be taken into account for completing a first reading of this book; more in particular, it is utterly suited for a university course of a year during a bachelor education. The teacher is advised to create additional exercises in order to guide the class through the assimilation and creation process. Brilliant students however, should be able to make all exercises and acquire a full understanding of this book by themselves. This deficit can be a stimulant for other students to discuss their findings in a classroom environment. This version is a revised version of the original manuscript which contained a few, not many, errors which have now been removed.

Chapter 1

Mathematics, an Interesting Language

Every child has its own vocabulary what makes it difficult to speak to one and another sometimes. Most words have a slightly different meaning for various persons leading to misunderstandings at several levels. Mathematics is an attempt to say something as unambiguous as possible, a formidable task indeed. Even the notion of equality between different objects proves to be a troublesome one given that one needs a *comparative operation* which slightly alters things. For example, are two bars of gold of the same shape? Transporting one bar to the other can cause deformations which means that defining a uniform measure stick can be troublesome, a problem which is related to the notion of temperature. Given that mathematics is an exact language, it is possible to find connections between different concepts, something which we call theorems. The latter are deeply hidden in the language itself and we, humans, have the capacity to discover them; it is often an art to succinctly formulate the concepts such that the theorems become self evident. Likewise, one should take care that the theorem is not vacuous or deprived of content such as is the case for the Riemann hypothesis. In case you wish to prove a result and you find no logical reasoning leading towards it, chances are high that you do not work with the appropriate concepts. Youngsters often encounter that problem, they do not find the right speech for uttering their thoughts. This is a matter of training and practice and the reader shall gradually become aware of a variety of ways to reason about the same thing; the choices we make, the so called axioms are diverse in nature and have to potency to lead to different kinds of mathematics. Indeed, we make the reader aware that realizing one's imagination can deepen one's understanding to a huge degree; sometimes, I shall explain highly advanced methods of reasoning and the interested reader may consult more specialist books regarding those matters. My personal goal is to guide youngsters into the art of adequately phrasing their thoughts as well as finding logical coherences between those; the build-up of this book is entirely logical and the student is advised to understand every immediate step deeply.

This book is in the first place pedagogical in the sense that it submerges the reader into the world of language and truth; that truthfulness is relative and often is restricted by our language. This has serious implications for the science of physics which still uses spoken language and this may be a matter of principle instead of a lack of understanding. Spoken language is blurry which suggests a limited form of magic. Science is not necessary an adversary of magic though it captures her boundaries leading to a better world view. This book has to be understood in this way also, as an attempt to control the environment; at that point, a new form of magic arises, that language itself determines your science. Hence, science also controls your imagination, one is easily convinced that girls have more trouble with that as boys do. From this point of view, this book is also suitable for girls as it is written from the magical point of view given that all proofs are procured in the most metaphysical way possible without delving too much into a computational form of evidence.

A proof is a logical reasoning and therefore logic ought to be a part of mathematics. Without logic, no correct proof and no theorems. The kind of logic employed in mathematics is of the easiest kind, a theorem is either true or false which we label under Boolean logic. Different kinds of logic will be discussed in this book also, but they are all defined within the framework of the Boolean one. Right means that it fits in all cases and wrong signifies that there can be found at least one counterexample. Hence, we introduce the symbols $\forall$ and $\exists$ which mean "for all", respectively, "there exists". Furthermore, the symbols $A \wedge B$ and $A \vee B$ apply whereas the first one is true if and only if both A and B are and the second is true if and only if at least one them is. In human language, this reads A and B, respectively A or B. The reader can now verify that

$A \wedge (B \vee C)$ is the same as $(A \wedge B) \vee (A \wedge C)$, a formula known as the rule of de Morgan. These constitute the defining rules of classical logic, which is usually supplemented with an absolute negation, in contrast to intuitionistic logic which leads to Heyting algebra's. By definition $\neg A$ if and only if A is false and obviously $\neg\neg A = A$. The reader may verify that $\neg(A \vee B)$ is equivalent to $\neg A \wedge \neg B$. It is also possible to devise a logic where certain sentences are true or false with a certain probability. This kind of logic is called quantum logic which requires the principle of superposition and therefore more advanced mathematics as mere set theory. In quantum logic, one needs to take into account that verification of truthfulness, in either a reality of some kind, requires an operation which *changes* the state of the system which is not the case in classical logic. There, the system is given by the unique Platonic space of all truisms. This implies that, in quantum logic, verification of the sentence A *after* B does not provide the same statistics as the one resulting from the verification of B after A unless both changes are compatible. This is never the case for classical logic where verification of truths are always interchangeable. Furthermore, there exists the possibility to posit that the truthfulness of a sentence is undetermined which is a further deepening of the fuzziness of logic. The Rosetta stone of classical logic is that the truth or falsity of any sentence within the language cannot always be proved from the axioms. This result, known as the Godel incompleteness theorem, is largely of philosophical interest.

The main idea behind the proof of this result is well known in linguististics; within the science of mathematics it is always assumed that the verification of the vericacity of a sentence is alway possible. If a sentence of the type A implies B is true, then you may want to prove this by showing that $\neg B$ implies $\neg A$. A sentence which does not provide for a relation of the above kind between A and B is meaningless; for example "this sentence is true" is of that this; in the mathematical language this reads $B = (\chi(B) = 1)$ where χ is the truth indicator. The problem is obvious given that the sentence is recursive in way, it appears on both sides of the equation; henceforth, it is impossible to *prove* whether it is right or wrong because sentences and truth evaluations of them *entirely* cannot be mixed up. Likewise, one has "this sentence is false". If you assume it to be true then it must be false; reversely, assuming its falsity leads to its truthfulness. These are all sentence with self reference, where the truthfulness of a part depends holistically upon the entire framework itself. It is, just as sets, where a set is an element of itself, infinitely recursive and no stopping to it.

Let it be clear; the foundations of mathematics can be disputed and we shall elucidate this in the course of the first chapter when dealing with the axiom of choice which is shown to be not compatible with the other, more reasonable, axioms. Even the way of reasoning, called logic, is susceptible for alternation as is exemplified by the notion of topi in which a sentence can be false and true at the same time. This constitutes an example of quantal logic where the notion of proof should be entirely different and the notion of reality is a much weaker one as is the case in classical mathematics. We shall address these issues in this book by means of lucid exercises. There exist mathematicians of the opinion that the ultimate goal is to transcribe mathematical proofs in such a fashion that a computer may verify its truth: those belong the Bourbaki club erected in intellectual Paris. Although I can appreciate the utility of the philosophy, I am not of the opinion that it is a very useful one for people who know what they are talking about given the workload attached to such endeavor. My approach is different and more practical; however, the reader should understand that, often, the most difficult part of mathematics consists of phrasing exactly ones thoughts: schooling and training is required here and in this sense the French method is useful initially. Once having acquired a sufficient level of competency, this laborious method can be abandoned.

This book builds up the entire edifice of mathematics from scratch: we start by introducing and motivating set theory, something which you have all studied in primary school but never properly understood. For example, how to define the relationship between element and set symbolically? What about sets with an infinite number of elements: how to formalize infinity? Are the operations of taking the intersection and union really not sensitive to the order of the sets or is some quantal or non commutative effect relevant here? More specifically $A \wedge B$ must be understood as A and B are intersected in that order instead of the collection of elements which belong to A as well as B. This last interpretation is not mandatory and the method of Venn diagrams is rather ridiculously restrictive in light of the first interpretation. However, we think as the French here, start with the most symmetrical situation and then look for constructive ways, within that setting, to break that symmetry. For example, every building of the Louvre has plenty of more symmetries as the entire Louvre itself. Nature appears to apply that strategy too: the building blocks are often simple and highly symmetric but collectively chaotic patterns may arise. There is also a principle of humility here;

first, one should learn to fully master simple things prior to moving on to more complex situations. After set theory has been understood, and that might take a while, we start by developing number theory starting at the natural numbers followed by the integer, rational, real, complex and Clifford numbers. In a way, we go beyond our human limitations here, real numbers can never be written down exactly in decimal form; however, we can capture them by means of geometrical concepts such as π versus half the circumference of a circle of unit radius.

After having gotten numbers under the belt, we delve into the world of topology which is kind of a restricted form of set theory. In a way, we forget about all the exotic sets and we construct geometrical objects from pasting charts together. A natural concept which arises is the one of homology which classically does not fully capture the topology for a specific class of spaces of interest; however, we suggest a more general definition which obviously captures the whole topological space. In way, the dimension of the homology modules, being natural numbers, codify a space allowing one to discern a tire from a sphere. In a way, the idea behind homology is a universal one belonging to the field of category theory and is grounded in the notion of the boundary operator which maps a compact space of dimension n to its $n-1$ dimensional boundary. One simply observes that the boundary of the boundary is empty and therefore this linear operator is nil-potent of degree two. There are various interesting things to say about operators with such property and homology captures it all. The most simple kind of space is like a sheet of paper infinitely extended in all directions; this is an example of a two dimensional real linear space. The thing to observe is that in linear spaces one can always add two displacements and the addition does not depend upon the order. Likewise, a displacement can be shrunk, inverted and expanded at will; the summation of two displacements v, w is denoted by

$$v + w = w + v$$

whereas the re-scaling is noted by rv where r is a real number. On a sphere is also possible to subsequently add two different displacements although the sum depends upon the order taken; moreover taking any nonzero displacement, one can always expand it such that it effectively results in the zero displacement which is written down as $rv = 0$ for a certain $r \neq 0$. Hence, a sphere is not a linear space. On linear spaces, one disposes of a natural class of functions, the so called linear operators, which preserve the properties of addition and scalar multiplication. These functions provide approximations to general ones on certain scales and constitute one of the defining objects of so called quadratic surfaces of which the circle is an example in two dimensions. These topics concern the abstract foundations of mathematics; next, we proceed with studying more practical affairs which are used on daily basis in the physical sciences. We pay due attention to analytical geometry which constitutes the basis for general relativity. Abstraction returns however at the end of this book where completely new topics are discussed which deal with geometry without analysis; indeed, the ideas presented there are much more abstract and intrinsic as have a much wider field of application. The passoniated student can become a real mathematician there and work him or herself on expanding on the gems of the theory presented there.

Another interesting aspect is that plenty of aspects, which are studied first in a more common approach, are dealt with at a later stage in this book. The presentation however is much deeper and suggests plenty of extensions towards distinct other subjects such as non commutative number theory. An example of this is provided by the definition of the exponential function as well as Euler's formula which, by virtue of the material studied previously, can be immediately extended to the quaternions as well as the general Clifford numbers which are used on a daily basis in physics. It is my conviction that we must depart from the fetishism around fake research topics such as the Riemann Hypothesis in the sense that an intrinsic, qualitative approach is ultimately the best one. Let me finally salute the reader, that he may enjoy this book and spend lots of hours on assimilating and mastering this material which should prove to be a source of inspiration for further studies.

Chapter 2

Set theory and First-Order Logic

The following three chapters are by far the most difficult ones in this book; although they are written in a compact way, the reader should grind his teeth and spend a sufficient number of days to it before he or she is entirely convinced that this is the right way to proceed. Indeed, laying out the foundations needs to be done with caution and the student of this book will return several times to these first chapters while digesting to other ones; there is this most peculiar interplay between practice, imagination and reflection and all three need to be dealt with with a sense for measure and humor. Therefore, this chapter is my personal brand resulting from these three processes: as the educated critical mind will undoubtedly utter, the level of abstraction here is way beyond any education in mathematics. But likewise is this so for the imaginative and reflective parts; given that young adults have plenty of more imaginative intelligence as adults do, I opted for immensely increasing their abstraction skills while preserving the student's imagination so that the very best may come out of the student teacher interaction instead of the very worst. This is an attitude which requires work and dialogue which is one of the reasons why this book pisses on the contemporary lazy so called didactic presentation techniques using stupid limiting visual representations as well as repetitive exercises. The latter are on the level of Fields medal winning topics god damned, the Nobel prize in mathematics. So be prepared and isolate yourself in a room with enough sunlight and natural peace.

Set theory is a very abstract and difficult domain indeed and as is the case for any taste, it requires practice to distinguish a Rothschild Bordeaux from a mere Chateauneuf du Pape. Since this ought to be your first moment of discovering what it means to speak in a formal way, which is understood by everyone in almost the same way, it is mandatory to practice one's mind in the world of incredible detail which hides in every phrase. Indeed, lawyers and judges are very much aware of the game language is and in this regard the mathematician is a bit as the police officer who likes it dry and clear. There is no prior way of telling here who is the most sane of mind given that the poet hiding in the judge is most capable of society re-conversion. However, most judges would become madmen if it were not for the police bringing them back to the most hard and simple of realities from time to time. Hard logic, proof and observation; nevertheless, the judge has a place in the police man's world and the blue shirt can speak intelligibly about the former in his "sprache". He can even do this in a way as to predict what the judge is going to do in an overwhelming majority of circumstances and those where his speech is not applicable is beyond the proper phrasing capabilities of the judge. Even for the judge, those things become fuzzy but nevertheless acceptable. This book is a policeman's view on the judge, at least this is so at the beginning. Gradually we shall develop gems of the logic of the judge's way of dealing with things but nevertheless so from the point of view of the police officer. You have to learn to fly but not end up as Icarus denying that the sun is not only bright but also hot. On the other hand, the simple policeman should not oversimplify either: the utmost attention to detail is required.

In mathematics, it is often so that the definition of a concept which we will submerge to logical scrutiny starts from playing around with examples; however, as we shall teach to the student, the formulation of one's ideas is the most important step in order to reach the Valhalla of lofty results. When introducing a set, people think about a collection of elements. But do the elements exist in the very definite way before I see them? For example, we all think about a shopping bag as containing fixed items such as a toothbrush, toothpaste, a bread, some cheese and so on; so called elements. However, it may be completely unnecessary to do this, maybe there is something in the bag beyond my reach. For example half a liter of spoiled milk, dripping on the bread and ham. Set theory does not contain such details and just posits the milk is there irrespective of the fact whether I can use it still or not. Maybe it is in our advantage not to start with a bag and items but just with "entities" which I can share $\cap$ and unite $\cup$, operations which are called the intersection and union. We still have not said anything useful here about these two operations given that we have not yet assigned any properties. In everyday life, we have the impression that $A \cap B = B \cap A$ as well as $A \cup B = B \cup A$ both properties being referred to as the commutativity of the respective operations. This is not necessarily so in nature, it does matter for example when I poor coffee in first in a bowl and then hot water later on. In this case the coffee dissolves and raises upwards causing for a homogeneous mixture. If I were to do it the other way around the coffee would most likely keep on floating on the water. So this commutativity of the union is not obvious, it refers to the fact that items are hard objects and no particular law holds between them. They are independent as to speak; this stance of individualism is required in science, we would not learn anything from a holistic perspective. We have to subdivide and believe in holy freedom otherwise nothing can be said about the I and its relations to others. We also have that $(A \cap B) \cap C = A \cap (B \cap C)$ and likewise so for the union, a property which we call associativity of the respective operation. Now, we can talk! Denote with $A, B, C, \ldots$ so called sets; we have no idea yet what they are but we shall further specify some properties regarding the operations $\cap$ and $\cup$. The operations satisfy for sure $A \cap A = A \cup A = A$ and we demand the existence of a unique empty set $\emptyset$ such that

$$\begin{aligned} A \cap \emptyset &= \emptyset \\ A \cup \emptyset &= A \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \end{aligned}$$

where this last rule is the same as the de-Morgan rule in Boolean logic. Set theory at this level is equivalent to the rules of classical logic where the A denote truisms and $\emptyset$ is given by "false". Then $A \cap A = A$ reads as A and A are both true is the same as A is true. A or A is true, denoted by $A \cup A$ is the same as A is true. A and false is always false. Finally A is true and B or C is true is the same as A and B is true or A and C is true. So set theory is classical logic, it is a definite speech about truisms. We will later on think of devilish ways to escape this definite way of speaking about things which hinges upon many assumptions which could equally well be false. But as mentioned earlier, the most simple rules can allow for very complicated ones to arise by means of building. The old Greek always described elements or atoms as things which cannot be further subdivided; hence the following definitions. We say that A is a subset of B is and only if the intersection of A and B equals A which reads as $A \subseteq B \leftrightarrow A \cap B = A$. An atom $A \neq \emptyset$ is called a primitive set, that is, A has the property that if $B \subsetneq A$ then $B = A$. The reader checks the obvious statement that $A \cap C \neq \emptyset$ is a subset of A; this follows from associativity and commutativity of the intersection because $A \cap (A \cap C) = (A \cap A) \cap C = A \cap C$ and therefore, by definition $A = A \cap C \subseteq C$ in case A is an atom or primitive set. Indeed, we can only speak of subparts when the operation of intersection is priceless. This suggests that primitive sets are as elements of a set and to emphasize that distinction we denote $A = \{\widehat{A}\}$ where $\widehat{A}$ is interpreted as an element and the brackets denote the bag. We use the symbolic notation $\widehat{A} \in B$ as an equivalent to the more primitive statement $A \cap B = A$.

The reader notices that we have *defined* elements from the operations $\cap, \cup$ whereas normally the opposite happens. This is a much more human way of dealing with language in the sense that the limitations attached to our operations define our notion of reality. The old approach starts from divine knowledge which nobody possesses; in order to make logic dynamical and attached to physical processes in space time mathematicians have invented the notion of a Heyting algebra instead of a Boolean one. We shall not go that far in this book but the interested reader should comprehend very well how this definition is tied to the one of classical relativistic causality. Our point of view also allows for quantal rules as long as the de-Morgan rule is suitably deformed; we shall discuss such logic in this book and make even further extensions towards non-associative and non commutative cases. Extension of the material presented is left to the fantasy of the gifted reader. For example, an infinite straight line does not need to consist out of points, the latter being mere abstractions. Let us first investigate further implications of our rules before we move on to further limitation of the setting at hand. It is true that if $B \subseteq C$ then every element $\widehat{A}$ in B belongs to C. Indeed, $\widehat{A} \in B$ if and only if $A \cap B = A$ and therefore $A \cap C = (A \cap B) \cap C = A \cap (B \cap C) = A \cap B = A$ proving that $A \cap C = A$ and therefore $\widehat{A} \in C$. Differently, $\widehat{A} \in B$ if and only if $A \cap B = A$ which is equivalent to $(A \cap C) \cap B = A$ and therefore $A \cap C \neq \emptyset$ from which follows that $A \cap C = A$ because A is an atom. Hence, elements of subsets belong to the set itself. What about the intersection of two sets? First, we show that if $\widehat{A} \in B, C$ then $\widehat{A} \in B \cap C$: this holds because $A \cap (B \cap C) = (A \cap B) \cap C = A \cap C = A$ and therefore $\widehat{A} \in B \cap C$. The other way around, we have that if $\widehat{A} \in B \cap C$ then $\widehat{A} \in B, C$ because the intersection is a subset of both. Hence, the elements in the intersection are precisely those which are in both of them. What about the union? We show that if $\widehat{A} \in B \cup C$ then either $\widehat{A} \in B$ or $\widehat{A} \in C$ because $A = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ implying that at least one of them is non empty and equal to A due to atomic property of the latter. Reversely, one has that if $\widehat{A} \in B$ then it is an element of $B \cup C$ because $A \cap (B \cup C) = A \cup (A \cap C)$ which equals $A \cup A$ or $A \cup \emptyset$ due to atomic property of A. In both cases we have that $A \cap (B \cup C) = A$ because $A \cup \emptyset = A = A \cup A$. Therefore, the elements in the union are in correspondence to the elements of one of the sets.

As suggested previously this does not imply that sets are fully specified by their elements nor that elements exist in the first place. For example, assume that $\mathcal{S}$ consists of $\emptyset, \{1\}, \{1, 2\}$, then $\{1\}$ is an atom, but $\{1, 2\}$ does not merely consist out of atoms. Standard set theory makes the assumption that

$$B = \{\widehat{A} | \widehat{A} \in B\}$$

meaning that a set equals a collection of its elements. In this case, we have just proved that $\cap$ and $\cup$ coincide with the usual operations of intersection and union. The reader might think this is all a bit abstract and utter “well, can I just not assume this without all these rules?”. The simple answer is “no”; mathematicians are very scarce on their assumptions indeed! Why writing an extra sentence into the constitution when the latter is already a consequence of the former rules?! The next question one could pose then is “well on then, but how do you make up for all these theorems as well as the formal proofs?”. The simple answer is that the results have to be in your mind prior to making up the concepts! A proof is no more as a logical confirmation of a kind of naturalistic observation in a way. Henceforth, it is merely an exercise to verify that the concepts lead to the appropriate results. This applies in the case of set theory due to the existence of the natural concept of an atom being equivalent to an element.

These are by far not the only rules of set theory which we shall slowly expand upon by means of more complicated objects and operations. Let us now deviate a bit and reflect further upon the commutation and associative properties of the intersection as well as union. We imagined that a set can be thought of as items in a bag; however, in reality our bag is a phantom bag given that the operations of emptying and resorting do not matter in taking the intersection or union. This would lead to complications involving the order of operations leading to a non-commutative logic which we shall study later on in this book. A true Frenchman would expect such rule to emerge in a way from the simple ones and indeed this is the case. Another field where such a thing happens is Riemannian geometry which is a generalization of flat Euclidean geometry.

We define the natural numbers n by means of the sum operation $n = 1 + 1 + 1 + 1 + \ldots + 1$ by means of the following prescription:

$$\begin{aligned} 0 &= \{\emptyset\} \\ n + 1 &= \{n, \emptyset\}. \end{aligned}$$

Hence, $1 = \{\{\emptyset\}, \emptyset\}$, $2 = \{\{\{\emptyset\}, \emptyset\}, \emptyset\}$ etcetera; this is a partial dictionary made out the symbols $\emptyset, \{, \}$ which are part of any set theory. I have warned the reader that symbolic notation often is the most difficult part of set theory and the latter notation allows for a definition comprehensible by a computer albeit the latter uses binary representations. We define in the same way $n+m$ by means of the prescription $n+(m+1) = \{(n+m), \emptyset\}$ where $n + 0 = 0 + n = n$. The reader shows that $n + m = m + n$ for every natural number m which is true by definition for $m = 0$. Indeed, suppose it is true for $m = k$, then we show it holds for $m = k + 1$. Indeed, $n + (k + 1) = \{n + k, \emptyset\} = \{k + n, \emptyset\} = \{k + (1 + (n - 1)), \emptyset\} = \{(k + 1) + (n - 1), \emptyset\} = (k + 1) + n$ where, in the first step, we have used the definition of the natural numbers, in the second the assumption that $k + n = n + k$ and finally, in the third step, the associativity of $+$. We pose that $\mathbb{N}$ is the set of all natural numbers, something which defines a set theory by means of taking all subsets of $\mathbb{N}$.

The operation $+$ maps two natural numbers onto a natural number; it is associative, commutative and has 0 as a neutral element implying that $0 + n = n + 0 = n$. For any n, it is possible to define an inverse $-n$ satisfying $n + (-n) = 0 = (-n) + n$ something we denote by $n - n = 0$; $n + (-m) = n - m$ is a natural number $n > m$ and minus a natural number if $n < m$. The set of natural numbers taken together with their inverse is called the entire numbers and is universally denoted by $\mathbb{Z}$. $\mathbb{Z}, +$ is called a commutative group given that the operation $+$ is interior, associative, has a neutral element and inverse.

As previously stated, one starts by making a distinction between elements of a set and sets themselves; we departed from the concept of an empty set $\emptyset$, the intersection and union and therefrom we deducted the first three axioms of set theory. The approach taken here is somewhat more general as we defined an element as a primitive set. Zermelo-Frankel set theory has plenty of more assumptions which have to do with infinity culminating into the axiom of choice. A fifth axiom deals with taking set theoretical differences

$$B \setminus C = \{\widehat{A} | \widehat{A} \in B \wedge \widehat{A} \notin C \cap B\}$$

and we shall always assume the difference set to exist. In the field of geometry, it is not only possible to take the union of two lines or the intersection theirof but we can also take the so called Cartesian product, defining a two dimensional sheet. More in particular, given two sets B, C, we define the Cartesian product $B \times C$ as the *set* of all tuples (x, y) such that $x \in B$ and $y \in C$ giving a sixth axiom in $\mathcal{S}$ and henceforth is this last one closed with respect to $\times$ from which holds

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

and

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

The existence of Cartesian products allows us to define relations where a *relationship* R between sets B and C constitutes a subset of $B \times C$. In case $B = C$ we can demand plenty of criteria. With the notation xRy we intend to say that x has a relation of type R to y if and only if $(x, y) \in R$; we call R reflexive if xRx for all $x \in B$, symmetric if xRy implies that yRx for all $x, y \in B$ and finally transitive if xRy and yRz imply that xRz. A reflexive, anti-symmetric, transitive relation is called to be a partial order and is noted by $\prec$ or $\leq$. A reflexive, symmetric and transitive relation is called an equivalence relation and is usually denoted by $\equiv$. One should think of an equivalence relation as a generalization of the equality sign given that it concerns objects with similar properties. One should prove that an equivalence relationship defined on a set A pulverizes it in equivalence classes $\overline{x}$ where

$$\overline{x} = \{y \in A | x \equiv y\}.$$

The reader verifies that $\overline{x} = \overline{y}$ if and only if $x \equiv y$ and therefore the intersection $\overline{x} \cap \overline{y} = \emptyset$ if they are not equivalent. A partial order is a generalization of a total order such as "Jon is larger as Elsa". A partial order allows for two objects to be not related at all.

We have defined the natural numbers by means of the operation $+$; $\mathbb{N}$ has a natural *total* order $\leq$ defined by $n \leq n$ and $n \leq n+1$ and one takes the *transitive closure* therefrom which is defined by imposing transitivity on the existing relationship. This can be compared with lacing a chain. From the natural numbers we constructed the entire numbers $\mathbb{Z}$ and the definition of $\leq$ has a natural extension towards $\mathbb{Z}$. We construct now the rational numbers starting from $\mathbb{Z} \times \mathbb{N}_0$ and imposing the equivalence relationship $(m, n) \equiv (m', n')$ if and only if there exist a $k, l \in \mathbb{N}_0$ such that $km = lm'$, $kn = ln'$ where $\mathbb{N}_0 = \mathbb{N} \setminus \{0\}$. The *rational* numbers are henceforth defined as the equivalence classes defined by means of this *equivalence* relation.

The six axioms discussed are by far the most important ones of set theory and allow one to construct the rational numbers; the remaining two axioms concern infinity and are in general added to generalize aspects of the rational numbers to the real ones. We shall be very cautious here with the kind of infinity we shall allow for culminating into a thorough discussion of the axiom of choice. In fact, we shall demonstrate that this highly contested axiom is wrongly chosen in the sense that it contradicts the existence of the real numbers. The kind of mathematics required here is not at its place in this chapter; however, it is presented here for matters of completeness and the reader is invited to thoroughly check the details later on. The seventh axiom allows one to define subsets of sets: given a set D, the power set 2^D of all nontrivial subsets of D is a set and belongs to $\mathcal{S}$. This axiom leads to the construction of the ordinary numbers by Cantor. The definition the Cartesian product is extended to so called "index" sets something which requires a partial order $\prec$. An index set I is a set equipped with a partial order $\prec$ such that for any $x, y \in I$ there exists a $z \in I$ such that $x, y \prec z$. This condition is required and sufficient if we want to take unique limits such any reader should check. If this is not valid, then several sub limits could exist; hence, we denote by

$$\times_{i \in I} A_i = \{(x_i)_{i \in I} | x_i \in A_i\}$$

where all I-tuples are partially ordered by $\prec$. Finally, we have the so called axiom of choice which can be formulated as follows: given sets A_i, $i \in I$, with I an index set, then the Cartesian product is nonempty.

Another, but equivalent formulation is that there exists a function f from I to $\cup_{i \in I} A_i$ such that $f(i) \in A_i$. So, one can constitute a set by drawing an element from each set. This axiom has plenty of ramifications in some parts of mathematics, in particular functional analysis although some mathematicians have refuted it because some results appear too strong and give the transfinite an equal status to the finite situation. I have stated it already a few times: mathematics as such is not open to proof; it is a language and we have to make some grammatical choices. The reader has to reflect about these rules and be conscious of the fact that commutativity, associativity as well as the formation of a power set are the most simple of all symmetrical rules. An example which does not obey these rules has been constructed from this ideal situation; for example, we shall study later on non commutative or associative operations and construct those from the simple commutative situation. This leads to non commutative groups, quantum groups etcetera. This reminds us about the Egyptian architectural art followed by the Roman and French symmetrical ones: super simple, magnificent and logical.

To clarify, the axiom of choice supports the idea that the Cartesian product is non empty whereas the Cartesian product axiom presupposes that the product is a set. We now show that this axiom leads to the most bizarre of situations; to this purpose, consider two rotations a, b with as angle $r2\pi$ where r is an irrational number, around the x and z axes respectively. One considers the free group F_2 constructed from a, b which can be split into five parts

$$S(a), S(a^{-1}), S(b), S(b^{-1}), e$$

where $S(a)$ contains all irreducible words starting with a. Clearly, one obtains that $S(a) \sim S(b)$ from a geometrical perspective applying the rotation ab^{-1}. The axiom of choice allows for the construction M of a set which contains exactly one representant from any F_2 orbit on the sphere S^2. The construction goes as follows: consider the set of all equivalence classes $\widetilde{M}$ of S^2 under F_2 and denote by $p : S^2 \to \widetilde{M}$ the associated projection. If one equips $\widetilde{M}$ with a trivial partial order $\prec$ by picking one element of $\widetilde{M}$ and putting it on top of all others which remain unrelated, then one arrives at an index set $(\widetilde{M}, \prec)$ and the axiom of choice is applied to $\times_{m \in \widetilde{M}} p^{-1}(m)$ giving rise to an element F. Consider the subsets

$$A = S(a)M, B = S(a^{-1})M, C = S(b)M, D = S(b^{-1})M, M$$

and observe that $bD = A \cup B \cup D \cup M$. The reader notices that $b^n D \subseteq b^{n+m} D$ for $n, m > 0$ and subsequently $\lim_{n \to \infty} b^n D = S^2$ meaning that one can define proper subsets which grow under a rotation and eventually cover up the entire S^2. The reader may enjoy finding out that similar observations lead to a decomposition of the sphere giving rise to two identical spheres (all three of the same radius).

Such paradoxes are at the heart of set theory and have far reaching consequences.

Finally, we comment upon a novel extension towards preposition logic and henceforth a new and much better alternative for the Peano axioms. The key idea is to refute existentialism; that is the French vocabulary of there exists or for all, nobody really knows what they mean and they are always contextual, meaning depending on some set of objects. For example, the easiest one is

$$\forall x \in A$$

which simply is replaced by

$$x \in A \to .$$

To define the quantor $\exists x \in A$, we must first appropriately define the negation $\neg^C$ with respect to some set C. It reads

$$\neg^C (x \in A) \leftrightarrow x \in C \setminus A.$$

Until so far, we have played a dirty game by defining

$$C \setminus A = \{x | x \in C, x \notin A\}$$

hence using the existential notion of $\notin$ which is precisely what we wanted to avoid. Much better is to define $C \setminus A$ as the unique set B satisfying

$$B \cap A = \emptyset, \; B \cup (A \cap C) = C.$$

Now, we define the logical disjunction:

$$(x \in A) \vee_o^C (x \in B) = \neg^C(((x \in A) \wedge (x \in B)) \vee (\neg^C(x \in A) \wedge \neg^C(x \in B)))$$

where $A \cup B \subsetneq C$ which is noting but the statement

$$x \in A\Delta B$$

where

$$A\Delta B = (A \cup B) \setminus (A \cap B).$$

By transfinity,

$$\exists x \in A \leftrightarrow \vee^C_{o,\emptyset \neq B \subsetneq A}(x \in B)$$

where again $A \subsetneq B$. The reader may verify that

$$\exists! x \in A \leftrightarrow \vee^C_{o,y \in A}(x = y).$$

As such, Russel's Paradox and all other well known paradoxes of preposition logic disappear.

Chapter 3

Advanced Number Systems

This chapter is already somewhat more practical as the previous one; abstract rules will now be applied to define and make specific computations. We commence with representations of the entire or rational number and extend those later on, by means of a countable completion procedure to the Archimedic field of real numbers. After this, we study the complex number system as well as the real Quaternions in order to finally introduce the Clifford algebra's. Applications of these number systems are plenty: the complex numbers have been of vital importance in the construction of quantum theory and extensions of the latter physical theory regarding the quaternions have been made also. We have up till now studied $\mathbb{Z}, +$ and prior to introducing higher number systems it is convenient to introduce some formal language. $\mathbb{Z}, +$ has the property that

- $+$ is internal implying that the sum of two entire numbers is again an entire number,
- $+$ is associative meaning that $(n+m)+k = n+(m+k)$, in other words the sum of a series of entire numbers only depends upon the numbers themselves and not how you split them into pairs,
- there exists a neutral element 0 such that $0+n = n+0 = n$,
- given the neutral element 0, for any n there exists a unique inverse $-n$ such that $n+(-n) = (-n)+n = 0$,
- the sum operation is commutative meaning that $n+m = m+n$. 77

More in general, a set G with operation $\star : G \times G \to G$ obeying those five properties is called a commutative group. One has to understand very well that those rules have been presented in the order of importance: the opposite or inverse cannot be defined without the neutral element as well as the internal character of the sum. It is easy to drop associativity as well as the existence of a unique inverse: commutativity is the easiest property to give up on and we shall study abundant computational tools later on. The entire, rational, real as well as complex numbers all obey these rules and we shall proceed by the definition of the multiplication. We declare by fiat that $n.1 = 1.n = n$ as well as the de Morgan rule $n.(m+1) = n.m + n = (m+1).n$; here from follows that

$$n.(m+k) = n.m + n.k = (m+k).n$$

for all $n, m, k \in \mathbb{N}$. Clearly, one obtains

$$n.(m+k) = n.((m+k-1)+1) = n.(m+k-1)+n = n.(m+k-2)+n.2 = \ldots = n.m + n.k$$

by repeatedly applying this elementary rule. One shows now that the same holds for all $n, m, k \in \mathbb{Z}$ by means of defining 0 to be the absorption $0.n = n.0 = 0$. 1 is by definition the identity element regarding the multiplication but not any element different from 1 has an inverse in $\mathbb{Z}$ given that it would need to satisfy $m.n = n.m = 1$. The multiplication is henceforth internal, associative, commutative and has an identity element. $\mathbb{Z}, +, .$ with those properties is called a ring; more precisely, we define a set G equipped with two operations $+, .$ where $G, +$ is a commutative group and . internal, associative and with identity 1 a ring if moreover

$$g.(u+v) = g.u + g.v,\ (u+v).g = u.g + v.g$$

a property which we call the distribuvity of the product regarding the sum. As stressed previously, this rule is precisely the same as the de-Morgan rule from classical logic and the reader is requested to construct a mapping given that false equals 0 and true is given by 1 with the supplementary statement that $1+1=0$ which is the same as taking the natural numbers modulo 2. The relationship mod_2 maps a natural number to its remainder after division by two. Roughly speaking, or coincides with plus and the conjunction by the multiplication. De reader appreciates that these rules are of a universal nature given that the large majority of known number systems satisfies these properties. Prior to introducing more general number systems, it is good to reflect on some nature "decimal" representations of them; the Arabs introduced so called powers n^m where n, m are two natural numbers. By definition holds that $n^0 = 1$ and $n^{m+1} = n.n^m$ so $n^1 = n, n^2 = n.n, n^3 = n.n.n$ and so on. In general, this reads as n multiplied m times with itself. It is fairly obvious that any natural number m can be uniquely decomposed as

$$m = \sum_k c_k.n^k$$

where c_k is a natural number between 0 and $n-1$. The proof of this result follows from $m = k.n + r$ with r between 0 and $n-1$ which simply means that any number is between subsequent multiples n. Repeated application of this rule gives

$$m = k_1.n + c_0 = (k_2.n + c_1).n + c_0 = k_2.n^2 + c_1.n + c_0 = \ldots = \sum_{k=0}^{l} c_k.n^k$$

where the last sum is finite because at a given moment $k_{l+1} = 0$. Hence, a number can be represented as $c_l c_{l-1} \ldots c_1 c_0$ and this code depends upon the base number n. In case the base number equals 2 the number system is a binary one and used in electronic devices such as microwaves and computers. Thus, one has to make the following exercises: show that the number $3 = 2+1$ has a binary code of 11 and the number $7 = 2^2 + 2 + 1$ as 111. The most popular base number is given by ten, perhaps because we have ten fingers or toes; the rest numbers are henceforth $0 = \emptyset, 1 = 1, 2 = 1+1, 3 = 1+1+1, 4, 5, 6, 7, 8, 9$. Therefore, we note numbers such as 923 and so on; now, why is such a notation extremely useful regarding the notions of addition and multiplication? The reason is that $0 \leq c_k + d_k \leq 2n-2$ as well as $0 < c_k c_l < n^2$ leading to the multiplication and substraction rules thought in the elementary school; a convenient notion here is the concept modulo " mod " where "$n \mod m$" is the leftover of n by means of division through m. This so far for the addition and multiplication of natural numbers; rules which can be trivially extended towards the entire number system and the reader is invited to make some exercises hereon. These are typically things which people do in elementary school; we finish this section with some novel concepts. A natural number $n > 1$ is called a prime number if and only if n and 1 are the only few divisors meaning that in case $n \mod q = 0$ and $0 \leq q \leq n$ then it holds that $q = 1, n$. One verifies now that $2, 3, 5, 7, 11, 13, \ldots$ constitute the first 6 prime numbers. The first exercise to make is proving that every number can be written as a unique product of prime numbers, something we call the prime number decomposition. For example, the decomposition of $6, 12, 21$ into prime numbers are $6 = 2.3, 12 = 2^2.3, 21 = 3.7$; calculate the prime decompositions of $37, 41, 56$.

We now proceed by the definition of the rational numbers starting from the Cartesian product $\mathbb{Z} \times \mathbb{N}_0$ equipped with an equivalence relationship: that is, the couples (m, n) and (m', n') are considered to be equivalent if and only if there exist nonzero natural numbers k, l such that $km' = lm, kn' = ln$ holds. This is not a convenient way to think about this relationship and historically the fraction

$$\frac{m}{n}$$

has been introduced. The equivalence law is then equivalent to a cancellation rule for common factors above and below the bar. More precisely,

$$\frac{km}{kn} = \frac{m}{n}$$

and the addition respectively multiplication laws are defined by

$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 n_2 + m_2 n_1}{n_1 n_2}$$

$$\frac{m_1}{n_1}\frac{m_2}{n_2} = \frac{m_1 m_2}{n_1 n_2}.$$

The identity element for the addition is given by $\frac{0}{n}$ where n can be arbitrary and the identity element for the multiplication is given by $\frac{n}{n}$ where $n \in \mathbb{N}_0$. The reader verifies, as a reassuring exercise, that the addition and multiplication laws are associative and commutative and that the distributivity rule is satisfied as well. So, $\mathbb{Q}, +, .$ is a ring; the inverse for the addition is clearly given by $\frac{-m}{n}$ and one has to find an inverse for the multiplication for any non zero element. With these additional properties, we call $\mathbb{Q}, +, .$ a number field; as far go the mathematical properties of the rational numbers. Now, it must be clear to the reader that the entire numbers $\mathbb{Z}$ can be represented as a rational number by $m := \frac{m}{1}$ and that the different definitions of addition and multiplication are preserved or commensurable. More precisely,

$$m + n := \frac{m+n}{1} = \frac{m}{1} + \frac{n}{1}, \; mn := \frac{mn}{1} = \frac{m}{1}\frac{n}{1}$$

and we summarize this property by stating that the mapping

$$\mathbb{Z} \to \mathbb{Q} : m \to \frac{m}{1}$$

is an algebraic homomorphism. The reader shows now that $\mathbb{Q}$ is the smallest field encompassing the integer numbers (hint: define inverses with respect to the multiplication). A second important property regards the extension of the order relationship $\leq$; we utter that

$$\frac{m}{n} \leq \frac{m'}{n'}$$

if and only if $mn' \leq m'n$. One controls that this definition is independent of the representants of the equivalence class meaning that if $\frac{r}{s} = \frac{m}{n}$ and $\frac{r'}{s'} = \frac{m'}{n'}$ then $rs' \leq r's$ if and only if $mn' \leq m'n$. As a more difficult exercise, one shows that it is impossible to enumerate the rational numbers in a way preserving the order; use for this purpose the property that every two sequential rational numbers have a rational midpoint.

This last property brings us to the construction of the real numbers $\mathbb{R}$; plenty of distinct methods exist for introducing them, for example by taking a metric closure of $\mathbb{Q}$ in the natural metric. Given that topology and metrics are merely studied in the subsequent chapter, we shall proceed in an alternative fashion; we simply employ the order relationship to get a closure. It is easy to see that the rational numbers contain holes with respect to the positive inverse of the square operation; for example, $\sqrt{2}$ is defined to be the unique positive number such that $\sqrt{2}^2 = 2$. We show that such a number, if we demand it to exist, is not rational. Suppose that $\sqrt{2} = \frac{m}{n}$ then $2n^2 = m^2$ which is impossible given that the prime number 2 occurs an odd number of times in the prime decomposition of the left hand side whereas the right hand side contains an even number of factors. Henceforth $\sqrt{2}$ does not belong to the rational numbers and therefore it contains "holes". Note that the prime numbers are important regarding the multiplication which is the mirror property of the intersection regarding sets. Indeed, prime numbers are the same as primitive sets in this regard and it is therefore not a surprise that they are important. So, in a way, prime numbers are the "elements" of the natural ones. The question now is how one constructs a number such as $\sqrt{2}$! To rehearse our logic so far: (a) the natural numbers are constructed by means of the addition given one elementary unit (b) the integer numbers follow by taking inverses regarding the sum (c) the rational number follow from the entire ones by taking inverses with respect to the multiplication law and finally (d) the real numbers follow from a completeness property.

The completion is defined by means of a Dedekind procedure: take an increasing sequence $q_n \leq q_{n+1}$ of rational numbers q_n which are all smaller as a rational number q. We define two such sequences $(q_n)_{n \in \mathbb{N}}$, $(p_n)_{n \in \mathbb{N}}$ to be equivalent if and only if for any p_n there exists a q_m such that $q_m \geq p_n$ and vice versa. The reader should verify that this relationship between increasing sequences is indeed an equivalence; it is likewise possible to work with decreasing sequences bounded from below. $\mathbb{R}, +, .$ is then defined as the set of these equivalence classes equipped with the operations $+$ and $.$ making $\mathbb{R}, +, .$ into a field. We shall construct by means of an example a sequence for the real number $\sqrt{2}$:

$$\sqrt{2} = \sqrt{4-2} = 2\sqrt{1 - \frac{1}{2}}$$

and

$$\sqrt{1+z} = 1 + \frac{1}{2}z + \sum_{n=2}^{\infty} (-1)^{n+1} \frac{1.3 \ldots (2n-3)}{2^n n!} z^n$$

and therefore $\sqrt{2}$ is approximated by means of a descending sequence

$$q_m = 1 - \frac{1}{4} - \sum_{n=2}^{m} \frac{1.3 \ldots (2n-3)}{4^n n!}$$

of rational numbers. Henceforth, the real numbers constitute a complete and totally ordered field satisfying the Archimedean property

$$ra \leq rb$$

if and only if $a \leq b$ provided $r > 0$. Number theory can be further extended by means of the following observation: a polynomial is a real function in terms of one variable x of the form

$$P(x) = \sum_{k=0}^{n} a_k x^k$$

where $a_k \in \mathbb{R}$; a root of P is a real number c such that $P(c) = 0$. In case c is a root of $P(x)$ one can write $P(x)$ as $P(x) = (x - c)Q(x)$ with $Q(x) = \sum_{k=0}^{n-1} b_k x^k$. Show it! In case P consists out of n zero's one can write

$$P(x) = a_n(x - c_1)\ldots(x - c_n).$$

In this case, we call $P(x)$ totally factorisable and $(x-c)$ a factor. One can easily see that not every polynomial of this form can be factorized in this way: for example

$$P(x) = x^2 + 1$$

is always strictly positive and has no zero's. To factorize it, one adds the number i such that $i^2 = -1$. In that case, we have

$$P(x) = (x - i)(x + i)$$

showing one has to extend the real numbers with i; in other words, define

$$z = a + bi$$

with $a, b \in \mathbb{R}$ and

$$\mathbb{C} = \{z | z = a + bi; a, b \in \mathbb{R}\}$$

the set of complex numbers. Consider subsequently the polynomials $P(z) = \sum_{k=0}^{n} a_k z^k$ with $a_k \in \mathbb{C}$ then one can prove that $P(z)$ can be factorized or $P(z)$ has n zero's. One remarks that it is sufficient that $P(z)$ has at least one zero c because then $P(z) = (x - c)Q(z)$ where $Q(z)$ is of degree $n - 1$ which has a zero again and so forth. A proof of the existence of at least one zero necessitates higher mathematics, more in particular complex analysis and we shall return to this issue further on in chapter 13. We equip $\mathbb{C}$ with an addition as well as multiplication defined as:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

and

$$(a + bi)(c + di) = (ac - bd) + (bc + ad)i.$$

One shows that $\mathbb{C}, +, .$ is a ring and one defines the complex conjugate as

$$\overline{z} = a - bi$$

implying

$$\overline{z}z = z\overline{z} = a^2 + b^2$$

which only vanishes in case $a = b = 0$. Therefore, the number $\frac{\overline{z}}{z\overline{z}}$ is the inverse for z regarding the multiplication and henceforth $\mathbb{C}, +, .$ is a field. The field is not Archimedean provided that the natural order defined by $a + bi \leq c + di$ if and only if $a < c$ or $a = c$ and $b \leq d$ does not satisfy the property mentioned before. In such case, one has $1 + 5i > 0$ but $(1 + 5i)^2 = -24 + 10i < 0$ which is in contradiction to the Archimedean property. We delve deeper into the world of complex numbers later on but note for now that

$$\overline{z + z'} = \overline{z} + \overline{z'}$$

en

$$\overline{zz'} = \overline{z}\,\overline{z'}$$

whose proof is left as an easy exercise to the reader. As an application we shall calculate zero's of $P(x)$ for P having degree one or two. For a function of degree one $ax + b$ the zero point is given by $x = -\frac{b}{a}$ whereas for a polynomial of second degree defined by

$$ax^2 + bx + c$$

we proceed by rewriting it as

$$a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c$$

which vanishes if and only if

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

In case $b^2 - 4ac \geq 0$ then

$$x_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

whereas in the other case

$$x_{\pm} = \frac{-b \pm i\sqrt{4ac - b^2}}{2a}$$

which proves explicitly that polynomials of degree two are factorisable over $\mathbb{C}$.

Now, one might suspect that the most interesting number systems have been studied already; however, there exists an interesting class of generalizations of the complex numbers which are special cases of so called Kac-Moody algebra's. The most simple case is provided by the real quaternions $\mathbb{RQ}$. This number system has a close alliance to geometry in three dimensional space or four dimensional spacetime. For now, we restrict ourselves by studying the algebraic properties and delay the discussion of the geometric aspects. $\mathbb{RQ}$ is generated by two imaginary units i, j such that $i^2 = j^2 = -1$ and $ij + ji = 0, ij = k$; henceforth, $k^2 = ijij = -i^2j^2 = -1$ and $jk = i, ki = j$. A general quaternion q is henceforth of the shape $q = a + bi + cj + dk$ where $a, b, c, d \in \mathbb{R}$. We define the quaternion conjugate as

$$\overline{q} = a - bi - cj - dk$$

and as such is $q\overline{q} = \overline{q}q = a^2 + b^2 + c^2 + d^2$ which is positive definite. The inverse of $q \neq 0$ is therefore given by $\frac{\overline{q}}{q\overline{q}}$ and therefore $\mathbb{RQ}, +, .$ is a non commutative ring for which every nonzero element has an inverse. Finally, we get to the last division algebra over the real numbers meaning that each nonzero element has an inverse regarding the addition and multiplication. We work again though geometrization by means of a positive definite scalar product to obtain the required inverse. The so called octonions are obtained by means of a doubling of the quaternions, just like the quaternions arose from the complex numbers by adding an anti-commuting imaginary unit. This is done by adding an imaginary unit l, satisfying $l^2 = -1$; in order for the definition of the inverse, by means of geometrization through an involution, to work out it is mandatory that

$$il + li = jl + lj = kl + lk = 0.$$

The reader immediately notices that the last condition violates associativity given that

$$kl = (ij)l = i(jl) = -i(lj) = -(il)j = (li)j = l(ij) = lk$$

which is a contradiction. Furthermore, taking products

$$li,\, lj\, lk$$

we need that

$$i(li) = l = -(li)i,\; i(lj) = -(lj)i = -lk,\; i(lk) = -(lk)i = lj$$

and the reader notices that associativity has been violated in the second and third series of equalities. Likewise, do we obtain similar results for j, k. Finally,

$$(li)^2 = -1,\; (li)(lj) = k = -(lj)(li),\; (li)(lk) = -j = -(lk)(li),\; (lj)(lk) = i = -(lk)(lj)$$

and

$$\overline{(li)} = \bar{i}\bar{l} = il = -li.$$

Henceforth, any quaternion q is written as

$$q = a + bi + cj + dk + el + f(li) + g(lj) + h(lk)$$

and

$$\overline{q} = a - bi - cj - dk - el - f(li) - g(lj) - h(lk)$$

so that

$$q\overline{q} = \overline{q}q = a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2 > 0$$

proving the claim that the non-commutative and non -associative octonions constitute a division algebra.

There is a natural extension of this number system towards the Clifford algebra's; those are generated by elements e_i for which holds that

$$e_i e_j + e_j e_i = \alpha_{ij}$$

with $\alpha_{ij} = 0$ if $i \neq j$ and $\alpha_{ii} = \pm 2$. We denote this algebra with $\text{Cl}(p, q)$ where p is the number of plusses and q the number of minusses. The quaternions are then given by $\text{Cl}(0, 2)$; one can show that with the exception of $\text{Cl}(0, 1)$ and $\text{Cl}(0, 2)$, the complex numbers and real quaternions, that in any of these rings a nonzero element can be found which has no inverse for the multiplication. This complicates matters given that $\text{Cl}(1, 3)$ is frequently used in theoretical physics. To wrap up the discussion, one shows that any Clifford number can be written as

$$q = a + \sum_{n=1}^{p+q} a_{j_1 \ldots j_n} e_{j_1} \ldots e_{j_n}$$

where $j_k \neq j_l$ if $k \neq l$. We shall later on study more properties of the Clifford numbers but this suffices for now. Remark that the Clifford numbers are non commutative regarding the multiplication and therefore require more than one real dimension to represent them geometrically. Making geometrical algebras is a very important branch of mathematics and we shall return to analysis with Clifford numbers later on in this book.

So far a first introduction to number theory: everything can appear to be rather abstract in nature but the reader will familiarize him or herself with those concepts in the future. Finally a bit of notation: we have already stated that the inverse of q regarding the sum is given by $-q$ whereas for the inverse this equals q^{-1}. If one did not have any previous experience with this calculus, it is necessary to show that

- $(\frac{5}{4})^{-1} = \frac{4}{5}$,
- $(1 + i)^{-1} = \frac{1-i}{2}$,
- $2^3 = 8$,
- $(\frac{1}{3})^2 = \frac{1}{9}$,
- $\frac{1}{i} = -i$,
- $(1 + 2i)^2 = -3 + 4i$.

Exercises regarding the Clifford numbers and octonions.

- Prove that the Clifford monomials $e_{i_1} \ldots e_{i_k}$ with $i_r \neq i_s$ for $r \neq s$ and k even constitute a subalgebra of the Clifford algebra.
- Consider the Clifford numbers $\beta_{ij} = e_i e_j$ with $i \neq j$ and define the commutator $[\beta_{ij}, \beta_{kl}] = \beta_{ij}\beta_{kl} - \beta_{kl}\beta_{ij}$. Show that the commutator is again of the form $\sum_{r \in \{i,j\}, s \in \{k,l\}} \pm \alpha_{r's'} \beta_{rs}$ where $r' = i$ if $r = j$ and $r' = j$ in case $r = i$.
- The β_{ij} constitute therefore a Lie algebra defined by means of the α_{ij} symbols.

- Classify all real Clifford algebra's for which every non zero element has a multiplicative inverse and such that no null divisors occur (there are three of them; the real numbers, complex numbers and quaternions). A null divisor is determined by a pair of numbers $a, b \neq 0$ such that $ab = 0$. Prove that the octonions also obey those facts but do not constitute a Clifford algebra since it is not associative.

- Determine the Lie algebra generated by the (inner automorphisms of the) octonions, called G_2, and hold in regard that a Lie algebra must obey that

$$[o, [p, q]] + [p, [q, o]] + [q, [o, p]] = 0.$$

Chapter 4

Topology

One has to contemplate about topology as a refinement of set theory; it is to say, we limit ourselves to special sets being the so called open sets. In nature, an open set is an abstraction, an imaginary concept which has no real existence. An open surrounding has to be thought of as a voluminous object: for example, a straight line segment is the set of all real numbers between two extremal values denoted by $(a, b) = \{x | a < x < b\}$ with a natural length of $b - a$. A point is an example of a closed set and has vanishing volume or length. We now consider some properties regarding the set theoretical operations on the open segments (a, b): the union of two open segments is declared open by fiat whereas the intersection of two open segments is an open segment anew. Note that the union of open segments can be written as a disjoint union. Given a set D, we call a set $\tau(D)$ of subsets of D a topology if and only if

- $\emptyset \in \tau$,
- $A, B \in \tau$ implies that $A \cap B \in \tau$,
- $A_i \in \tau$ implies that $\cup_{i \in I} A_i \in \tau$ for every second countable index set I.

I stress again that this definition depends upon the commutativity as well as associativity of the intersection and union; it is possible to define a non-associative and non commutative topology by means of deformations. We shall study this from the viewpoint of logic further on and the reader may repeat these constructions almost ad verbatim here. In this chapter, we start pedestrian by studying the classical case where taking the union can be seen as putting landscape maps together; typically such charts overlap and all we demand is that the intersection of two charts is again a chart and that arbitrary many of them can be put together. There exist special subsets $E \subseteq D$ such that it is

- *closed* if and only if $E^c := D \setminus E \in \tau(D)$,
- *compact* if and only if for any coverage by means of open sets O_α of E there exists a finite sub coverage $O_i; i = 1 \ldots n$ such that $E \subseteq \cup_{i=1}^{n} O_i$.

Henceforth, the compact sets are those which can always be covered by means of a finite sub cover such as for example a globe: irrespectful of how small you make the charts, the globe is covered by a finite number of them. Given a point $p \in D$, we say O is an open environment of p if and only if $p \in O$. Given a point p, a basis of open environments is given by a countable collection of open neighborhoods O_i of p, such that for any open V encompassing p it holds that there exists an index i such that $O_i \subseteq V$. One could moreover demand that $O_{i+1} \subseteq O_i$ by taking intersections but this is not mandatory however. Regarding the closed sets X, Y one has to verify the following truisms: (a) $\emptyset, D$ are closed (b) $X \cup Y$ is closed (c) $\cap_{i \in I} X_i$ is closed if and only if all X_i are as such. Sets such as $\emptyset, D$ which are open and closed at the same time are dubbed cloped. Given $B \subseteq D$, the intersection of all closed sets X encompassing B is closed and called the closure of B which we denote as $\overline{B}$. The closure of a set is therefore the smallest closed set encompassing the latter itself. In other words, one adds elements or points which are limits of elements in B. More concretely, we call x a limit point of a sequence $(x_i)_{i \in I}$ if and only if for every open neighborhood $\mathcal{O}$ of x holds that there exists an index j such that $\forall j \prec i$ it holds that $x_i \in \mathcal{O}$. Now one shows that, using the properties of an index set, if y were another limit point then the open neighborhoods of x and y coincide. This motivates the following definition: a topology is Hausdorff if and only if all disjunct points x and y have open neighborhoods each with empty mutual intersection. It is to say that $x \in \mathcal{O}, y \in \mathcal{V}$ and $\mathcal{O} \cap \mathcal{V} = \emptyset$. For Hausdorff topologies holds that the limit point of a sequence is unique. We now prove the following result for topologies with a countable basis: a set is closed if and only if it contains all its limit points. Indeed, suppose that B is closed, and $(x_i)_{i \in I}$ is a sequence in B with limit point $x \in D$, then it holds that $x \in B$ otherwise one can find an open neighborhood B^c of x which is disjoint with $(x_i)_{i \in I}$, something which contradicts the definition of a limit point. Reversely, suppose that any limit point of B belongs to B, then we show that B is closed; suppose it is not, then we find an $x \in \overline{B} \setminus B$ such that for any basis-open neighborhood $\mathcal{O}_n$ of x we find an element $x_n \in B \cap \mathcal{O}_n$ and as such it holds that x is a limit point of $(x_n)_{n \in \mathbb{N}} \in B$ and henceforth, by assumption, an element of B which leads to a logical contradiction. Later on, we give an example of a compact set in a non-Hausdorff topology with a sequence containing no sub-sequence with a limit point (in case you want to think about this; find an example in an infinite number of dimensions). We shall study further characteristics of compactness in the so called metric topologies, which are determined by a distance function d.

So far, the treatment of topology appears to be very abstract and not very useful at all, one can think of any topology one wants to and indeed, all subsets of the real number system for example constitute a topology called the discrete topology. Indeed, all sets are cloped there which suggests a huge triviality. The physical reality we live in appears by very close inspection much more peculiar given that we speak about distance functions and spheres such as for example the circle with radius of 10 kilometer around Brussels measured from the Grand Place in bird flight. On earth this procedure only goes wrong when one traverses half of the circumference; one step further in the same direction would replace that journey by a different one where one originally departs in the opposite direction. Therefore, at large distances, one can expect problems of this global nature and in quantum geometry, one suspects those issues can occur at small distances too. Typical scales here are much smaller as those of an atom. By definition, a distance function $d : X \times X \to \mathbb{R}^+$ defined on a set X satisfies

- $d(x, y) = 0$ if and only if $x = y$,
- $d(x, y) = d(y, x)$ for each $x, y \in X$,
- $d(x, z) \leq d(x, y) + d(y, z)$ the so called triangle inequality.

A distance function defines a so called Hausdorff topology with countable basis by means of the open balls

$$B(x, \epsilon) = \{z | d(x, z) < \epsilon\}$$

giving rise to a countable basis defined by $B(x, \frac{1}{n})$ where $n \in \mathbb{N}_0$. Two points x, y separated by means of a distance $d(x, y) > 2\epsilon$ can be surrounded by means of two disjoint balls $B(x, \epsilon), B(y, \epsilon)$ respectively. This representation of affairs is still a bit abstract given that one wants to measure angles as well contemplate a notion of orthogonality which is not so simple in this formalism. In other words, we require further specialization extending beyond the distance function only. Nevertheless, one can prove plenty of theorems in this primitive language relying solely upon those three axioms. A generalization consists in specifying that the distance function has a local origin; it is to say that the distance between two points can be chopped into arbitrarily small pieces. This leads to the notion of a path metric: d is a path metric if and only if the property holds that for any two points x, y there exists a z such that

$$d(x, z) = d(y, z) = \frac{d(x, y)}{2}.$$

In other words, every two points define at least one midpoint. We shall later on give a better representation of those facts.

We will study now an equivalence relationship between two topological spaces; in other words, when are two topological spaces the same? To determine that, we shall study topological mappings between two topological spaces X, Y. A mapping $f : X \to Y$ is defined by means of a subset F of the Cartesian product $X \times Y$; F obeys the law that for any $x \in X$ there exists exactly one $y \in Y$ such that $(x, y) \in F$. y is then denoted as $f(x)$ and F is the graph of f. In human language, this signifies that each element chosen from X has precisely one image in Y. Concerning mappings, we formulate still the following extremal properties: (a) f

is injective if and only if $f(x) = f(x')$ implies that $x = x'$ or each x has a different image (b) f is surjective if and only if for each $y \in Y$ there exists an $x \in X$ such that $f(x) = y$ or, in other words, every potential image is realized effectively. Finally, we say that f is a bijection if and only if it is injective as well as surjective; bijective mappings are equivalences between sets as we shall see now. Suppose $f : X \to Y$ and $g : Y \to Z$ then $g \circ f : X \to Z : x \to g(f(x))$ is the composition of these two mappings. Show that $g \circ f$ is injective if and only if g has this property on $f(X)$ and f obeys this law on X. Show that $g \circ f$ is surjective if and only if g is on $f(X)$; finally, show that $g \circ f$ is a bijection if and only if g is and f is injective. In case $f : X \to Y$ is a bijection, it becomes possible to define a unique inverse $f^{-1} : Y \to X$ by means of

$$f^{-1}(f(x)) = x$$

or $f^{-1} \circ f = \mathrm{id}_X$ where id_X constitutes the identity mapping on X. Derive here from that

$$f \circ f^{-1} = \mathrm{id}_Y$$

using the surjectivity of f. Finally, one shows that f^{-1} also is a bijection; we say henceforth that X and Y are equivalent if and only if there exists a bijection from X onto Y. Using the previous properties, one shows that this relation is reflexive, symmetric and transitive. Now, we are in position to define topological equivalences $f : X \to Y$; f is continuous if and only if the inverse of each open set O in Y, denoted by $f^{-1}(O)$, is open in X. For a continuous bijection, one has that f^{-1} is continuous if and only if $f(V)$ open is in Y for any open V in X. In case a function f satisfies this property, we call it an open mapping. An example of a continuous bijection for which the inverse is not continuous, is given by $f : (-1, 0) \cup (0, 1) \to (-1, 0) \times \mathbb{Z}_2 : x \to (-|x|, \theta(x))$ where $|x| = -x$ if $x < 0$ and x if $x \geq 0$. $\theta(x) = 0$ for $x \leq 0$ and 1 otherwise; finally, $\mathbb{Z}_2 = \{0, 1\}$. The topology defined on $(-1, 0) \times \mathbb{Z}_2$ is the natural one of $(-1, 0)$ and is henceforth not Hausdorff on $\{0, 1\}$. One has that $f((-1, 0)) = (-1, 0) \times \{0\}$ which is not open whereas $(-1, 0) \times \mathbb{Z}_2$ is. A topological equivalence is given by means of a bijection f which is continuous and open. Such mappings are called homeomorphisms and the reader verifies that this definition obeys all requirements of an equivalence relationship indeed.

We return to our study of metric topologies and in particular alternative characterizations of compactness. A sequence $(x_i)_{i \in I}$ is called Cauchy if and only if for each $\epsilon > 0$, there exists an i, such that for all $i \prec j, k$ one has that $d(x_j, x_k) < \epsilon$. In human language, this reads: if one proceeds sufficiently far in the sequence then the points reside arbitrarily close together. Such a property suggests the existence of a unique limit point x; a metrical space (X, d) for which any Cauchy sequence has a limit point is called complete. In case K is a compact set, then one shows that any sequence $(x_i)_{i \in I}$ has a subsequence with a limit point in K. The proof is simple, consider arbitrary finite (due to compactness) covers with balls of radius $\frac{1}{n}$; then one finds a sequence of balls $B(y_n, \frac{1}{n})$ such that finite intersections $\cap_{n=1}^{m} B(y_n, \frac{1}{n})$ contain an infinite number of $x_i \in K$. This defines a sub-sequence with as limit point

$$x = \cap_{n=1}^{\infty} B(y_n, \frac{1}{n})$$

in K. Reversely, suppose that any sequence in K has a Cauchy sub-sequence with a limit point in K, then K is compact. Choose a cover of K of open balls - without limitation of validity- $B(y_n, \epsilon_n)$ where $n \in \mathbb{N}$ and suppose that no finite sub cover exists. Define then $B_m = \cup_{n=1}^{m} B(y_n, \epsilon_n)$, we henceforth arrive at the conclusion that for any m there exists an $m' > m$ such that $B_{m'} \cap \overline{B_m}^c \cap K \neq \emptyset$. In particular, we construct a sequence (x_m) with the property that for any m there is an $m' > m$ such that $x_{m'} \in \overline{B_m}^c$. This sequence cannot contain a Cauchy sub-sequence with some limit point x because $x \in B_m$ for m sufficiently large which is a contradiction. We just proved that a set is compact in a metric topology if and only if any sequence contains a Cauchy sub-sequence with limit point in K. Prove the following properties:

- define on $\mathbb{R}$ the function $d(x, y) = |y - x|$, show that this defines a metric (easy exercise),
- prove that in the metric topology on $\mathbb{R}$, the closed interval $[a, b]$ is compact (hint: use the decimal representation of real numbers) (difficult),
- suppose two topological sets X, Y, then the product topology $\tau(X \times Y)$ is the smallest topology containing $\tau(X) \times \tau(Y)$, where the last contains elements $U \times V$ with $U \in \tau(X)$ and $V \in \tau(V)$,
- show that the Cartesian product $K_1 \times K_2$ of two compact sets is compact in the product topology (average),

- a metrical space (X, d) is bounded if and only if there exists an $M > 0$ such that $d(x, y) \leq M$ for all $x, y \in X$; show that a compact space is closed and bounded (easy).

Again, the reader might utter that this kind of considerations are far too general and that our world is much more detailed in the sense that light rays bend and twist around one and another and that this behavior is geometrical and continuous in nature. To describe these features in detail, one needs the notion of a local scalar product which we shall study further on in chapter six giving further rise to analytical geometry. Note the following: suppose that $\gamma : [a, b] \to X$ is a continuous curve joining x and y and define the length functional $L(\gamma)$ of γ where

$$L(\gamma) = \sup_{a=t_0<t_1<t_2<\ldots<t_{n+1}=b} \sum_{k=0}^{n} d(\gamma(t_k), \gamma(t_{k+1}))$$

and sup means taking the supremum of this sum over all finite partitions $a = t_0 < t_1 < t_2 < \ldots < t_{n+1} = b$ of the closed interval $[a, b]$. The supremum of a set of real numbers A is the smallest number larger or equal as any number $x \in A$. The supremum is also called the upper bound and the reader shows that by definition the supremum always exists and is unique by means of addition of the number $+\infty$. Likewise, one defines the infimum or under bound and one shows again it exists and is unique. Concerning the sum, one notices that breaking up an interval $[t_k, t_{k+1}]$ into two disjoint pieces by means of addition of an intermediate point $t_k < t_{k+\frac{1}{2}} < t_{k+1}$ the sum increases by means of the triangle inequality. Henceforth, splitting up an interval $[a, b]$ leads to a higher sum by means of the triangle inequality.

Now, we will formulate our main result; a complete metric space (X, d) defines a path metric d if and only if

$$d(x, y) = \min_{\gamma:[a,b]\to X, \gamma(a)=x, \gamma(b)=y} L(\gamma).$$

In other words, when the distance between two points equals the minimal length of a curve joining x to y we speak about a path metric space. The reader is advised to show this by means of using the midpoint property in order to construct such curve using that $L(\gamma) \geq d(x, y)$. Reversely, in case such a curve exists, one automatically finds a midpoint. A curve minimizing length is called a geodesic and in a path metric space, the length of a geodesic equals the distance between two points. Later on, we shall arrive at a more detailed characterization of geodesics when imposing more structure. Again, those primitive notions allow one to obtain a substantial amount of results some of which have been obtained by Mikhail Gromov and Peter Anderson. Studying those primitive metric spaces further on requires consultation of their work.

As one notices, our language is not rich enough to speak about notions such as perpendicularity, angles etcetera. One gradually learns that this book will become more and more specific, that the language gets more rich and complex allowing for stronger connections and results. Compactness or local compactness is an important notion because the (local) topology is finite in a way. Spaces which are not locally compact often do not allow for certain mathematical structures to exist because there is too much "room" or space such as is the case for integrals. We now arrive at very special building blocks: line segments, triangles and pyramids as well as higher dimensional generalizations thereof. We shall use those to describe certain topological spaces and characterize those: a central element herein is the concept of homology which leads to further categorical abstractions.

Exercises.
The intention of these short exercises is to show to the reader that path metric spaces carry a notion of continuity which shall be further restricted to a "smooth" structure which facilitates the definition of an angle as well as perpendicularity. This is also the case here, but the sum of angles around a point does not need to equal 2π something which is for example the case with the top point of a cone. This characterization translates into the fact that one cannot stratify the top which is nevertheless possible for points on the mantle. The latter define a tangent plane or an infinitely small Euclidean structure which is not the case for the top for which the following definitions hold also

- Take a point x and consider a sequence of neighboring points y_n, z_n equidistant, meaning $d(x, y_n) = d(x, z_n)$ from x at two different geodesics emanating from x such that in the limit for n to infinity, the sequence approaches x as a limit point. In case the limit

$$\lim_{n\to\infty} \frac{d(x,y_n)^2 + d(x,z_n)^2 - d(y_n,z_n)^2}{2d(x,y_n)d(x,z_n)}$$

exists, we put it equal to $\cos(\theta_x(y,z))$ whereby $\theta_x(y,z)$ is the angle in radials between the two geodesics.

- Show that the total angle around the top of a cone is (a) well defined (does not depend upon the partition into triangles) and (b) smaller as 2π.

- Construct with cutting and pasting a space with opening angle greater as 2π; spaces of the previous type are called spherical and of the latter hyperbolic.

- Alexandrov curvature: in a flat geometry, one has the property that the vector to the midpoint in a triangle is given by

  $$\vec{xr} = \frac{1}{2}(\vec{xa} + \vec{xb}).$$

  Consequently, the distance is equal to

  $$d(x,r)^2 = \frac{1}{4}(d(x,a)^2 + d(x,b)^2 + 2d(x,a)d(x,b)\cos(\theta_x(a,b))).$$

  Considering the definition of an angle and taking midpoints r_n between y_n, z_n and likewise the limit for n to infinity, we are in position to study the limit R of the quantities

  $$R_n(y,z) = -\frac{d(x,y_n)^2 + d(x,z_n)^2 + 2d(x,y_n)d(x,z_n)\cos(\theta_x(y_n,z_n)) - 4d(x,r_n)^2}{d(x,y_n)^2 d(x,z_n)^2 \sin^2(\theta_x(y_n,z_n))}$$

  or alternatively

  $$R_n(y,z) = \frac{-2d(x,y_n)^2 - 2d(x,z_n)^2 + d(y_n,z_n)^2 + 4d(x,r_n)^2}{d(x,y_n)^2 d(x,z_n)^2 \sin^2(\theta_x(y_n,z_n))}$$

  with dimension one over meter squared. Symmetrize this expression over all sides and one arrives at an expression depending upon the triangle only.

- In case $R > 0$, then is spherical, otherwise it is flat ($R = 0$) or hyperbolic $R < 0$.

Chapter 5

Simplicial Homology

Whereas the previous three chapters were very abstract, we shall now continue to work with more tangible objects, things we know from everyday life. We shall use abstraction of these objects to deal with them in a more appropriate way. This has its advantages because it allows us to *calculate* with them; this actually is the main miracle of abstraction, that it allows us to *do* things. The topological spaces to be studied in this chapter are those which are modeled by means of the n-dimensional real space

$$\mathbb{R}^n = \times_{i=1}^n \mathbb{R} = \{(x_i)_{i=1}^n | x_i \in \mathbb{R}\}$$

which is the set of n-tuples of real numbers equipped with the product metrical topology of $\mathbb{R}$. One can extend the notion of a sum by means of the definition

$$(x_i) + (y_i) = (x_i + y_i)$$

and likewise can one define the *scalar* multiplication of a real number with an n-tuple *vector* by means of

$$r.(x_i) = (rx_i).$$

More in general, let R be a field and $G, +$ a commutative group, then we say that G is an R module in case there exists a scalar multiplication such that

$$1.g = g;\ (rs).g = r.(s.g);\ (r+s).g = r.g + s.g;\ r.(g_1 + g_2) = r.g_1 + r.g_2$$

for all $r, s \in R$ and $g, g_1, g_2 \in G$. In case $R = \mathbb{R}$ we call the module a real vector space. In $\mathbb{R}^n, +$, we have special vectors e_i, defined by the number 1 on the i'th digit and zero elsewhere; herefore, it holds that

$$\sum_{i=1}^n r_i.e_i = 0$$

if and only if it holds that all $r_i = 0$ and moreover all vectors can be written uniquely as

$$\sum_{i=1}^n r_i.e_i.$$

In case these properties hold for a set of vectors $\{v_i | i = 1 \ldots m\}$, then we call $\{v_i | i = 1 \ldots m\}$ a basis. One notices that we have used two integer numbers here, n for the e_i and m for all v_j; it is now a piece of cake to show that $n = m$. The reason is the following, because e_i is a basis, one can write the v_j uniquely as

$$v_j = \sum_{i=1}^n v_j^i e_i$$

and reversely

$$e_i = \sum_{j=1}^m e_i^j v_j.$$

Henceforth,

$$\sum_{i=1}^{n} v^i_j e^k_i = \delta^k_j;\ j,k:1\ldots m$$

and

$$\sum_{j=1}^{m} e^j_i v^l_j = \delta^l_i;\ i,l:1\ldots n$$

where $\delta^k_j = 1$ if and only if $j = k$ and zero otherwise. This system of equations is symmetrical in e and v and therefore $m = n$ given that both mappings are injective. Henceforth n is a basis *invariant* and called the dimension of $\mathbb{R}^n, +$. Now, we have a sufficient grasp upon real vector spaces and we proceed by defining special building blocks mandatory for the construction of simplicial manifolds.

What follows is a generalization of simple cutting and pasting of higher dimensional triangles and pyramids. We may construct so called Euclidean bodies in this way and the old fashioned approach towards a classification of topological spaces upon a homeomorphism has been made as such. However, different lines of argumentation which are less constructivist can lead towards such classification too. Consider the space $\mathbb{R}^{n+1}$ and consider a basis $v_i; i = 0\ldots n$, then the n simplex $(v_0v_1\ldots v_n)$ is defined by means of the closed space

$$(v_0v_1\ldots v_n) = \{\sum_{i=0}^{n} \lambda_i v_i | \lambda_i \geq 0, \sum_{i=0}^{n} \lambda_i = 1\}.$$

This is all a bit abstract and in order to get a picture of how such space looks like, one imagines the $0, 1, 2, 3$ dimensional cases. A zero dimensional simplex (v_0) is simply a point, a one dimensional simplex is given by the line segment (v_0v_1) which may be embedded into the plane $\mathbb{R}^2$. A two dimensional simplex $(v_0v_1v_2)$ is given by a triangle which can be embedded into $\mathbb{R}^2$ whereas finally $(v_0v_1v_2v_3)$ describes a pyramid in $\mathbb{R}^3$. In general, the simplex $(v_0v_1\ldots v_n)$ is a convex space meaning that the line segment between two points $x, y \in (v_0v_1\ldots v_n)$ completely belongs to $(v_0v_1\ldots v_n)$. The line segment between two points x, y is the set

$$\{\lambda x + (1-\lambda)y | 0 \leq \lambda \leq 1\}.$$

Points of the simplex which do not belong to the interior of a line segment belonging entirely to the simplex are called extremal. Show by means of exercise that the only extremal points of $(v_0v_1\ldots v_n)$ are given by v_i. One calls the simplex the convex hull of the extremal points $\{v_i | i = 0\ldots n\}$. We know now how a module is defined as well as a simplex which allows us for the definition of a linear operator. A mapping $A : V \to W$ between two R modules V, W is linear if and only if

$$A(rv_1 + sv_2) = rA(v_1) + sA(v_2)$$

for all $r, s \in R$ and $v_i \in V$. Show that A is injective if and only if $A(v) = 0$ implies that $v = 0$. Now, we shall work a bit more abstractly: we do not need at this point the property that $v_i \in \mathbb{R}^{n+1}$, something which was required for matters of representation. We shall temporarily proceed by insisting that the v_i, w_j are merely points which are not necessarily associated to vectors in some linear space. Note that a simplex $(v_0v_1\ldots v_n)$ naturally possesses an orientation defined by the order in which the vertices appear and that swapping two vertices reverses the orientation, meaning for example

$$(v_0, v_1) = -(v_1, v_0).$$

An n dimensional simplcial complex is defined as a collection of n distinct simplices such that any sub-simplex also belongs to it. We shall be interested in taking formal sums of simplices of the same dimension $k \leq n$; ab initio, you might want to impose certain constraints such as (a) no branching meaning that no more as two k dimensional simplices share the same $k-1$ dimensional sub-simplex. Also (b) you might want for every k simplex to appear exactly once into such a sum so that we can think of it as being single valued. Also, you might insist upon it being (c) oriented which in its most general sense would mean that the contribution of internal $k-1$ dimensional sub-simplices vanishes. This means that, upon taking a formal sum

$$\sum_i a_i(v^i_1, \ldots, v^i_k)$$

where all $(v^i_1, \ldots, v^i_k)$ are different, we have that in case

$$\sum_w \partial_w \partial_{w_1} \ldots \partial_{w_{k-1}} (v^l_1, \ldots, v^l_k) = \pm \sum_w \partial_w \partial_{w_1} \ldots \partial_{w_{k-1}} (v^s_1, \ldots, v^s_k) \neq 0$$

for at least two values $s \neq l$, then $\sum_w \partial_w \partial_{w_1} \ldots \partial_{w_{k-1}} \sum_i a_i(v^i_1, \ldots, v^i_k) = 0$ where ∂_v is the linear operator attached to any vertex v defined by $\partial_v(vv_0 \ldots v_i) = (v_0 \ldots v_i)$ in case none of the v_j equals v and zero otherwise, here it is assumed[1] that $() = 1$. This, taken together with condition (a) simply means that if precisely two k dimensional simplices share the same $k-1$ dimensional sub-simplex then the induced orientations differ. Let us for now keep things in the middle and see if those concerns really matter. We define the boundary operator $\partial_n : Z_n \to Z_{n-1}$ as the linear operator over $\mathbb{Z}$ mapping a simplex $(v_0 v_1 \ldots v_n)$ to

$$\partial_k(v_0 v_1 \ldots v_k) = \sum_{i=0}^{k} (-1)^i (v_0 \ldots v_{i-1} v_{i+1} \ldots v_k) = \sum_w \partial_w (v_0 v_1 \ldots v_k).$$

One verifies that $\partial_{k-1}\partial_k S_k = 0$ given any sum of simplices. From the definition, it follows that the boundary of any linear combination of simplices is oriented in the previous sense. Also, by the same virtue, any closed sum S_k of simplices, meaning $\partial_k S_k = 0$, is oriented since $\sum_w \partial_w \partial_{w_1} \ldots \partial_{w_{k-1}} S_k = (-1)^{k-1} \partial_{w_1} \ldots \partial_{w_{k-1}} \left(\sum_w \partial_w S_k\right) = 0$ given that $\partial_v \partial_w = -\partial_w \partial_v$. Now, one may wonder whether any closed S_n can be written as a linear combination of closed simplicial complexes satisfying (a) and (b). We shall first prove that this is the case for $n = 1$; take any one dimensional complex $S_1 = \frac{1}{2} \sum_{ij} a^{ij}(v_i v_j)$ where $a_{ij} = -a_{ji}$ where a factor $\frac{1}{2}$ has been included because each simplex is summed over twice. Assume now that the simplex is closed meaning that $\sum_i a_{ij} = 0$ and choose the smallest positive a_{ij}. Then at the vertex j, one certainly has some k such that $a_{jk} \geq a_{ij}$, proceed towards k and substract a_{ij} from a_{jk}. One repeats this procedure a sufficient number of times until the curve comes back to itself defining a_{ij} times a canonical loop obeying a and b. Now, the remainder contains at least one edge less and is also closed; hence, upon repetition of this procedure we arrive at our result. Now, a one dimensional simplicial complex is rather trivial as each such structure can be consistently oriented. This is no longer true in two dimensions and we shall generalize here the construction of a Mobius strip in order to provide for a counterexample. I will not provide all details but the reader will see how it works. Take an oriented square with four corner boundary points in order (according to the orientation of the boundary) given by $1, 2, 3, 4$ and identify the line segments 12 with 34 and 23 with 41, then the reader notices that, given a representation in terms of a simplicial complex, all the interior lines are cancelled when taking the boundary of this simplicial complex, but the "boundary lines" between 12 and 23 are doubled. So, this sum is not closed; to compensate for this, take a second identical construction but now with opposite orientation on the "boundary lines" and glue them together. Then, the boundary of this doubled complex vanishes but there is no way to undo the bifurcation at the lines 12 and 23 which are now adjacent to four half planes instead of two. It appears that we have to live with such "anomalies" as there is no way to exclude them, therefore we consider all formal linear combinations of k dimensional simplices. To repeat, we consider a formal linear combination T_k of k simplices closed if and only if $\partial_k T_k = 0$ and exact if and only if $T_k = \partial_{k+1} S_{k+1}$ for some S_{k+1}. It is clear that exact simplicial complexes are closed using the crucial property of a boundary operator and we define accordingly the $\mathbb{Z}$ modules $C_k(S_n)$ of all closed k sums and $E_k(S_n)$ of all exact k sums, where $E_n(S_n) = \{0\}$ and $C_0(S_n)$ equals $\mathbb{Z}^{V-1}$ with V the number of points or vertices in S_n. Clearly it holds that $E_k(S_n) \subseteq C_k(S_n)$ and we define the homology classes $H_k(S_n)$ as the quotient module

$$H_k(S_n) = \frac{C_k(S_n)}{E_k(S_n)}$$

being the $\mathbb{Z}$ module of $E_k(S_n)$ equivalence classes in $C_k(S_n)$. We say that two closed sums T_k, Y_k are equivalent if and only if $T_k - Y_k \in E_k(S_n)$. So far for the general theory of simplical complexes, we now arrive to the very important sub theory of topological spaces A homeomorphic to a simplicial complex S_n; the important step herein consists in proving that $H_k(A)$ is well defined because homeomorphic simplicial sums define the same homology module. The reader may try to show this fact by him or herself as a kind of difficult exercise but it it clear that the statement is rather obvious. Indeed, the boundary operator is defined independently of the simplicial decomposition. The dimension of $H_k(S_n)$ plus one, in case $k = 0$, is called the k-th Betti number b_k of the simplicial complex S_n. The reader now makes the following exercises:

[1] Note that we deviate here slightly with the convention in the literature where $() = 0$. This will result in a zero'th homology froup with one generator less; it is worthwhile keeping this in mind when discussing the definition of the Betti numbers.

take a two dimensional spherical surface and show that $b_2 = 1, b_1 = 0, b_0 = 1$. The two torus T_2 is defined by taking an oriented square and glue opposite sides to one and another; show that $b_2 = 1, b_1 = 2, b_0 = 1$. In general, one defines the Euler number of a two dimensional simplicial complex S_2 as

$$\chi(S_2) = D - L + V$$

where D is the number of triangles and L the number of line segments. One can show that the Euler number is a topological invariant; calculate that the Euler number of a two sphere is given by $2 = b_2 - b_0 + b_1 = 1 - 0 + 1$ and that of a torus by $0 = 1 - 2 + 1$. In general, one shows that

$$\chi(S_n) := \sum_{i=0}^{n} (-1)^i V_{n-i} = \sum_{i=0}^{n} (-1)^i b_{n-i}$$

where V_i equals the number of i dimensional sub-simplices. To start with the calculation of the dimension of a homology class, note that an element of $H_k(S_n)$ corresponds to a closed k dimensional connected surface which cannot be contracted to a point. Concerning the calculation of b_1 on the two sphere, it is clear that any closed curve can be reduced to a point whereas on the two torus two fundamental circles do exist which are not the boundary of a two dimensional simplicial complex. Consider two closed surfaces A_2 and B_2 and remove a two disk from both of them; now, paste each of the remainders along the circular boundaries resulting in a new closed surface denoted by $A_2 \diamond B_2$. Show that the operation $\diamond$ is associative as well as commutative with as identity element the two dimensional surface S^2. Calculate that the Euler number of the n-fold cross-product of T_2 equals $2 - 2n$; more in particular, it holds that

$$\chi(A_2 \diamond B_2) = \chi(A_2) + \chi(B_2) - 2.$$

Later on, we shall study the notion of a manifold and one of the most important results is that any closed, compact, connected and oriented two dimensional topological space is homeomorphic to S^2 or an n-fold product $T_2 \diamond T_2 \diamond \ldots \diamond T_2$. This formula can be generalized towards any dimension, where the connected sum is then defined by means of cutting out the interior of a ball and identifying the boundaries; the reader verifies that in general

$$\chi(A_n \diamond B_n) = \chi(A_n) + \chi(B_n) - 2$$

for n even and

$$\chi(A_n \diamond B_n) = \chi(A_n) + \chi(B_n)$$

for n odd. This implies that closed, compact, connected as well as orientable two dimensional manifolds are completely characterized topologically by means of the Euler number. For closed manifolds, one shows that $b_{n-i} = b_i$ something which is called Betti duality, a result which may be proved by definition of a duality operator $\star$ on the simplicial complexes such that $S_n^\star$ is homeomorphic to S_n and $H_k(S_n)$ is mapped bijectively to $H_{n-k}(S_n^\star)$. One can imagine $\star$ as a natural generalization of the following operation on a one dimensional simplicial complex S_1: it maps every line segment r to a point $r^\star$ and each point p to a line segment $p^\star$ such that $\star$ interchanges the operation $\subseteq$ meaning $r^\star \subseteq p^\star$ if and only if $p \subseteq r$. $S_1^\star$ is a closed simplicial complex if and only if S_1 is in case no branching occurs; the Euler number changes in case S_1 shows branching as the reader verifies. Henceforth, the manifold condition is mandatory and Betti duality does not hold for general closed simplicial complexes. The reader should prove that two circles having a common point show bad behavior under the duality transformation. The notion of a variety is henceforth really special and our result, that closed two dimensional and oriented varieties are classified by the Euler number only does not hold in higher dimensions. Here ends our discussion of simplicial homology which can be summarized by a chain of operations $\partial_k : Z_k(S_n) \to Z_{k-1}(S_n)$ met $\partial_0 : Z_0(S_n) \to \mathbb{Z}$ and[2] $\partial_{k+1}\partial_k = 0$. Such a structure is called a chain and those objects enjoy plenty of beautiful characteristics which are much more primitive as the topological point of departure. An initial point for higher mathematics therefore!

It is clear, from the simplicial point of view, that topological spaces of dimension n cannot be classified by means of the Betti numbers. The reader is invited to show this by means of braiding three closed surfaces in different ways. Later on, we shall study the Euler number from the viewpoint of vector-fields, akin Morse theory, as well as closed differential forms determined by the homology classes.

Exercise: the Poincaré Conjecture

The conjecture of Poincaré is that every 3 dimensional compact, closed topological space $\mathcal{M}$ which is path

[2] In the literature, this is zero given that there one takes another definition of ∂_0.

connected and has trivial first homotopy class, is homeomorphic to the 3 dimensional sphere. Note that I am speaking of homotopy instead of homology which is another and much crazier way of constructing topological invariants; the reader is encouraged to wade through the literature on homotopy which is full of rich results (for example, the homotopy groups are also labeled by a discrete index, referring to dimension, but they may well be nontrivial beyond the dimension of the embedding space). The first homotopy group consists out of all equivalence classes of continuous closed curves with a base point which cannot be continuously deformed into one and another while keeping the base point fixed. Obviously, for the torus, the homology group is $\mathbb{Z} \oplus \mathbb{Z}$ which basically means that you consider linear combinations over $\mathbb{Z}$ of two independent generators, whereas the first homotopy group is given by the free group in two letters a, b since there is no relationship between the two generators. However, this is no longer true upon considering the torus and cutting out a small disc. The boundary of the latter is obviously the boundary of the complement of the disc (and is therefore trivially zero in the homology group) but it cannot be deformed into any other existing element in the homotopy group of the torus. Nevertheless one may trivially prove that if the first homotopy group vanishes, then the first homology group must vanish too; the reverse is not true however. Note that this theorem does not hold in higher dimensions as for example $S^2 \times S^2$ provides for a counterexample. Taking the classification of closed, compact and connected two dimensional topological spaces we have just given, I once constructed the following simple argument. As of today, I do not know where my error resides and I encourage the reader to think about it.

- Show that $\mathcal{M}$ allows for a path metric d.
- Consider an arbitrary point p and show that for sufficiently small r, the surface $L_r := \{x | d(p, x) = r\}$ is homeomorphic to the 2 dimensional sphere S^2.
- Show that there exists a critical point r_0 such that L_{r_0} is no longer a sphere.
- In case L_{r_0} is a point, the theorem is proved; otherwise we have a compact 2 dimensional topological space obtained from the sphere by means of identification of k dimensional subspaces where k can range from 0 to 2.
- Show that the subsequent connected components of the topological space for $r > r_0$ are again two dimensional connected, closed spaces[3] which can only close up to a point in a three dimensional closed space in case they are homeomorphic to the sphere S^2.
- Subsequently, to close the topological space, all components different from some S^2 and possibly the S^2 themselves must be pasted together leading to a nontrivial first homotopy class which is forbidden.
- Consequently $\mathcal{M}$ is a 3 dimensional sphere.

Simplicial Gravitation

Simplicial metric spaces are very simple and entirely characterized by means of distances $d(v_0 v_1)$ defined on the line segments $(v_0 v_1)$. One defines the following operators: $x_w(v_0 \ldots v_i) = (w v_0 \ldots v_i)$ and $\partial_w(w v_0 \ldots v_i) = (v_0 \ldots v_i)$ in case none of the v_j equals w. The remaining cases where this last condition is violated lead to the null simplex with as boundary conditions $\partial_w(w) = \mathbf{1}$, $x_w \mathbf{1} = (w)$ where $\mathbf{1} = ()$ is the empty simplex. From this, it follows that $(x_w)^2 = 0$ as well as $(\partial_w)^2 = 0$. One verifies that the operator $\partial = \sum_{w \in S} \partial_w$ is the usual boundary operator what shows that ∂_w constitutes the appropriate derivative operator defined by means of the boundary operator ∂. The empty simplex constitutes the neutral element regarding the cross product $*$ defined by means of

$$(v_0 \ldots v_i) * (w_0 \ldots w_j) = (v_0 \ldots v_i w_0 \ldots w_j).$$

One simply verifies that $x_w x_v = -x_v x_w$ and likewise for the operators ∂_v, ∂_w. Henceforth, the creation operators associated to a vertex generate a Grassmann algebra; moreover, it holds on the vector space of simplices that

$$\partial_v x_w + x_w \partial_v = \delta(v, w)$$

[3] This seems to be the crucial step! It is certainly true for the theorem in two instead of three dimensions where a circle possibly bifurcates into two circles which, in case they rejoin, gives rise to a nontrivial homotopy. The reader may convince himself of that by studying the example of a torus versus a long "sausage". In both cases, we have that for generic points x the circles of radius r around x identify at some points but split later again into two distinct circles which in case of the torus rejoin and in case of the sausage individually collapse to a point.

such that the ∂_v represent Grassmann annihilation operators. Bosonic line segment operators are consequently defined by means of

$$\partial_{(vw)} = \partial_w \partial_v$$

and such operators satisfy

$$\partial_{(vw)}(yz) = \delta(v,y)\delta(w,z) - \delta(v,z)\delta(w,y)$$

giving rise to an oriented derivative. The simplex algebra is henceforth defined by means of polynomials spanned by monomials which are formal products of simplices $(v_0 \ldots v_j)$ for all $j : 0 \ldots n$. Mind that this formal product does not equal the cross-product implying that $\mathbf{1}$ does not constitute the neutral element. Given that on general spaces bi relations carry an evaluation by means of the metric d it is natural to limit the function algebra to two simplices $(v_0 v_1)$ given that other simplices do not procure for independent variables. The bosonic character of $\mathbf{1}$ implies that the ∂_v, x_w constitute Fermionic Leibniz operators on the function algebra. Indeed, one has that

$$\partial_v((w)Q) = \partial_v((x_w \mathbf{1})Q) = \partial_v x_w(\mathbf{1}Q) - \partial_v(\mathbf{1}x_w Q) =$$

$$(k+1)\delta(v,w)\mathbf{1}Q - x_w(\mathbf{1}\partial_v Q) - \partial_v(\mathbf{1}x_w Q)$$

which reduces to

$$(k+1)\delta(v,w)\mathbf{1}Q - (x_w)\partial_v Q - \mathbf{1}x_w\partial_v Q - \mathbf{1}\partial_v x_w Q = \delta(v,w)\mathbf{1}Q - (x_w)\partial_v Q$$

where k denotes the degree of the monomial Q given by the number of factors. This follows immediately from the Leibniz rule given that the operator

$$x_w \partial_v + \partial_v x_w = \delta(v,w)$$

is bosonic. Henceforth, the even simplex variables behave bosonically whereas the odd ones fermionic. Indeed,

$$\partial_v((wz)Q) = \partial_v((x_w(z))Q) = \partial_v(x_w((z)Q) + ((z)x_w Q)) = -x_w\partial_v((z)Q) - (z)(\partial_v x_w Q)$$

which reduces to

$$= x_w((z)\partial_v Q) - (z)(\partial_v x_w Q) = (wz)\partial_v Q.$$

Given that the usual derivatives of a function are defined by means of the infinitesimal intervals $(x - |\epsilon|, x + |\epsilon|)$ where $f(v+\epsilon, v-\epsilon)$ gets identified with the coordinate function $f(x)$. This is logical given that the $v \pm \epsilon$ are fermionic and independent such that the intervals $(v-\epsilon, v+\epsilon) \sim x$ are bosonic. Note that products of the form $(v-\epsilon)(v+\epsilon)$ can be further derived such that

$$\partial_x f(x) = \mathbf{L}\left[\partial_{(v-\epsilon,v+\epsilon)} f(v-\epsilon, v+\epsilon)\right]$$

where $\mathbf{L}$ merely retains the monomials depending exclusively of the line segments. This phenomenon clearly occurs in $(vw)^2$ whose (vw) derivative equals

$$2(vw) - 2(v)(w).$$

To obtain the standard commutation-relations on the function algebra generated by (vw) we define

$$\widehat{x}_{(vw)}Q := x_{(vw)}x_{\mathbf{1}}Q$$

where Q is a polynomial defined on the edges (r,s) and $x_{(vw)}$ is a bosonic Leibniz operator defined by

$$x_{(vw)}(v_0 \ldots v_j) = (vwv_0 \ldots v_j).$$

By definition, one has that

$$x_{(vw)}(rs) = 0$$

if and only if r or s equals v, w and moreover

$$(x_{(vw)} + x_{(rs)})((vw) + (rs)) = 2(vwrs)$$

which vanishes unless (r, s) is the opposite side of a pyramid which we shall forbid from now on. In particular, this does not apply to geodesics

$$\gamma(v_0 v_i) := (v_0 v_1) + (v_1 v_2) + \dots (v_{i-1} v_i)$$

which satisfy

$$x_{\gamma(v_0 v_i)} := \sum_{j=1}^{i} x_{(v_{j-1} v_j)}$$

and therefore

$$x_{\gamma(v_0 v_i)} \gamma(v_0, v_i) = 0.$$

Next, we define the derivatives

$$\partial_{\gamma(v_0, v_i)} := \sum_{j=1}^{i} \partial_{(v_{j-1} v_j)}$$

and consider the operator

$$\widehat{\partial}_{\gamma(v_0, v_i)} = \mathbf{L} \circ \partial_{\gamma(v_0, v_i)}$$

and one calculates that

$$\widehat{\partial}_{\gamma(v_0, v_i)} \widehat{x}_{\gamma(v_0, v_i)} - \widehat{x}_{\gamma(v_0, v_i)} \widehat{\partial}_{\gamma(v_0, v_i)} = 1$$

on the function algebra generated by the monomials Q of the form $(\gamma(v_0, v_i))^k$ where $k > 0$. We have now a tool to do physics; in particular,generated by the monomials Q of the form $(\gamma(v_0, v_i))^k$ where $k > 0$. We have now a tool to do physics; in particular,

$$\mathbf{E} P(\gamma(v_0, v_i)) = P(\sum_{j=1}^{i} d(v_{j-1} v_j))$$

is the evaluation function. The reader is invited to expand this theory further as well as to implement the Fourier transformation from chapter fourteen on conic tangent spaces. Hint: integrate in "hyperbolic" or "spherical" coordinates by replacing the $n-1$ sphere with the level surface $H^{n-1}(\epsilon, v_0) = \{x | d(v_0, x) = \epsilon\}$ for ϵ sufficiently small such that $H^{n-1}(\epsilon, v_0)$ belongs to the star neighborhood of v_0. See chapter thirteen for more information.

Betti numbers.
Give an example of two oriented spaces with the same Betti numbers and develop the homology concept further on with the purpose of distinguishing both (very difficult).

Chapter 6

Linear Spaces and Operators

The reader profits from reconsidering the path taken so far towards further specialization of our mathematical language: we started from set theory which allows for huge constructions, the next step was to narrow our scope to general topology and in particular simplicial homology. Finally, we stumbled upon the notion of a manifold which we have treated on an intuitive level so far and shall make exact further on in this book. The advantage of more specific structures is that they allow for more results, that they are better under control: herein, the mathematician often anticipates an unproved stability result namely that structures nearly satisfying the idealized rules also obey the same properties to a good approximation. That is the real virtue of subtle simplification, that it allows us to understand things which are not valid from a higher, more general point of view. Albert Einstein often spoke about this in a way that one needs to represent things as easy as possible but not too simplistic; too many limitations often involve a huge risk that things get too narrow whereas too little assumptions lead towards the problem that even the most simple of observations do not enjoy a stringent explanation within your framework. Science is the art to explore that very fine boundary in a better and better way, little by little without dogmatic certainty but with reason and intelligence.

In this philosophy, linear spaces are too specific but we shall later on study manifolds which do look like linear spaces on a "small scale" and henceforth constitute a wide generalization of the former. So, we apply now the opposite strategy and study first the "simple" case prior to studying the more general one. A linear space is a bi-module defined over the field of the real or complex numbers; the prefix "bi" refers to the fact that the scalar multiplication can occur from the left as well as from the right and that both are equal in this case which is logical because both number systems are commutative regarding the multiplication. For quaternionic modules, we could speak about a left and right or bi one and those situations do differ in plenty of cases. We say that the linear space is finite dimensional if and only if the dimension over the field is finite and infinite dimensional otherwise. The dimension is still well defined as the number of basis vectors given that the existence of a basis can be shown from set theoretical considerations.

A linear mapping $A : V \to W$ from one linear space to another over the same field is a function satisfying

$$A(r.v + s.w) = r.A(v) + s.A(w)$$

where the dot denotes scalar multiplication. Evidently one has that $A(0) = 0$ and $A(v) = A(w)$ if and only if $A(v - w) = 0$. Henceforth, the so called nucleus of A, defined by $\mathrm{Ker}(A) = \{v | A(v) = 0\}$ measures the deviation from injectivity of A. Every image $A(w)$ has as inverse $w + \mathrm{Ker}(A)$. The nucleus is henceforth itself a linear subspace of V. In the same way, one has that the so called image

$$\mathrm{Im}(A) = \{A(v) | v \in V\}$$

constitutes a subspace of W. It is now evidently true that

$$\mathrm{Im}(A) \cong \frac{V}{\mathrm{Ker}(A)}$$

meaning that both linear spaces are isomorphic to one and another. Indeed,

$$A : \frac{V}{\mathrm{Ker}(A)} \to \mathrm{Im}(A) : w + \mathrm{Ker}(A) \to A(w)$$

is linear and bijective which are the defining characteristics of an isomorphism. A trivial consequence of this theorem is that

$$\dim(\mathrm{Ker}(A)) + \dim(\mathrm{Im}(A)) = \dim(V)$$

where "dim" stands for dimension. Linear mappings can be represented by means of matrices defined with respect to basis vectors e_i in V and f_j in W respectively. The definition is given by

$$A(e_i) = \sum_{j=1}^{m} A_i^j f_j$$

where j is called the row index and i the column-index; taking a general vector $v = v^i e_i$ gives rise to the matrix multiplication

$$A(v) = \sum_{j=1}^{m} (\sum_{i=1}^{n} A_i^j v^i) f_j.$$

The composition of two linear mappings $A : V \to W$ en $B : W \to Z$ results into the matrix product

$$(BA)_i^j = \sum_{k=1}^{m} B_k^j A_i^k$$

where m represents the dimension of W. From now on, we dispose of the summation-signs, a convention which has been named after Einstein; so

$$\sum_{i=1}^{n} A_i^j v^i$$

is noted as

$$A_i^j v^i.$$

A 2×3 matrix, or a matrix with 2 rows and 3 columns is represented as

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

and regarding the matrix product BA one has the rule that the column dimension of B has to be equal to the row dimension of A. Show by means of a computational exercise that

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 13 \\ 19 & 19 \end{pmatrix}.$$

Show that in general for 2×2 matrices A, B one has that

$$AB - BA \neq 0$$

where 0 denotes the zero matrix. This result shows that the matrix multiplication is in general non-commutative and hitherto such operators constitute a non-commutative ring. The latter has been constructed as an object formed by elements which belong to a field. One can justifiably wonder whether the non-commutative number systems such as the quaternions and Clifford algebra's can be represented as matrices over the complex numbers. The answer is yes and one can obtain representations in different dimensions; regarding the quaternions q one has that

$$q = \begin{pmatrix} a + ib & ic - d \\ ic + d & a - ib \end{pmatrix}$$

where $a, b, c, d \in \mathbb{R}$. Another way of writing those in terms of Pauli matrices is provided by

$$q = a.1 + ic.\sigma_1 + id.\sigma_2 + ib.\sigma_3.$$

Now that we have understood a few essentials of matrix calculus, we arrive at the following natural question regarding matrix representations of linear operators: is it possible to find a basis e_i in V associated to a linear operator $A : V \to V$ such that A has a particularly simple matrix representation regarding e_i? Evidently, the formulation is somewhat vague up till now but try to ensure yourself that for an arbitrary $n \times n$ matrix A it almost always holds that $A = ODO^{-1}$ where $OO^{-1} = 1_n = O^{-1}O$ and $D_i^j = \lambda_i \delta_i^j$ with $\delta_i^j = 1$ if and only if $i = j$ and zero otherwise. D is a so called diagonal $n \times n$ matrix and the λ_i are called the eigenvalues such that

$$A(Oe_i) = \lambda_i(Oe_i)$$

which translates as the statement that Oe_i constitutes an eigenvector of A met eigenvalue λ_i. O is called an invertible or reversible $n \times n$ matrix. The reasoning behind it is very simple: in general, it holds that almost any square matrix O is invertible such that O has n^2 degrees of freedom; the mapping $O \to ODO^{-1}$ reduces exactly n dimensions in case all λ_i in D are different because the equation $VDV^{-1} = ODO^{-1}$ implies that $(V^{-1}O)D = D(V^{-1}O)$ such that $V = OD'$ with D' diagonal and henceforth any D "orbit" is $n^2 - n$ dimensional. Given that the number of degrees of freedom in D also equals n we have in total n^2 degrees of freedom and henceforth we obtain a "generic" $n \times n$ matrix. Prior to proceeding, we study the effect of a change of basis on the matrix representation of A. Denoting $e'_i = O(e_i)$ then one has

$$A'^i_j e'_i = A(O(e_i)) = A(O^j_i e_j) = A^k_j O^j_i e_k$$

and as such $A'^i_j = (O^{-1})^i_k A^k_l O^l_j$. So, generically, one can find a basis with respect to which the matrix representation for A is diagonal. The reader should show that all eigenvalues are unique as well as the eigenvectors (upon a normalization constant) in case all λ_j differ. One can find exceptions to this rule! Show that the matrix

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

satisfies $N^2 = 0$ and therefore cannot be diagonalized. This is a simple consequence of the fact that any eigenvalue must be equal to zero and henceforth $N = 0$ in case N can be diagonalized which is a contradiction. In two dimensions, one can by means of a suitable choice of basis ensure that an operator can be exclusively represented by one of the following matrices:

$$\begin{aligned} A &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ A &= \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}. \end{aligned}$$

In case the reader wishes to prove such a result, as well as a suitable extension towards higher dimensions, then I advise further reading up to the end of the chapter prior to dealing with this challenge. An $n \times n$ matrix A can still be interpreted in a different way as being merely the representation of a linear operator with respect to a vector space basis. One can see A as a collection of n ordered column vectors $A = (v_1, \ldots, v_n)$. This viewpoint allows one to interpret A as a simplex or the multi-dimensional cube determined by the column vectors v_i. The determinant $\det(A)$, to be defined below, calculates then the oriented volume of that cube which is just the product of the lengths of the basis vectors e_i if the latter are perpendicular to one and another. We derive a formula for $\det(A)$ from conditions the oriented volume needs to satisfy. First of all, det is multi-linear in the columns; it is to say:

$$\det(v_1, \ldots, v_{i-1}, a.v_i + b.w_i, v_{i+1}, \ldots, v_n) =$$

$$a \det(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + b \det(v_1, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_n)$$

as well as nilpotent in the sense that if $v_i = v_j$ for some $i \neq j$ then the determinant vanishes. This last condition merely reflects that if some axis coincide then the matrix defines a lower dimensional object with vanishing volume. Finally, one imposes the normalization condition that $\det(1_n) = 1$. These three conditions fully determine the functional description for the determinant: from the first and second condition one derives that the determinant is fully anti-symmetrical; it is to say that $\det(v_1, \ldots, , v_i, \ldots, v_j, \ldots, v_n) = -\det(v_1, \ldots, , v_j, \ldots, v_i, \ldots, v_n)$. Combining this fact with the third and first condition one arrives at

$$\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) A^1_{\sigma(1)} \ldots A^n_{\sigma(n)}$$

where σ is a so called permutation and sign denotes the sign thereof. Due to the anti-symmetrical nature of the determinant, each index is allowed to appear exactly once which is encoded in the above formula by means of a permutation. The latter is a bijection of $\{1, 2, \ldots, n\}$ onto itself whereas the sign denotes the even or odd nature of the number of swaps one has to perform to arrive from the identity mapping to σ, where an even number results in the value one and the odd number in minus one. One shows that permutations constitute a non commutative group S_n with $n.(n-1).(n-2)\ldots 3.2.1$ elements and we show now that the sign function is well defined meaning no odd and even number of swaps can occur. The proof is a bit technical; denote with (ij) the swapping operation of the i'th and j'th index leaving the remainder invariant, then it holds that

$$\begin{aligned}(ik)(ij) &= (jk)(ik)\\ (ik)(jl) &= (jl)(ik)\\ (ik)(ij) &= (ij)(jk)\end{aligned}$$

for distinct i, j, k, l. First of all, it is clear that any permutation can be written as a product of such swapping operations. Given a non trivial product of such swaps equivalent to the identity, then it is a simple matter to prove that it contains an even number of swaps using above swapping rules. Indeed, given $\sigma = l(ik)s(jk)$ where l, s are products of swaps and l does not contain a swapping with the index k then the reader shows that it is possible to rewrite this decomposition as $\sigma = ls'$ where s' does not contain the index k and has precisely the same number of swaps as s has modulo two. In this way, one proves that σ contains an even number of swaps. From this it follows that two different products l, s for any permutation σ always differ by an even number of swaps by denoting that ls^{-1} is equivalent to the identity. Henceforth, the function sign is well defined; show that the determinant of a 2×2 matrix is given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Prove now that

- $\text{sign}(\rho\sigma) = \text{sign}(\rho)\text{sign}(\sigma)$
- $\det(AB) = \det(A)\det(B)$.

This last rule holds due to

$$\begin{aligned}\det(AB) &= \sum_{\sigma \in S_n} \text{sign}(\sigma)(AB)^1_{\sigma(1)} \ldots (AB)^n_{\sigma(n)}\\ &= \sum_{\sigma \in S_n} \sum_{m_1,\ldots,m_n} \text{sign}(\sigma) A^1_{m_1} \ldots A^n_{m_n} B^{m_1}_{\sigma(1)} \ldots B^{m_n}_{\sigma(n)}\end{aligned}$$

where $m_1, \ldots, m_n$ is another notation for a permutation. One easily understands this as follows: assuming that $m_i = m_j$ then for every permutation σ it holds that the associated term is compensated by the one associated to the permutation $\sigma(ij)$. Henceforth, we have that

$$\begin{aligned}\det(AB) &= \sum_{\sigma,\rho \in S_n} \text{sign}(\sigma) A^1_{\rho(1)} \ldots A^n_{\rho(n)} B^{\rho(1)}_{\sigma(1)} \ldots B^{\rho(n)}_{\sigma(n)}\\ &= \sum_{\sigma,\rho \in S_n} \text{sign}(\sigma\rho) A^1_{\rho(1)} \ldots A^n_{\rho(n)} B^1_{\sigma(1)} \ldots B^n_{\sigma(n)}\\ &= \det(A)\det(B).\end{aligned}$$

This implies in particular that $\det(A^{-1}) = (\det(A))^{-1}$. Hence, the determinant of $A = (v_1, \ldots, v_n)$ differs from zero if and only if the v_i constitute a basis which is equivalent to invertibility of A. Show that the inverse of

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is provided by

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The reader is advised to explicitely write out the determinant for 3×3 matrices as well as to develop a suitable formula for the inverse of such matrix.

Now, we return to the study of the classification of matrices in "standard form" by means of a basis transformation, the so called Cartan problem which requires the proof of existence of eigenvalues λ as well as associated eigenvectors v_λ satisfying

$$A(v_\lambda) = \lambda v_\lambda.$$

Another way of phrasing this is to say that the nucleus of $A - \lambda 1_n$ is nontrivial which is true if and only if

$$\det(A - \lambda 1_n) = 0.$$

At this point, determinants become useful because this formula can be interpreted as a root equation for a polynomial of the n'th degree. As we know, this polynomial can be entirely factorized over the field of complex numbers $\mathbb{C}$ and we obtain in general n distinguished complex eigenvalues showing that almost any matrix can be diagonalized. The reader is now advised to consider the previous example in two dimensions where two eigenvalues coincide and consider further examples of operators of a higher nil-potency in three or more dimensions.

The vigilant reader has meanwhile noticed that that determinant of a matrix is a basis invariant and henceforth associated to a linear mapping; that is,

$$\det(O^{-1}AO) = \det(O)^{-1}\det(A)\det(O) = \det(A).$$

Therefore, it is noticed that the eigenvalue polynomial $\det(A - \lambda 1_n)$ is an operator invariant. In particular, it is shown that the functional coefficient of k'th degree corresponding to the $n-k$'th power of λ constitutes an invariant under basis transformations. For $k = 1$ this gives $(-1)^{n-1}\mathrm{Tr}(A)$ where the so called trace Tr is defined by means of

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A^i_i.$$

Verify as an exercise in an explicit way that the trace is indeed a basis invariant and study the specific functional form of the higher invariants as well. One might try to write those as polynomials of traces of powers of the matrix; in particular in two dimensions it holds that

$$2\det(A) = (\mathrm{Tr}(A))^2 - \mathrm{Tr}(A^2).$$

Show that, in case one replaces the real number λ by the matrix A in the eigenvalue polynomial that it holds then that the resulting matrix equals the zero matrix. This is known as the theorem of Cayley Hamilton (hint: suppose first that A can be diagonalized and use then the definition of an eigenvalue as a root of the eigenvalue polynomial and finally employ that any matrix can be arbitrarily well approximated by one which is diagonalizable) which reads in two dimensions as

$$A^2 - \mathrm{Tr}(A)A + \det(A)1_2 = 0.$$

Show finally that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and that this implies that no pair of matrices A, B exists such that

$$AB - BA = 1_n$$

a formula which is known as the bosonic Heisenberg relation and requires an infinite number of dimensions for operators having infinite traces. Note that it is possible to find two by two matrices such that

$$AB + BA = 1_2$$

known as the fermionic Heisenberg relationship. Bosons require henceforth an infinite number of dimensions whereas fermions live in dimensions equal to $n = 2^d$ where the reader should find a realization for $d = 1$. Finally, we define the notion of transposition A^T as well as the complex conjugation $\overline{A}$ of a matrix A

$$(A^T)^i_j = A^j_i, \; (\overline{A})^i_j = \overline{A^i_j}.$$

Show that

$$(AB)^T = B^T A^T, (A^T)^T = A, (rA + sB)^T = rA^T + sB^T, (A^{-1})^T = (A^T)^{-1}$$

and similar properties for the complex conjugation. The hermitian conjugate, which is of fundamental importance in this book, is given by $A^\dagger = \overline{A}^T$ and the reader may verify that

$$(AB)^\dagger = B^\dagger A^\dagger, (A^\dagger)^\dagger = A, (rA + sB)^\dagger = \overline{r}A^\dagger + \overline{s}B^\dagger, (A^{-1})^\dagger = (A\dagger)^{-1}$$

In particular, it holds that for

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

$$A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

swapping rows as well as columns. Prove that for

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

it holds that $N^T N + N N^T = 1_2, N^2 = 0$ giving rise to the designation that N constitutes a fermionic creation-operator. This suffices for a first encounter with linear spaces and operators; the next chapter treats the subject in more depth and we continue now with succinct exercises.

Exercises regarding Hermitian projection operators.

- Let P, Q be two Hermitian projection operators meaning that $P^2 = P$, $Q^2 = Q$, $P^\dagger = P, Q^\dagger = Q$. Show that $P + Q$ constitutes a Hermitian projection operator if and only if $PQ = QP = 0$. Show that the same holds for PQ if and only if $PQ = QP$.
- Two Hermitian projection operators P, Q are orthogonal if and only if $PQ = 0$; we define the partial order $\leq$ by means of $P \leq Q$ if and only if $QP = PQ = P$. Prove explicitly that $\leq$ defines a partial order on the set of Hermitian projection operators. In particular, it holds that $P \leq Q$ and $Q \leq P$ implies that $P = Q$. Also, $P \leq Q$ and $Q \leq R$ leads to $P \leq R$.
- We call the set of Hermitian projection operators on a vector space, equipped with $\leq$, a raster. Show that for any P, Q there exists a minimal projection operator $P \vee Q$ such that $P, Q \leq P \vee Q$ and any R such that $P, Q \leq R$ satisfies $P \vee Q \leq R$. On the other hand, one may construct a maximal projection operator $P \wedge Q \leq P, Q$. Show that $\vee, \wedge$ do not in general obey the rule of de Morgan:

$$P \wedge (R \vee Q) \neq (P \wedge R) \vee (P \wedge Q).$$

- Show that the raster possesses a unique minimum as well as maximum provided by 0 and 1 respectively.
- Show that there exist minimal nonzero Hermitian projection operators, called atoms. Every Hermitian projection operator may be written as a sum of orthogonal atoms.

Quantum logic.
Given that in the previous exercise $\vee$ and $\wedge$ may be conceived as "or" and "and" respectively, it becomes possible to understand quantal logic by means of using Hermitian projection operators as propositions. Reflect on this and retrieve classical pointer propositions.

Hilbert space.
Let v and w be two complex vectors and denote by

$$\langle v|w\rangle = v^\dagger w \in \mathbb{C}$$

the so-called scalar product of v and w. Prove that

$$\langle v|w\rangle = \overline{\langle w|v\rangle},\ \ \langle v|v\rangle \geq 0 \text{ and equality holds if and only if } v = 0$$

$$\langle v|aw + bz\rangle = a\langle v|w\rangle + b\langle v|z\rangle$$

and the reader verifies that these equalities imply that

$$\langle av + bz|w\rangle = \overline{a}\langle v|w\rangle + \overline{b}\langle z|w\rangle.$$

As a challenging exercise, the reader proves that

$$|\langle v|w\rangle| \leq ||v||||w||$$

where $||v|| = \sqrt{\langle v|v\rangle}$. Prove from hereon that

$$||v + w|| \leq ||v|| + ||w||$$

the so-called triangle inequality. Finally, let A be an operator, then show that

$$\langle v|Aw\rangle = \langle A^{\dagger}v|w\rangle.$$

Dirac notation: a vector v is also denoted as $|v\rangle$ and a conjugate vector $v^{\dagger}$ as $\langle v|$ so that $|v\rangle\langle v|$ is the Hermitian projector on v in case $\langle v|v\rangle = 1$.

Non-commutative Quantum logic.
We generalize the operations $\wedge$ and $\vee$ to a context in which they are no longer commutative; this procedure holds as well for the classical Boolean logic or the quantual logic explained above where the de Morgan rule gets a minor blow. It is natural to interpret $\wedge$ as well as $\vee$ as mappings $\wedge, \vee : P \times P \to P : (x, y) \to x \wedge y$, $(x, y) \to x \vee y$ where P denotes the lattice of propositions defined by means of a linear Euclidean space in the quantal case. Define the mapping $S : P \times P \to P \times P : (x, y) \to (y, x)$ and consider $\wedge^{'(V,W)} := W \circ \wedge \circ S \circ V$ as well as $\vee^{'(V,W)} = W \circ \vee \circ S \circ V$ where $V : P \times P \to P \times P$ is required to be invertible as well as is the case for $W : P \to P$. Requiring $\wedge^{'(V,W)}$ to satisfy $\left(\wedge^{'(V,W)}\right)^{'(V,W)} = \wedge$ it is sufficient and mandatory that $W^2 = 1$ as well as $S \circ V \circ S \circ V = 1$. This demand is of a special algebraic nature which we dub by the name of an involution; so we are going to study involutive deviations from quantal logic. An involution gives rise to a notion of duality; in particular self-duality is defined by the condition that

$$\wedge^{'(V,W)} = \wedge, \vee^{'(V,W)} = \vee.$$

It is natural to propose first S symmetrical logics; these are given by

$$\wedge^{'(V,W)} \circ S = \wedge^{'(V,W)}, \vee^{'(V,W)} \circ S = \vee^{'(V,W)}.$$

This can only happen by choosing V such that

$$V \circ S = S \circ V$$

reducing a previous condition to

$$V^2 = 1$$

whereas it still holds that

$$\wedge^{'(V,W)} = W \circ \wedge \circ S \circ V.$$

In case $\wedge$, $\vee$ coincide with the standard Boolean or Quantal operations denoted by $\wedge_d$, $\vee_d$ where $d = c, q$ one has that

$$\wedge_d \circ S = \wedge_d,\ \vee_d \circ S = \vee_d.$$

In such a case,

$$\wedge := \wedge_d^{'(V,W)} = W \circ \wedge_d \circ V$$

a small simplification of the previous formula and $\vee$ is defined in a similar way. Now, to remain entirely clear, it is so that the d index should be the same in $\wedge$, $\vee$ but (V, W) becomes (R, T) for $\vee$ whereas the former pertains to $\wedge$. We now isolate the "de Morgan expression" $a \wedge (b \vee c)$:

$$\wedge \circ (1 \times \vee)(a, b, c) = W \wedge_q V(1 \times T \vee_q R)(a, b, c).$$

It is subsequently natural to call T - $(\wedge_q, V)$ compatible if and only if $\wedge_q V(1 \times T) = T' \wedge_q V$ for some $T' : P \to P$. Likewise, it is natural to call V - $\vee_q$ compatible if and only if $V(1 \times \vee_q) = (1 \times \vee_q)V'$ for some $V' : P^3 \to P^3$. Under these assumptions, the previous expression reduces to

$$WT'(\wedge_q(1 \times \vee_q))V'(1 \times R)$$

which was the desirable separation. It is furthermore natural to suggest further restrictions

$$WT' = 1,\ V'(1 \times R) = 1_3.$$

Truth evaluators ω

The material presented below constitutes an extension of the notes I have received once from Rafael Dolnick Sorkin; in classical Boolean logic one disposes of truth evaluator ω of logical sentences which constitutes a homomorphism from the set of propositions $P, \vee_c, \wedge_c$ to $\mathbb{Z}_2, +, .$ where 0 is interpreted as false and 1 as true and $\vee_c$ is the so called exclusive *or* in the sense that $a \vee_c b$ is true if and only if exactly one of them is true. It is to say that

$$\omega(a \vee_c b) = \omega(a) + \omega(b),\ \omega(a \wedge_c b) = \omega(a)\omega(b).$$

In quantum logic, there is no such thing as a truth evaluator; one can only say whether a particular assertion is true or false with a certain probability. A quantum reality is then a particular choice of mapping from P to $\mathbb{Z}_2$ but it makes no sense any longer to speak about a homomorphism because the de-Morgan rule fails in general: the lattice is not distributive. As such, it may very well be that you have a quantal reality ω for which $\omega(a) = \omega(b) = 1$, but $\omega(a \wedge_q b) = 0$. To get an idea of what more general realities are about, let us describe a classical system in a quantum mechanical fashion. An example is give by means of the weather, "the sun shines", modeled by $|l\rangle$, or "it is dark" given by $|d\rangle$. Quantum mechanically, one disposes of a complex two dimensional Euclidean space spanned by the extremal vectors $|l\rangle, |d\rangle$. Consider now a general state

$$|\psi\rangle = \alpha|l\rangle + \beta|d\rangle$$

and study the class of truth functionals ω which merely depend upon

$$\frac{|\alpha|^2}{|\alpha|^2 + |\beta|^2}, \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2}$$

something which reduces to a parameter $0 \leq \lambda \leq 1$ due to

$$\frac{|\alpha|^2}{|\alpha|^2 + |\beta|^2} + \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2} = 1.$$

When all truth evaluators merely depend upon this parameter only, the complex plane may be reduced to the line segment connecting both extremal vectors $|l\rangle, |d\rangle$ to one and another. An example of such a generalized reality is provided by

$$\omega^l_\epsilon : [0, 1] \to \mathbb{Z}_3$$

given by means of the prescription

$$\omega^l_\epsilon(\sqrt{\lambda}|l\rangle + \sqrt{(1-\lambda)}|d\rangle) = \chi(\lambda + \epsilon - 1) + 2\chi(\lambda - \epsilon)\chi(1 - \epsilon - \lambda).$$

ω^l and is henceforth connected to the question whether the light shines and ϵ is the tolerance of the observer. This truth evaluator says "yes", given by means of 1, in case $1 - \epsilon \leq \lambda \leq 1$, under determined or "vague" 2 when $\epsilon \leq \lambda \leq 1 - \epsilon$ and no, given by 0, when $0 \leq \lambda \leq \epsilon$. We have that χ is the so called characteristic function defined on the real numbers by means of $\chi(x) = 1$ in case $x \geq 0$ and zero otherwise. The issue is that we departed from a quantum mechanical description of the weather and by reduction of the allowed questions arrived to a classical system where, moreover, ω^l_ϵ is nonlinear.

Most physicists would suggest at this moment that we did not make a sufficient distinction between classical and quantum logic as yet because $\wedge_q, \vee_q$ are commutative, assiociative but $\wedge_q$ is not distributive with regard to $\vee_q$ which is the case for $\wedge_c$, $\vee_c$. In our most general setting, one has that $\wedge$ and $\vee$ are neither commutative, nor associative

$$\vee(1 \times \vee)(a, b, c) = T \vee_d R(1 \times T \vee_d R)(a, b, c) \neq T \vee_d R(T \vee_d R \times 1)(a, b, c) = \vee(\vee \times 1)(a, b, c)$$

and likewise so for $\wedge$. The main distinction between classical and quantum logic resides in the fact that the set of propositions constitutes a distributive lattice in the former case whereas it does not in the latter; this results in the statement that the classical rule

$$\mu(a|b)\mu(b) = \mu(b|a)\mu(a)$$

is no longer true in the quantal case. Here, μ is the probability measure that a is true; in other words, the truth determinations of a and b depend upon the order in which they occur. This has so far not been accounted given that a homomorphism $\vee_{c,q}, \wedge_{c,q}$ does not make any distinction in the order of the factors. Therefore, classically, for our homomorphism $\omega_c(a \wedge_c b)$ is determined by the unordered tuple $\{\omega_c(a), \omega_c(b)\}$. Quantum mechanically, it is as such that the reality $\omega_q(a \wedge_q b)$ is not provided by the ordered couple $(\omega_q(a), \omega_q(b))$ as elements of $\mathbb{Z}_2$ but also depends upon a, b themselves. It is not so that

$$\mu_{|v\rangle}(a|b) = \frac{\mu_{|v\rangle}(a \wedge_q b)}{\mu_{|v\rangle}(b)}$$

due to commutativity of $\wedge_q$ as well as $a \wedge_q b = 0$ for distinct one dimensional Hermitian projection operators a, b on a Hilbert space $\mathcal{H}$. The exact formula is given by

$$\mu_{|v\rangle}(a|b) = \frac{\mathrm{Tr}(|v\rangle\langle v|bab)}{\mathrm{Tr}(|v\rangle\langle v|b)}$$

and the reader notices that the non-commutativity of a and b is of vital importance. Henceforth, the ontological mapping defined in quantum theory is given by $\kappa : P \to \mathbf{L}(\mathcal{H})$ where P is the set of prepositions with a yes or no answer onto the lattice of Hermitian projection operators defined on the Hilbert space of states of the system. The classical Lagrange formula

$$\mu(a|b)\mu(b) = \mu(b|a)\mu(a)$$

where μ is determined by the state of the system is abandoned upon provided that $\wedge_q$ a la Von Neumann offers no alternative. The natural question henceforth is whether we may find a natural $\wedge$ as well as a consistent set of realities

$$\omega_q^\rho : P \to \mathbb{Z}_2 \times [0, 1]$$

attached to density matrices ρ defined on $\mathcal{H}$, such that

$$\omega_q^\rho(a) = (1, \lambda)$$

and

$$\omega_q'^\rho(a) := (0, 1 - \lambda)$$

is defined as the complementary observation. It is clear that ω_q is not always given by a homomorphism; prior to proceeding, it is important to understand $\vee_q$. It is clearly so that in quantum theory, we have an extended ontology; we do not only pose the question "what is the probability that $a \wedge_c b$ holds given that a as well as b are true" such as the case in classical logic, but we insist on the formulation "what is the chance that $a \wedge_q b$ holds given that a after b has been experimentally established". The right answer is easy if $a \wedge b$ is represented by the Hermitian operator bab which is logical given that the order of measurement matters. In general, one shows that

$$a \wedge_q b = \lim_{n \to \infty} \left(\frac{1}{2}(ab + ba) \right)^n$$

and in the framework of our deformation theory $\wedge$ is given by means of

$$V(a, b) = (1, bab)$$

at least this is so for atomic elements a, b. For atomic elements, $\frac{bab}{\mathrm{Tr(ab)}}$ is again a rank one Hermitian projection operator; however for projection operators of general rank, this is no longer the case. Here, we have to extend our definition of V as going from $P \times P \to C \times C$ where C are the so called positive operators on Hilbert space. An operator A is positive if and only if A is self adjoint and

$$\langle v|A|v\rangle > 0$$

for all $v \neq 0$. As an exercise, the reader understands that the definition of $\leq$ extends to the Hermitian operators by means of $A \leq B$ if and only if

$$\langle v|(B-A)|v\rangle > 0.$$

Show that in such a case, the definitions of $\wedge_q$ and $\vee_q$ can be extended as the largest Hermitian operator smaller or equal than A, B and the smallest Hermitian operator greater or equal to A, B respectively. As an exercise, the reader understands that the definition of $\leq$ extends to the Hermitian operators by means of $A \leq B$ if and only if

$$\langle v|(B-A)|v\rangle > 0.$$

Show that in such a case, the definitions of $\wedge_q$ and $\vee_q$ can be extended in a rathr aritrary fashion (we shall provide for a construction below). Indeed, it is impossible to define $A \wedge B$ as the largest Hermitian operator smaller or equal than A, B and $A \vee B$ as the smallest Hermitian operator greater or equal to A, B respectively (see the following example). Nevertheless, it is possible to define a weaker notion of maximality for $A \wedge B$ meaning that if $A \wedge B \leq C \leq A, B$ then $C = A \wedge B$. As an example, consider

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}, B = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

the latter having eigenvalues $1, 5$ and the reader may verify that the operators

$$C = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}, D = \begin{pmatrix} b & 0 \\ 0 & \frac{5}{2} \end{pmatrix}$$

for $32b = \frac{89}{4} - (\frac{27}{2})^2$ are both intersections in the second sense but not in the first as the reader may verify. A rather canonical construction hinges on the spectral theorem for Hermitian operators, something which we shall study in the next section. Briefly, it says that any Hermitian operator A can be written as

$$A = \sum_i \lambda_i P_i$$

where the λ_i are the real eigenvalues and the P_i Hermitian projection operators such that $P_i P_j = \delta_{ij} P_i$. Therefore, take A, B and order all eigenvalues

$$\lambda_0 < \lambda_1 \ldots < \lambda_k$$

with $k \leq 2n$ where n is the dimension of Hilbert space. Note that some of the λ_i may belong to A as well as B; in that case, we consider the projection operators $R_i := P_i \vee_q Q_i$ where the Q_i refer to B otherwise R_i equals P_i or Q_i. Start now with λ_0, the smallest eigenvalue, and consider the operator $C_0 = \lambda_0 R_0$; clearly $C_0 \leq A, B$. Proceed now towards the minimal λ_j such that $S_j := \vee_{i=2}^{j} R_i$ obeys $[S_j, R_0] = 0$ and consider the projection operator

$$T_1 := S_j(1 - R_0)$$

then the reader verifies that this is an Hermitian projection operator and that $T_1 R_0 = 0$. In case no such j exists, then define $A \wedge_q B = \lambda_0 R_0 + \lambda_1(1 - R_0)$, otherwise proceed with $C_1 := \lambda_0 R_0 + \lambda_1 T_1$. The reader now understands that he has to look at λ_{j+1} and construct the smallest $S_k := \vee_{i=j+1}^{k} R_i$ such that

$$[S_k, R_0 + T_1] = 0.$$

In case no such k exists $A \wedge_q B = \lambda_0 R_0 + \lambda_1 T_1 + \lambda_{j+1}(1 - R_0 - T_1)$ otherwise we consider

$$C_2 = \lambda_0 R_0 + \lambda_1 T_1 + \lambda_{j+1} T_2$$

where $T_2 = S_k(1 - R_0 - T_1)$ and the procedure continues. It is obvious that the final result is no necessarily an optimal Hermitian operator which is smaller or equal to both A, B. The construction of $\vee_q$ is similar, but then one starts at the largest eigenvalue of both operators. W is henceforth determined on the rank 1 matrices by means of the identity. Therefore, for rank one projectors a, b it holds that

$$a \wedge b = T \circ \wedge_q \circ R(a, b) = bab.$$

Subsequently, one has that

$$\omega_q^\rho(a) = (1, \mathrm{Tr}(\rho a))$$

or

$$\omega_q^\rho(a) = (0, 1 - \mathrm{Tr}(\rho a))$$

for a of rank one. Clearly, by definition

$$\omega_{q,1}^\rho(a|b) := \frac{\pi_2(\omega_q^\rho(a \wedge b))}{\pi_2(\omega_q^\rho(b))}$$

equals the probability that a is measured after b. Here π_j equals the projection on the j'th factor. Elaborate further on this theory and determine a suitable $\vee$ operation. Hint: the latter is cannot be given by $a \vee b = a + b$ in the deformation framework provided that $\vee_q$ does not allow one to determine the projection of a on b as is given by $\mathrm{Tr}(ab)$. This is something which is mandatory to extract the sum operation. To define $\vee$ it is advised to use the classical rule

$$\neg(a \vee_c b) = (\neg a) \wedge_c (\neg b)$$

and using $\neg\neg = 1$, it holds that

$$a \vee b = \neg((\neg a) \wedge (\neg b)).$$

In quantum theory, $\neg(a)$ is provided by $1 - a$ and henceforth, we arrive at

$$a \vee b = 1 - (1 - a) \wedge (1 - b)$$

which leads to a violation of the de Morgan rule given that

$$a \wedge (b \vee c) = a \wedge (1 - (1 - b) \wedge (1 - c)) = (1 - (1 - c)(1 - b)(1 - c))a(1 - (1 - c)(1 - b)(1 - c))$$

whereas

$$(a \wedge b) \vee (a \wedge c) = 1 - (1 - cac).(1 - bab).(1 - cac).$$

General exercise.
Determine matrix representations of deformed logic's in terms of commutative albeit possible non-associative ones. It is to say that

$$\wedge = (\tilde{\wedge}_{ijk})_{i,j,k:1\ldots n}$$

where

$$\tilde{\wedge}_{ijk}(a_j, b_k) = \tilde{\wedge}_{ijk}(b_k, a_j)$$

constitute S symmetrical logics on the product space $\times_n P$ where P provides for elementary propositions. Classify first the S symmetric deformations of Boolean logic on general proposition sets.

Chapter 7

Hilbert Spaces and some Important Theorems

Until now, we have been silent about the subject of topologies on linear spaces as well as on spaces of linear operators defined upon the former. The reason is very simple: all such spaces have been equivalent to $\mathbb{R}^n$ from the set theoretical point of view and all "natural" topologies which spring to ones mind are equivalent to the in product norm topology. In a countable infinite number of dimensions, these topologies become in-equivalent and we shall study those at an early stage of this chapter. We shall commence with studying and in product geometry and see how it connects to probability theory as well as topology: the philosophy then is that such flat geometry precedes all these concepts in a well defined sense.

There exist many distinguished means of presenting the material below but I am of the opinion that the succinct presentation beneath is the most efficient one. A distinguished feature of Euclidean geometry is that the underlying set is given by means of linear space, this is no longer true when studying curved geometry. This very feature shaped a too limited characterization for two thousand years of several geometrical concepts such a the one of an oriented line segment connecting two points x, y. The old view was that those could be connected by means of a *free* vector $y - x$ which is then assumed to be "tight" to the point x. Crucial herein is the minus sign as an operation suggesting that it is possible to add vectors without caring about their "anchoring" to particular points. Mathematically, this results in the notion of a linear space with the zero displacement 0 as a neutral element mistaken dubbed as the "origin" of the latter space. This preferred origin has been long subject of "theological debate" which has its philosophical side too: is earth the center of the universe which never in motion? Or must one speak about the sun or another heavenly body in this regard? Newton and his friends were the first to cut the Gordian knot: they introduced the concept of an affine space by allowing for translations removing any preference of origin whatsoever. Indeed, the mapping $x \to x + a$ does not commute with the addition given that

$$(x + y) + a \neq (x + a) + (y + a).$$

It leaves however the difference invariant in the sense that

$$(y + a) - (x + a) = y - x$$

such that vectors, bound or free, have a significantly distinct status from points. In Newton's world, nothing is fixed and that was a grand realization by itself. Mathematicians such as Gauss, Riemann and Cartan did proceed even further on: modern cosmos has no translation symmetry any longer and cannot be described any more in the language of affine spaces. The importation of these realization into physics has been the great achievement by Albert Einstein by means of theory of general relativity which constitutes by far a superior explanation behind everyday large scale observations in the universe. Euclidean space or an (in)finite dimensional flat geometry is defined henceforth by means of a real vector space $\mathcal{H}$ as well as scalar product $\langle v|w\rangle$ where $v, w \in \mathcal{H}$. The scalar product between v and w is supposed to be equal to the product of the oriented length of the projection of w upon v times the length of v. This quantity satisfies, by means of simple experience, the following properties:

$$\begin{aligned}\langle v|w\rangle &= \langle w|v\rangle \\ \langle v|aw+bu\rangle &= a\langle v|w\rangle + b\langle v|u\rangle \\ \langle v|v\rangle &\geq 0 \text{ where equality holds if and only if } v=0.\end{aligned}$$

The scalar product henceforth determines the notion of perpendicularity; the very fact that we have here on earth a preferred notion of perpendicularity is of a *physical* nature. Albert Einstein discovered that this information is encoded partially into the gravitational field. It could be that an alien would experience this gravitational field differently and that it would suggest a different local geometry. One can speak about complex geometries: in such a case, one defines in exactly the same fashion a sesquilinear form where now

$$\langle v|w\rangle = \overline{\langle w|v\rangle}$$

with the complex conjugation defined as usual by means of

$$\overline{a+bi} = a - bi.$$

For example, $\mathbb{C}$ constitutes a one dimensional Hilbert space with as scalar product $\overline{v}w$. As stated in the introduction, a Hilbert space carries some natural topologies; to define those, we show that the scalar product defines in a canonical fashion a metric d. We first prove that the quantity $||v||$ defined by

$$||v|| = \sqrt{\langle v|v\rangle}$$

and called a norm has identical properties to those of a modulus of a complex number. An important step herein is the so called Cauchy-Schwartz identity

$$|\langle v|w\rangle| \leq ||v||||w||$$

signifying that the projection of w on v multiplied with the length of v is less or equal to the product of the lengths of v and w, a result one expects to hold trivially. The formal proof goes as follows:

$$0 \leq ||v+\lambda w||^2 = ||v||^2 + |\lambda|^2\, ||w||^2 + 2\text{Re}\left(\overline{\lambda}\langle w|v\rangle\right)$$

where $\text{Re}(a+ib) = a$ is the real part of the complex number $z = a+bi$. One verifies that the real part of the complex number z may be written as $\frac{1}{2}(z+\overline{z})$ whereas the imaginary part equals $-i\frac{1}{2}(z-\overline{z})$. The modulus of a complex number is defined by means of

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

and satisfies

$$|z+z'|^2 = |z|^2 + |z'|^2 + (z\overline{z'} + \overline{z}z')$$

whereas the last term equals, up to a factor of two,

$$aa' + bb'$$

and the absolute value is bounded from above by $|a|\,|a'| + |b|\,|b'|$. The square of this last expression is given by

$$a^2a'^2 + b^2b'^2 + 2\,|a|\,|a'|\,|b|\,|b'| \leq \left(a^2+b^2\right)\left(a'^2+b'^2\right) = |z|^2\,|z'|^2$$

and consequently one has that

$$|z+z'|^2 \leq (|z|+|z'|)^2$$

and hitherto

$$|z+z'| \leq |z| + |z'|$$

a formula known as the triangle inequality. Consequently, we may define a metric on the complex plane by means of

$$d(z,z') = |z-z'|\,.$$

Returning to the proof of the triangle inequality, one notices that we may pick λ such that

$$\mathrm{Re}\left(\overline{\lambda}\langle w|v\rangle\right) = -\left|\lambda\right|\left|\langle v|w\rangle\right|$$

whereas, in general, the left hand side is always larger than the right hand side. Therefore, we have that

$$0 \leq ||v||^2 + \left|\lambda\right|^2 ||w||^2 - 2\left|\lambda\right|\left|\langle v|w\rangle\right|$$

which is a quadratic polynomial inequality in the positive variable $|\lambda|$. The existence of at most one positive root demands that

$$0 \leq 4\left|\langle v|w\rangle\right|^2 - 4||v||^2||w||^2$$

which proves the result and equality only holds if and only if $w = -\lambda v$. Consequently,

$$||v+w||^2 \leq ||v||^2 + ||w||^2 + 2\left|\langle v|w\rangle\right| \leq ||v||^2 + ||w||^2 + 2||v||||w|| = (||v|| + ||w||)^2$$

which proves the triangle inequality for the norm. Consequently, each Hilbert space $\mathcal{H}$ defines a canonical metric topology by means of

$$d(v,w) = ||v-w||$$

and we demand that $\mathcal{H}$ is complete in this topology. This condition is extremely important for the theory of linear operators but let us start by making some preliminary observations. Two non-zero vectors v, w are perpendicular to one and another if and only if $\langle v|w\rangle = 0$ and we say v is normed if and only if $||v|| = 1$. Due to the axiom of choice, any Hilbert space has an orthonormal basis $(e_i)_{i\in I}$ meaning $\langle e_i|e_j\rangle = \delta_{ij}$ where δ_{ij} equals 0 if $i \neq j$ and 1 otherwise. The mindful reader notices that δ^i_j constitutes a basis invariant whereas δ_{ij} is only invariant under orthogonal or unitary transformations. For finite dimensional Hilbert spaces, one has that, with $v = \sum_{i=1}^n v^i e_i$, it holds

$$\langle v|w\rangle = \sum_{i,j=1}^n \overline{v^i} w^j \delta_{ij}$$

which constitutes a generalization of the standard in product in three dimensional Euclidean geometry. Show that by means of a basis transformation $e'_i = O^j_i e_j$ we have that $\delta'_{ij} = \langle e'_i|e'_j\rangle = \overline{O}^k_i O^l_j \delta_{kl}$. Exercise: define Hilbert spaces over the real quaternions.

We now consider some operations or constructions one can perform with real or complex Hilbert spaces. The best known ones are applied in the theory of quantum mechanics and are given by the tensor product $\otimes$ as well as direct sum $\oplus$. Given two Hilbert spaces $\mathcal{H}_i$, the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ constitutes again a Hilbert space spanned by pure vectors $v_1 \otimes v_2$ where $v_i \in \mathcal{H}_i$. Regarding sums $\sum_{i=1}^n z_i v^i \otimes w^i$, the following equivalences are in place

$$\begin{aligned} z(v\otimes w) &\equiv (zv)\otimes w \equiv v \otimes (zw) \\ v\otimes w_1 + v\otimes w_2 &\equiv v\otimes(w_1 + w_2). \end{aligned}$$

We define $\mathcal{H}$ as the linear space of such equivalence classes and make a completion in the metric topology defined by means of the scalar product

$$\langle v_1 \otimes w_1 | v_2 \otimes w_2\rangle := \langle v_1|v_2\rangle\langle w_1|w_2\rangle.$$

In a similar vein, the direct sum $\mathcal{H}_1 \oplus \mathcal{H}_2$ is defined by means of the equivalences

$$\begin{aligned} z(v\oplus w) &\equiv (zv)\oplus(zw) \\ v_1\oplus w_1 + v_2 \oplus w_2 &\equiv (v_1+v_2)\oplus(w_1+w_2) \end{aligned}$$

with as scalar product

$$\langle v_1 \oplus w_1 | v_2 \oplus w_2\rangle := \langle v_1|v_2\rangle + \langle w_1|w_2\rangle.$$

One verifies that a basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ is provided by means of $v_i \otimes w_j$ where the v_i constitute a basis of $\mathcal{H}_1$ and w_j of $\mathcal{H}_2$. A basis for $\mathcal{H}_1 \oplus \mathcal{H}_2$ is provided by $v_i \oplus 0, 0 \oplus w_j$.

In a vector spaces, a basis defines a scalar product and the mapping of bases to Hilbert spaces is surjective. Bases connected by means of a transformation O satisfying

$$\overline{O}^k_i O^l_j \delta_{kl} = \delta_{ij}$$

determine the same scalar product and reversely alike scalar products define separate bases connected by such a transformation. One verifies that those matrices constitute a group, $U(n)$ for n dimensional complex Hilbert spaces and $O(n)$ in the real case, the so called unitary respectively orthogonal groups. The above formula reads in matrix language

$$O^H O = 1$$

whereas $O^H = (\overline{O})^T = \overline{(O^T)}$. Show that in two dimensions the unitary matrices are explicitly given by

$$O = \frac{1}{\sqrt{|a|^2 + |b|^2}} \begin{pmatrix} a & -\overline{b} \\ b & \overline{a} \end{pmatrix}$$

with $a, b \in \mathbb{C}$. This group has three real parameters; the reader is advised to determine an alike representation for $O(2)$. Given linear operators $A : \mathcal{H}_1 \to \mathcal{H}_3$ and $B : \mathcal{H}_2 \to \mathcal{H}_4$ then we may define operators

$$A \oplus B : \mathcal{H}_1 \oplus \mathcal{H}_2 \to \mathcal{H}_3 \oplus \mathcal{H}_4$$

as well as

$$A \otimes B : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \mathcal{H}_3 \otimes \mathcal{H}_4$$

by means of

$$A \oplus B(v_1 \oplus v_2) = A(v_1) \oplus B(v_2)$$

and

$$A \otimes B(v_1 \otimes v_2) = A(v_1) \otimes B(v_2).$$

The reader should reflect for a moment and convince himself that $\otimes$ serves for the purpose of combining separate systems; it is to say functions in n real variables $f_k : (x_1, \ldots, x_n) \to \mathbb{C}$ and m real variables $g_k : (y_1, \ldots, y_m) \to \mathbb{C}$ define functions in $n + m$ real variables by means of

$$F = \sum_k a_k (f_k \otimes g_k) : \mathbb{R}^{n+m} \to \mathbb{C} : (x_1, \ldots, x_n, y_1 \ldots, y_m) \to$$

$$\sum_k a_k f_k(x_1, \ldots, x_n) g_k(y_1, \ldots, y_m).$$

Here, one should not regard $\mathbb{R}^{n+m}$ as a vector space but as a set ; in the vector space language, it holds that $\mathbb{R}^{n+m} = \mathbb{R}^n \oplus \mathbb{R}^m$. It is a result from real analysis that $F : \mathbb{R}^{n+m} \to \mathbb{C}$ may be written as $\sum_k a_k (f_k \otimes g_k)$; in other words, one has a complex vector space of functions $L_2(\mathbb{R}^{n+m})$ which equals $L_2(\mathbb{R}^n) \otimes L_2(\mathbb{R}^m)$.

One now makes the following exercises: be $A : V \to V$ and $B : W \to W$ operators on finite dimensional vector spaces; show that

$$\mathrm{Tr}(A \oplus B) = \mathrm{Tr}(A) + \mathrm{Tr}(B),\ \mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B)$$

and

$$\det(A \oplus B) = \det(A)\det(B),\ \det(A \otimes B) = \det(A)^m \det(B)^n$$

where $n = \dim(V)$ and $m = \dim(W)$. In case V, W constitute moreover Hilbert spaces; show that

$$(A \oplus B)^H = A^H \oplus B^H,\ (A \otimes B)^H = A^H \otimes B^H.$$

Prove that the operations $\oplus, \otimes$ are associative with $\{0\}, \mathbb{C}$ as identity element respectively; denote with $\otimes_{\mathcal{F}}$ the mapping on the space of Hilbert spaces defined by $\otimes_{\mathcal{F}}(\mathcal{H}) = \mathcal{H} \otimes \mathcal{F}$. Construct a $i_{\mathcal{F}}$ such that $i_{\mathcal{F}} \circ (\otimes_{\mathcal{F}}) = \mathrm{id}$ where id is given by the identity transformation. Show that $\otimes_{\mathcal{F}}$ is not surjective unless $\mathcal{F} = \mathbb{C}$ which shows that there does not exist any $p_{\mathcal{F}}$ obeying $(\otimes_{\mathcal{F}}) \circ p_{\mathcal{F}} = \mathrm{id}$. Make a similar construction for $\oplus_{\mathcal{F}}$ and notify that nor $\oplus, \otimes$ are commutative. Here, we have found an example of a mapping, derived from an operation, with a left but no right inverse. Introduce now the concept of an anti-Hilbert space $\mathcal{F}^{\otimes}$ as a formal right inverse for $\mathcal{F}$; it is to say that

$$\mathcal{F} \otimes \mathcal{F}^{\otimes} = \mathbb{C}.$$

In that case $i_{\mathcal{F}}$ equals $\otimes_{\mathcal{F}^{\otimes}}$ on the image of $\otimes_{\mathcal{F}}$. This procedure is entirely analogous to taking negative integers or fractions starting from the natural numbers. Do the same for $\oplus$ and reflect further hereupon. More in particular, denote with a_i bosonic annihilation operators defined by

$$a_i a_j^\dagger - a_j^\dagger a_i = \delta_{ij} 1$$

and posit that

$$\mathcal{F} \equiv \{v = \sum_{i=1}^{\infty} \lambda^i a_i | \langle 0 | v v^\dagger | 0 \rangle < \infty \,\text{with scalar product}\, \langle v | w \rangle = \langle 0 | w v^\dagger | 0 \rangle \}$$

where $|0\rangle$ constitutes the so called Fock vacuum defined by $a_i|0\rangle = 0$. $\mathcal{F}^{\otimes}$ equals then for example

$$\{v = \lambda^j a_j^\dagger | \text{with as scalar product}\, \langle v | w \rangle = \langle 0 | v^\dagger w | 0 \rangle \}$$

such that

$$v \in \mathcal{F} \otimes \mathcal{F}^{\otimes}$$

is given by $\sum_i \lambda_i \mu_j a_i a_j^\dagger$. The scalar product is given by

$$\sum_i |\lambda_i|^2 |\mu_j|^2 \langle 0 | a_j a_i^\dagger a_i a_j^\dagger | 0 \rangle$$

which equals $|\lambda_j|^2 |\mu_j|^2$. Therefore, all modes in $\mathcal{F}$ with $i \neq j$ are killed such that the positive norm requirement is restored. These phantoms need to be eliminated with the purpose of retaining a one dimensional space.

Remark that this non-commutative "product" also appeared in set theory by means of $\times$. More precisely, given a set A, an anti-set obeys

$$A \times A^{\times} = \{1\}$$

where the last one is a set with one element 1 and henceforth serves as the identity element for $\times$. To represent an anti-set in the set like fashion; denote that if $A = \{x | x \in A\}$ and $A^{\times} = \{\omega_A^\star\}$ where $\omega_A : A \to \{1\}$ is the constant mapping onto 1 and $\star$ is the associated duality relation, then

$$A \times \{\omega_A^\star\} = \omega_A(A) = \{1\}.$$

Later on, the reader shall deepen his understanding of the fact that Hilbert spaces are employed in physics to describe separated entities such as elementary particles whereas the concept of an anti-Hilbert space can be used to describe particle collisions to create novel types of particles. To collide or not to collide could be mere approximations due to the point description of a particle and the reader is invited, as an exercise of colossal difficulty, to search for a concept of touching.

With this knowledge at hand, it becomes possible to solve standard problems from flat geometry; very strong results are possible here which do not hold in general due to topological as well as metrical complications. The magic of flat geometry is entirely hidden into the vector space structure. For example, on the surface of a ball, any two straight lines, defined as the intersection of the spherical surface with a two dimensional plane containing the barycenter of the sphere, intersect at a length of pi times the radius. In the two dimensional plane on the other hand, there exists a preferential class of parallel lines defined by the property that they do not intersect. In the three dimensional Euclidean space, we call a two dimensional space a plane, a one dimensional a line and a zero dimensional one a point. In Euclidean space, there is only one zero dimensional subspace constituting the neutral element for the addition, denoted by $\{0\}$, also called the origin. A straight line or geodesic is parametrized as follows $r = \{\lambda.v + a | \lambda \in \mathbb{R},\ v, a \in \mathbb{R}^3\}$ and a plane as $vl = \{\lambda.v + \mu.w + a | \lambda, \mu \in \mathbb{R},\ v, w, a \in \mathbb{R}^3\}$ where the free vectors v, w can be chosen to be orthonormal. A straight line can always be written as the intersection of two planes and a plane is completely determined by means of a point and a perpendicular vector. To understand this at a higher level, we introduce the totally anti-symmetrical symbol ϵ_{ijk} where $\epsilon_{123} = 1$ and $\epsilon_{ijk} = \text{sign} \begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix}$ which is merely a convenient notation for the sign of a permutation mapping 1 to i, 2 to j and 3 to k. Henceforth, in this notation,

$$\det(A) = \epsilon_{ijk} A_i^1 A_j^2 A_k^3$$

for a 3×3 matrix A. Here i, j, k constitute indices with respect to vectors belonging to an orthonormal basis and therefore, the ϵ symbol has a geometrical significance. Indeed, $\delta^{ik} \epsilon_{klm} v^l w^m = (v \times w)^i$ is a vector which is orthogonal to v, w (use the anti-symmetry for that) and δ^{kl} is the inverse of the δ_{ij} symbol. It is to say that

$$\delta^{ik}\delta_{kj} = \delta^i_j, \;\; \delta_{ik}\delta^{kj} = \delta^i_j.$$

The square length

$$(v \times w)^2 = \epsilon_{lmn}\delta^{li}\epsilon_{ijk}v^m v^j w^n w^k = (\delta_{mj}\delta_{nk} - \delta_{mk}\delta_{nj})v^m v^j w^n w^k$$

which equals

$$v^2 w^2 - (\langle v|w\rangle)^2$$

and this has the geometrical significance of the surface squared of the parallelipid spanned by the vectors v, w. Henceforth, we have construed a unit vector

$$n = \frac{v \times w}{||v \times w||}$$

perpendicular to the two dimensional subspace spanned by v, w equipped with an orientation such that v rotates right handedly into w. The plane

$$vl = \{\lambda.v + \mu.w + a|\lambda, \mu \in \mathbb{R}, \;\; v, w, a \in \mathbb{R}^3\}$$

then consists precisely out of the points x satisfying the equation

$$\langle n|x - a\rangle = 0$$

which is a linear system in x. In this case $x = (x_1, x_2, x_3)$ satisfies an equation of the form

$$n_1(x_1 - a_1) + n_2(x_2 - a_2) + n_3(x_3 - a_3) = 0.$$

Determine the vector perpendicular to the plane determined by $2x - 3y + z - 12 = 0$ and compute the point which is the closest to the origin.

Other important equations are given by the so called quadratic equations with as an important example, the n sphere. The latter is defined as the set of all points x located at a fixed distance r from the point a. The corresponding equation is given by

$$||x - a||^2 = r^2$$

which reduces in three dimensions to

$$(x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_3 - a_3)^2 = r^2.$$

In exactly the same way, the equation of a circle embedded in $\mathbb{R}^2$ is provided by

$$(x - a)^2 + (y - b)^2 = r^2.$$

Show in two different ways that the intersection of the two sphere with a plane in three spatial dimensions is either empty, a point or a circle. Prove that the same result holds for the intersection of two spheres. These properties are not valid any longer for so called curved geometries which we shall study later on. In a similar vein, we shall study the concept of a triangle as well as some theorems regarding properties of them in general curved geometries for which the flat case is of special symmetric nature. Due to the symmetry, extremely sophisticated results exist in flat geometry: old books will serve the reader well who is willing to study those. I am however of the opinion that at this point it is much more important to understand the general setup which reveals the "true" inner workings of general geometry. This is indeed much more gratifying than becoming a specialist in studying linear and quadratic equations, an art which can be further generalized, in an intermediate step towards analytic geometry, provided by algebraic geometry. We now elaborate further on simplicial geometry which is so called piece-wise flat. Show that the number of simplices $a(n)$ in which an n dimensional cube can be partitioned equals $n!$. A simple proof consists in showing that the volume of an n simplex is given by $\frac{1}{n!}$; indeed, the volume of a n dimensional simplex with length r is given by $b(n)r^n$. Henceforth,

$$b(n+1) = b(n)\int_0^1 dr r^n = \frac{b(n)}{n+1}$$

which proves by iteration that $b(n) = \frac{1}{n!}$.

This wraps up our discussion about Hilbert spaces; we now return to an elaboration on the theory of linear operators as well as delicate topologies defined on such algebra's. This subject is of extreme importance regarding the old operational formulation of quantum mechanics construed by Heisenberg, Jordan and associated gangsters such as Von Neumann. First, we study two distinct topologies on general Hilbert spaces $\mathcal{H}$ prior to engaging into further discussion of the space of linear operators. On $\mathcal{H}$, we did study the norm topology and one proves now the vericacity of the following two statements:

- A set in a finite dimensional Hilbert space is compact in the norm topology *if and only if* it is closed and bounded; hence, the reader must prove the reverse of the more general result in metrical spaces that a compact space is closed and bounded.
- In an infinite dimensional Hilbert space with countable basis $(e_n)_{n \in \mathbb{N}}$ we have that the unit sphere is no longer compact in the norm topology. Hint: argue briefly that the sequence $(e_n)_{n \in \mathbb{N}}$ has no convergent sub-sequence.

We now arrive at a weaker topology having all advantages of the finite dimensional norm topology and which coincides with the latter in the finite dimensional case. It is clear that the norm topology is too strong in infinite dimensions and we require a weaker one spanned by linear functionals ω, defined as mappings from $\mathcal{H}$ to $\mathbb{C}$, a one dimensional lens through which one perceives the Hilbert space. The space of linear functionals constitutes a vector space called the algebraic dual; we are merely interested in those functionals which are continuous in the norm topology. Such functionals constitute again a vector space called the topological dual $\mathcal{H}^\star$. Show that for a finite dimensional Hilbert space, the topological and algebraic dual coincide. An important characterization of continuous functionals is that they are bounded, meaning that

$$|\omega(v)| \leq C||v||$$

for a certain $C > 0$; reversely, it is clear that any bounded linear functional is continuous in the norm topology. We shall give a proof of the former statement: assume that the functional is *not* bounded, then our task is to show that it is not continuous either. More in particular, there exists a sequence of unit norm vectors v_n such that $\omega(v_n) \to \infty$ in the limit for n to ∞. By taking a sub-sequence, we may assume that $\omega(v_n) > n^2$ and the sequence of vectors $w_k = \sum_{n=0}^{k} \frac{1}{n^2} v_n$ converges to $w = \sum_{n=0}^{\infty} \frac{1}{n^2} v_n$ of finite norm (show that the sequence $\sum_{n>0} \frac{1}{n^2}$ converges) whereas $k < \omega(w_k) \to \infty$ in contradiction to continuity.

Because a continuous linear functional provides one with a one dimensional view upon Hilbert space, it has to coincide with a projection on a vector v. It is to say that

$$\omega(w) = \langle v | w \rangle$$

with $||v|| < \infty$ and the reader is encouraged to provide for a formal proof of this theorem. This viewpoint is evident from the geometrical view given that ω is completely determined by means of its nucleus $W = \{w | \omega(w) = 0\}$ as well as the action upon its normal vector $\frac{v}{||v||}$. This motivates the following definition, the sets

$$\mathcal{O}_{\epsilon; v_1, \ldots, v_n}(w) = \{w' \, | \, |\langle w - w' | v_i \rangle| < \epsilon \, \text{for} \, i = 1 \ldots n\}$$

constitute open neighborhoods of w in dimensions determined by v_j and constitute a basis for the *weak* or $\star$-topology.

Open neighborhoods of w in the weak topology control henceforth the modulus of the projection of the difference vector $w - w'$ on a finite dimensional subspace and leave the components perpendicular to it invariant. Given that the norm topology controls all dimensions, it is therefore stronger as the weak one is; in particular, every open set in the weak topology is open in the strong one, a result which follows from the Cauchy Schwartz inequality

$$|\langle w - w' | v_i \rangle| \leq ||w - w'|| ||v_i||.$$

It is henceforth obvious that the same results hold in the weak topology for all Hilbert spaces and that those coincide with the norm topology in the finite dimensional case. In particular, it holds that a set is compact in any Hilbert space if and only if it is bounded in norm and closed in the weak topology. Show that in case a set is bounded in the norm that it is closed in the weak topology if and only if it is so in the norm topology.

This is obvious given that boundedness controls an infinite number of dimensions leaving one with a finite number and those are controlled by means of the weak topology. Henceforth, the unit sphere is closed and compact in the weak topology but merely closed in the norm topology a result known as the Hahn Banach theorem. The reverse is also true, a set which is compact in the weak topology is always bounded in norm. We leave the proofs of these statements as challenging exercises for the reader.

We shall now deal with topologies on spaces of linear mappings $A : \mathcal{H} \to \mathcal{H}$ as well as prove some important theorems regarding operators having a special geometrical significance such as the unitary operators out of the previous chapter. In particular, we are interested in situations where one disposes of an orthonormal basis of eigenvectors as well as some limitations on the eigenvalues. One disposes of plenty of topologies on specific classes of operators all of which are equivalent in a finite number of dimensions. We start with the supremum norm topology:

$$||A||_{\text{sup}} = \sup_{||v||=1} ||Av||.$$

In case the latter is finite, we call the operator A bounded (which is again equivalent to continuous) and the entire edifice of bounded operators is poured into the framework of so called $C^\star$-algebra's. This theory is an abstraction of the concrete situation delineated below and we are not going to pay too much attention to this given that the operators useful in physics are of an unbounded nature. To deal with those devilish objects, we require weaker topologies to probe them, called the strong and weak $\star$ topologies to name two of them. The first one is defined by means of the open sets

$$\mathcal{O}_{\epsilon;v_1,\ldots v_n}(A) = \{B| \; ||(B-A)v_k|| < \epsilon \, \text{for} \, k = 1 \ldots n\}$$

whereas the latter is defined by means of

$$\mathcal{O}_{\epsilon;v_1,\ldots v_n,w_1,\ldots,w_n}(A) = \{B| \; |\langle (B-A)v_k|w_k\rangle| < \epsilon \, \text{for} \, k = 1 \ldots n\}.$$

One shows that both topologies satisfy the Hausdorff property and that the weak-$\star$ topology is weaker as the strong one.

We first introduce some important notions regarding linear operators on Hilbert spaces. The reader may suspect that some subtleties arise which have to do with infinity and were not present in a finite number of dimensions. For example, operators A do have a domain $\mathcal{D} \subset \mathcal{H}$, which we assume to be dense in the norm topology, on which A is well defined. The adjoint operator $A^\dagger$ of A is then retrieved by means of the following procedure. Consider a subspace $\mathcal{D}^\star$ of vectors v such that

$$|\langle v|Aw\rangle| < C(v)||w||$$

for all $w \in \mathcal{D}$. Then, we have that the functional $w \to \langle v|Aw\rangle$ has a unique continuous extension to $\mathcal{H}$ due to the density of $\mathcal{D}$. We obtain the existence of a vector z such that

$$\langle v|Aw\rangle = \langle z|w\rangle$$

and we define $A^\dagger v = z$ and subsequently it easily follows that $A^\dagger$ is a linear operator. Henceforth, the domain of $A^\dagger$ is given by $\mathcal{D}^\star$. Next cases are of extreme importance:

- $A = A^\dagger$ and $\mathcal{D} = \mathcal{D}^\star$ in which case the operator is self adjoint,
- $AA^\dagger = A^\dagger A$ and $\mathcal{D} = \mathcal{D}^\star$ in which case the operator is normal,
- $UU^\dagger = U^\dagger U = 1$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator is unitary,
- $P^2 = P = P^\dagger$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator constitutes a Hermitian projection.

One verifies that in the finite dimensional case it holds that $A^\dagger = A^H$ and moreover, unitary operators constitute generalizations of $U(n)$. Determine the domain of the operator defined by $Ae_n = ne_n$ for $n \in \mathbb{N}$ where e_m constitutes an orthonormal basis and show that it is dense in $\mathcal{H}$; prove that $\mathcal{D} \subseteq \mathcal{D}^\star$ and that

$A = A^\dagger$ on $\mathcal{D}$. We progress now towards the proof of two different theorems: the first one concerns the extension of a special class of operators to Hermitian ones, where the extension of an operator is a new one with a larger domain coinciding with the old operator on its domain. A second result reveals that a normal operator can be decomposed into sums of scalar multiples of Hermitian projection operators in the weak-$\star$ topology.

The importance of the first theorem resides in the second one; this one states that in the finite dimensional case any normal matrix can be diagonalized with respect to an orthonormal basis of eigenvectors. This last aspect is of primary importance to have a probability interpretation such as is the case in quantum theory. Show, by means of exercise, that in a finite number of dimensions Hermitian operators have only real eigenvalues whereas unitary operators have eigenvalues located on the unit circle in the complex plane. Finally, normal operators can have any complex eigenvalue whatsoever. As said before, one has a connection between unitary and self adjoint operators and in that vein it is easier to deal with the problem of unitary extensions of so called partial isometries V with as domain $\mathcal{D}$ which is not necessarily dense. A partial isometry is defined by means of the property that

$$\langle V(v)|V(w)\rangle = \langle v|w\rangle$$

for all $v, w \in \mathcal{D}$. By means of continuity, we can extend V to the closure $\overline{\mathcal{D}}$ of $\mathcal{D}$ resulting in a unitary mapping between $\overline{\mathcal{D}}$ and $\overline{\text{Im}(V)}$ where $\text{Im}(V) = \{Vw|w \in \mathcal{D}\}$ constitutes the image of V. It must be clear to the reader that only in case the orthogonal complements

$$\mathcal{D}^\perp = \{w|\langle w|v\rangle = 0 \; \forall v \in \mathcal{D}\}$$

and

$$(\text{Im}(V))^\perp$$

have identical dimension that we are in position to extend V to a unitary operator U by means of $W : \overline{\mathcal{D}}^\perp \to (\text{Im}(V))^\perp$ where $U = V \oplus W : \mathcal{H} \to \mathcal{H}$. The reader notices that given a subspace W, $W^\perp$ is closed in the weak and therefore also norm topology; the sub space $W^{\perp\perp} := (W^\perp)^\perp$ is moreover equal to the weak closure of W.

One notices therefore that a partial isometry has many unitary extensions in case the dimensions of the orthogonal complements are the same or none whatsoever in case this is not true. Now, we return to the mapping connecting Hermitian to unitary operators; Von Neumann knew the so called Cayley transform between Hermitian and unitary operators in finite dimensional Hilbert spaces. A self adjoint operator A is mapped to

$$U = (A - i1)(A + i1)^{-1}$$

where $(A \pm i1)$ is invertible in a finite number of dimensions given that $Av = \mp iv$ which has no solution. One understands this by means of observing that

$$\mp i||v||^2 = \langle v|Av\rangle = \langle Av|v\rangle = \pm i||v||^2$$

implying that $v = 0$. Moreover, $(A + i1)$ commutes with $(A - i1)$ leading to unitarity of U as is confirmed by means of a small computation. Von Neumann wondered which conditions A should obey such that U is a partial isometry. In such a case, an extension can be made towards a unitary operator defining a Hermitian one by means of the inverse Cayley transformation:

$$A = -i(U + 1)(U - 1)^{-1}.$$

The operator $A \pm i1$ has to be injective so that it becomes possible to take an inverse which suggests that the conditions $\mathcal{D} \subseteq \mathcal{D}^\star$ and $A = A^\dagger$ on $\mathcal{D}$ have to be obeyed which is coined by the term of a symmetric operator. In the infinite dimensional case, it is not necessarily so that $A \pm i1$ is surjective. The Cayley transform is henceforth a linear mapping

$$U : \text{Im}(A + i1) \to \text{Im}(A - i1)$$

and we have to prove three things : (a) verify that U constitutes a partial isometry (b) close the operator $\text{Im}(A \pm i1)$ and finally (c) verify whether $\text{Im}(A + i1)^\perp$ and $\text{Im}(A - i1)^\perp$ have the same dimension. Regarding (a) one notices that

$$\langle U(A + i1)v|U(A + i1)w\rangle = \langle (A - i1)v|(A - i1)w\rangle = \langle Av|Aw\rangle + i\langle v|Aw\rangle - i\langle Av|w\rangle + \langle v|w\rangle$$

and this last expression equals, using the symmetry of A,

$$\langle Av|Aw\rangle + \langle v|w\rangle = \langle (A+i1)v|(A+i1)w\rangle$$

for all $v, w \in \mathcal{D}$. In the standard literature, one closes the operator A prior to taking the Cayley transformation although this is not mandatory; U extends trivially to an operator

$$U : \overline{\mathrm{Im}(A+i1)} \to \overline{\mathrm{Im}(A-i1)}$$

and one requires (c) to extend U to a unitary operator on $\mathcal{H}$. This last condition may be formulated as

$$\mathrm{Im}(A \pm i1)^{\perp} = \mathrm{Ker}(A^{\dagger} \mp i1).$$

Indeed,

$$\langle w|(A \pm i1)v\rangle = 0$$

for all $v \in \mathcal{D}$ is equivalent to $w \in \mathcal{D}^{\star}$ and

$$\langle (A^{\dagger} \mp i1)w|v\rangle = 0.$$

The latter is true if and only if $(A^{\dagger} \mp i1)w = 0$ because $\mathcal{D}$ is dense in $\mathcal{H}$; by definition it holds that $\mathrm{Ker}(B) = \{w|Bw = 0\}$ which produces the right result.

We have just shown our first deep result: symmetric, densely defined operators have self adjoint extensions if and only if the dimensions of the sub-spaces $\mathrm{Ker}(A^{\dagger} \mp i1)$ are equal to one and another. Now we work towards our second main result regarding normal operators A with as special cases Hermitian and unitary operators. That is, there exists a projection valued measure $dP(z)$ on the complex plane such that in the weak $\star$ topology holds that

$$A = \int_{\mathbb{C}} z\, dP(z).$$

We first encounter here an integral, something which we shall study in full detail in the next chapter; subsequently we shall merely touch upon the most fundamental definitions. Suppose one would like to achieve such a result, then it is logical that the operator $(A - z1)$ is not invertible in case $dP(z) \neq 0$; logically, one has three possibilities:

- $(A - z1)$ is not injective, nor surjective; in such a case z belongs to the discrete spectrum,
- $(A - z1)$ is not injective, but surjective; in such a case z belongs to the residual spectrum,
- $(A - z1)$ is injective, but not surjective; in such a case z belongs to the continuous spectrum.

Regarding normal operators, the reader may first show that the residual spectrum is empty. Note that if A is normal, then $A_z = A - z1$ obeys this property too; moreover, A is injective if and only if $A^{\dagger}$ is also which is equivalent to the statement that $Av = 0$ if and only if $A^{\dagger}v = 0$. Mind that surjectivity of A does not imply surjectivity of $A^{\dagger}$. Suppose that z belongs to the residual spectrum then we have that

$$\langle v|A_z w\rangle = 0$$

for all w implies that $v = 0$ due to surjectivity of A_z. This implies that $\mathrm{Ker}(A_z^{\dagger}) = \mathrm{Ker}(A_z) = 0$ which is the necessary contradiction. Henceforth, we have shown that the residual spectrum is empty. In case z belongs to the discrete spectrum, one can find a unique Hermitian projection operator P_z on $\mathrm{Ker}(A_z)$. P_z commutes with A, $AP_z = P_z A = zP_z$ because $\langle v|AP_z w\rangle = z\langle v|P_z w\rangle = \langle \overline{z}P_z v|w\rangle = \langle A^{\dagger}P_z v|w\rangle = \langle v|P_z Aw\rangle$ and the same commutation relations hold between $A^{\dagger}$ and P_z given that the projector is Hermitian. Moreover, suppose that $z \neq z'$ and both belong to the discrete spectrum, then it holds that $P_z P_{z'} = 0$ which follows from

$$zP_z P_{z'} = AP_z P_z' = z' P_z P_{z'}.$$

This strongly resembles the result we wish to obtain in the sense that on infinite dimensional Hilbert spaces, the discrete spectrum consists at most out of a countable number of points. We procure an example of a bounded linear operator with a compact spectrum (which one can show to be always the case). Given that $Ae_n = \frac{1}{n}e_n$ for $n > 0$ and e_m an orthonormal basis: the discrete spectrum is given by $\{\frac{1}{n}|n \in \mathbb{N}_0\}$ and 0 belongs to the continuous spectrum given that the vector $\sum_{n=1}^{\infty} \frac{1}{n}e_n$ does not belong to the image of A. Henceforth, the continuous spectrum may have "measure zero" and henceforth not contribute to the spectral decomposition.

The continuous spectrum is clearly void for normal operators on finite dimensional Hilbert spaces and the reader shows as an easy exercise that

$$A = \sum_{z\in\sigma_d(A)} zP_z$$

where $\sum_{z\in\sigma_d(A)} P_z = 1$ and $\sigma_d(A)$ denotes the spectre consisting entirely out of discrete eigenvalues. One should get used to the following notation: given a unit vector v, define by means of the expression

$$P = vv^\dagger$$

the operator with as action $Pw = v\langle v|w\rangle$. Prove that P is a rank one Hermitian projection operator with property $AP = zP$ and in particular $Av = zv$ implying that v is an eigenvector. The entire complexity of the theorem regarding the infinitesimal aspect having to do with the integral resides entirely in the treatment of the continuous spectrum in an infinite number of dimensions. We shall not present the matter here at the fullest level of generality because this brings along some technical complications muddling with the main line of argumentation. Note that in the finite dimensional case, we did use the fundamental theorem of complex algebra, namely that every polynomial defined over $\mathbb{C}$ can be factorized.

In case z belongs to the continuous spectrum, then we have in particular that the image of the unit sphere under A_z does not contain an open neighborhood of the origin. Otherwise, we have the property that A_z is surjective: henceforth, there exists a sequence of unit vectors v_n such that

$$||A_z v_n|| \to 0$$

in the limit for n towards ∞. Therefore, elements in the continuous spectrum contain approximate eigenvectors. Henceforth, $\text{Im}(A_z)^\perp$ vanishes due to injectivity of A_z implying that $\text{Im}(A_z)$ is dense in $\mathcal{H}$. It holds moreover that for $z \neq z'$,

$$\lim_{n,m\to\infty} \langle v_n|w_m\rangle = 0$$

where $(v_n)_{n\in\mathbb{N}}$ corresponds to A_z and $(w_n)_{n\in\mathbb{N}}$ with $A_{z'}$ giving a generalization of the standard orthogonality property for Hermitian projection operators associated to discrete eigenvalues.

Finally, we deal with the construction of the spectral measure: given the measurable set $\mathcal{O} \subseteq \mathbb{C}$, one defines $P_\mathcal{O}$ as the smallest Hermitian projection operator having the property that for each $z \in \sigma(A)\cap\mathcal{O}$ and sequence of approximating eigenvectors $(v_n)_{n\in\mathbb{N}}$ associated to z, then it holds that $||P_\mathcal{O}(v_n) - v_n|| \to 0$ in the limit for $n \to \infty$. A measurable Borel set is defined by means of:

- every open set can be measured,
- the complement of a measurable set is measurable,
- any union of measurable sets can be measured.

We postpone the delicate aspects of measure theory to the next chapter and now proceed with the closure of the proof. It may be clear that

$$P_\mathcal{O} P_\mathcal{V} = P_{\mathcal{O}\cap\mathcal{V}}$$

and the diligent reader delivers a proof. Given a countable partition $(B_n)_{n\in\mathbb{N}}$ of $\mathbb{C}$ by means of measurable sets[1] we consider

$$A_{(B_n)_{n\in\mathbb{N}}} = \sum_{n=0}^{\infty} z_n P_{B_n}$$

[1]A partition satisfies the property that $B_n \cap B_m = \emptyset$ for $n \neq m$ as well as $\cup_{n=0}^{\infty} B_n = \mathbb{C}$.

where $z_n \in B_n$. The integral is then defined by means of refining the partition and the remainder exercise consists in showing that the sum converges in the weak-$\star$ topology towards the integral as well as A. The first assertion is true by definition whereas the latter follows from prudent estimates.

This extremely important theorem, known as the spectral theorem, allows one to define measurable functions $f : \mathbb{C} \to \mathbb{C}$ replacing the complex variable by the normal operator A. We have that

$$f(A) := \int_{\mathbb{C}} f(z) dP(z)$$

where we have used the spectral decomposition

$$A = \int_{\mathbb{C}} z dP(z).$$

There exist two important generalizations of this theorem: the first one consists in replacing the complex numbers by means of the quaternions $\mathbb{RQ}$ and to consider quaternion bi-modules with a quaternion valued scalar product. A second generalization consists in dropping the condition

$$\langle v|v\rangle \geq 0$$

and to allow for this quantity to become negative. This kind of generalization is much more subtle and requires amongst others the introduction of conjugated null pairs. These comments wrap up our discussion about linear spaces and functions; as usual, there is much more beef into the cow as made explicit above but these constitute the main results indeed. In the next chapter, we study non-linear function theory as well as geometry and cosmology. The reader is now invited to make some exercises.

Exercises on Von Neumann Extensions of Linear Operators

Consider the operator $i\frac{d}{d\theta}$ on the space of *differentiable* functions on the unit circle S^1 with circumference 2π. Show that this operator is essentially self adjoint and densely defined on the Hilbert space of square integrable functions on the circle. Consequently, this operator has a unique Von Neumann extension. As an additional exercise, prove that

$$\left[i\frac{d}{d\theta}, \theta \right] = i((2\pi - 1)\delta(\theta) + 1)$$

where $\delta(\theta)$ is defined by means of

$$\int d\theta \, \delta(\theta) f(\theta) = f(0)$$

for any continuous function f on the unit circle.

Perform now the same study for $i\frac{d}{dx}$ defined on complex valued functions with as domain $[a, b]$ by imposing boundary conditions $f(a) = f(b) = 0$. Show that the operator on this function domain $\mathcal{D}$ is symmetrical and determine the adjoint domain $\mathcal{D}^\star$ (differentiable functions on the line segment without boundary conditions). The closure of $i\frac{d}{dx}$ requires some weaker boundary conditions. To calculate those, note that the kernels of the operators $\frac{d}{dx} \pm 1$ are provided by $a_{\mp N} e^{\mp x}$ where $a_{\mp N}$ is a suitable normalization constant. One obtains therefore a one parameter group of unitary operators

$$U(\theta) : a_{-N} e^{-x} \to e^{i\theta} a_{+N} e^{x}$$

providing for a one parameter family of Von Neumann extensions.

Delta-Dirac Distributions

Let $x \in \mathbb{R}$ be a real variable and $f : \mathbb{R} \to \mathbb{C}$ a continuous function, then we define a distribution $\delta(x)$ by means of

$$\int_{\mathbb{R}} \delta(x) f(x) dx = f(0).$$

The integral representation of $\widehat{\delta}$ constitutes a linear functional on the complex vectors space of complex valued functions $f : \mathbb{R} \to \mathbb{C}$ provided by

$$\widehat{\delta}(f) = f(0).$$

The latter is a weak limit of a sequence of continuous functionals construed by means of

$$g_n := n\chi_{[-\frac{1}{2n},\frac{1}{2n}]}$$

where

$$\chi_A(x) = 1$$

if and only if $x \in A$ and zero otherwise. More precisely

$$\widehat{\delta}(f) = \lim_{n\to\infty} \int_{\mathbb{R}} g_n(x)f(x) := \int_{\mathbb{R}} \delta(x)f(x)$$

whereby this last notation constitutes a formal representation. Likewise, one may define $\widehat{\delta}_z$ by means of $\int_{\mathbb{R}} \delta(x-z)f(x) = f(z)$. Prove that

$$\int_{\mathbb{R}} \delta(x-z)\delta(x-y)dx = \delta(y-z)$$

by insisting on

$$\int_{\mathbb{R}} dz \int_{\mathbb{R}} \delta(x-z)\delta(x-y)dxf(z) = \int_{\mathbb{R}} dx\delta(x-y)\int_{\mathbb{R}} \delta(x-z)f(z)dz$$

for any continuous function $f(z)$. Let f, g be two continuous functions from $\mathbb{R}$ onto $\mathbb{C}$ differing from zero only on a compact set such that

$$\langle f|g\rangle = \int_{\mathbb{R}} \overline{f(x)}g(x)dx$$

is well defined. Show that the latter expression provides for a scalar product and define $L^2(\mathbb{R}, dx)$ as the Hilbert space defined by means of this scalar product by taking the completion. Define subsequently the following linear operator $X(f)$ by means of

$$(X(f))(x) = xf(x).$$

Show that the latter is densely defined, essentially self adjoint (vanishing deficiency indices) and that the spectre is continuous and equals the entire $\mathbb{R}$. Finally, the projective measure P is given by means of

$$P((a,b)) = \chi_{(a,b)}$$

as well as

$$(dP(z)f)(x) = \delta(x-z)f(z)dz$$

such that finally

$$(X(f))(x) = \int_{\mathbb{R}} z\delta(x-z)f(z)dz = xf(x).$$

Heisenberg equations.

In the traditional operational formulation of quantum theory, one has the so called Heisenberg pair (X, P), modeled by means of Hermitian operators on an infinite dimensional Hilbert space

$$[P, X] = i1.$$

Herein, one considers the so called Schroedinger representation on $L^2(\mathbb{R}, dx)$ where X has been defined previously and

$$P = i\frac{d}{dx}.$$

In chapter ten, we will explicitly verify that the spectrum of P is given by $\mathbb{R}$ and that the so called distributional eigenvectors are provided by e^{-ikx}. The latter define the so called Fourier transformation.

Chapter 8

Higher Dimensional Analysis

In this chapter, we study special functions $g : \mathbb{R}^n \to \mathbb{R}^m$; in particular, we define the differential Dg, vector-fields, dual fields, general tensor fields, the exterior derivative as well as general integration theory. All these issues may be generalized towards the infinite dimensional context where one defines the so called Fréchet derivative, but we shall avoid these intricacies associated to in-equivalent topologies. It is of extreme importance to have a solid grasp upon tensor calculus in terms of index manipulation as well as to have a deep understanding of the geometrical significance of algebraic relationships. The degree of sophistication of the calculus, in particular the Lie bracket, is entirely due to the structural properties of the space $\mathbb{R}^n$. We shall return to this issue in chapter fifteen when we revise the status of the torsion tensor; in general spaces, the notion of torsion is interchangeable with the one of a "deformed" Lie bracket where the latter has to be seen as a kind of vector space calibration given that it has the same symmetries as the torsion tensor, although it is no tensor, and moreover satisfies the Jacobi identity which is not the case for general torsion tensors. In what follows, the norm we shall employ is poured into standard form associated to the standard scalar product. The coordinates in $\mathbb{R}^m$ associated to an orthonormal basis may be noted by x'^μ, x'^ν where those regarding an orthonormal basis in $\mathbb{R}^n$ are denoted by x^α, x^β. Given $g : \mathbb{R}^n \to \mathbb{R}^m$, the notation $g^\mu(x^\alpha)$ has a unique meaning and we commence by defining partial derivatives $\partial_\alpha = \frac{\partial}{\partial x^\alpha}$. These are fixed by the condition that

$$\lim_{h \to 0} \frac{||f(x + he_\alpha) - f(x) - \partial_\alpha f(x)h||}{h} = 0$$

where h is a real number. In case $n = m = 1$ one disposes merely of one partial derivative called the derivative; show that

- $\partial_x x^n = nx^{n-1}$ for $n \geq 1$,
- $\partial_x c = 0$ for a constant c,
- $\partial_x y^m = 0$ for another variable y.

Prove now that in case all partial derivatives exist in a point x, then f is continuous in that point. More in general, the function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable if and only if there exists a unique linear mapping $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ such that it holds

$$\lim_{||h|| \to 0} \frac{||f(x + h) - f(x) - Df(x)h||}{||h||} = 0$$

for a nonzero $h \in \mathbb{R}^n$. It is child's play to show that $Df(x) = \partial_\alpha f(x) dx^\alpha$ where $dx^\alpha(h) = h^\alpha$ and consequently all partial derivatives do exist in x in case the derivative does. The reverse is not necessarily true: consider for example any continuous function in two variables with as restrictions $f(x, 0) = x^2, f(0, y) = y^2$ and $f(x, x) = |x|$ for x, y sufficiently close to zero. Then all partial derivatives do exist in zero (with value zero) but not so for the total derivative as the reader may easily verify by probing the $(1, 1)$ direction. In case all partial derivatives exist in a neighborhood of a point x and are moreover continuous in x, then the derivative of the function exists in x. Try to deliver this proof for yourself.

Given that partial derivatives constitute so called Leibniz operators on functions $f : \mathbb{R}^n \to \mathbb{R}^m$, a natural step consists in studying the commutativity properties. The following result holds: in case all second order derivatives $\partial_\alpha \partial_\beta f$ exist and are continuous then it holds that

$$\partial_\alpha \partial_\beta f = \partial_\beta \partial_\alpha f.$$

We leave an easy proof of this statement to the discretion of the reader. One knows now that the matrix

$$H^\alpha_\beta(x) = \delta^{\alpha\gamma} \partial_\gamma \partial_\beta f(x)$$

is symmetric for good functions and therefore one has a spectrum of real eigenvalues. Define now $\Delta(x)$ as the number of positive minus the number of negative eigenvalues of $H^\alpha_\beta(x)$, the so called deficiency index of the Hessian and call a point x critical in case $\partial_\alpha f(x) = 0$ for all α. Critical point and their deficiency indices play an important part and carry topological information as has been revealed by Brouwer and Morse. Finally, one shows that the following rules hold

- $\partial_\alpha (f \otimes g)(x) = (\partial_\alpha f) \otimes g(x) + f \otimes (\partial_\alpha g)(x)$,
- $\partial_\alpha (af + bg)(x) = a\partial_\alpha f(x) + b\partial_\alpha g(x)$,
- $\partial_\alpha (f(g(x))) = \partial_{g^\beta(x)} f(g(x)) \partial_\alpha g^\beta(x)$.

The first identity is known as the Leibniz rule, the second one expresses linearity whereas the third follows from the previous two.

We now study differentiable generalizations of topological homeomorphisms: $g : \mathcal{O} \subseteq \mathbb{R}^n \to \mathcal{V} \subseteq \mathbb{R}^n$ is called a C^n diffeomorphism for $n \in \mathbb{N}_0$ if and only if g is a homeomorphism and for each $k \leq n$ the derivatives $D^k g$ as well as $D^k g^{-1}$ exist. In the sequel, we chiefly study C^2 or C^∞ diffeomorphisms but exceptional circumstances may occur. It is natural to consider the mapping $f \circ g : \mathcal{O} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ whose notation can be abbreviated to

$$f^\mu(g^\beta(x^\alpha))$$

and the task at hand is to take x^α derivatives keeping in mind that $x'^\beta(x^\alpha) := g^\beta(x^\alpha)$. We already know that

$$\partial_\alpha f(x'^\beta(x^\gamma)) = \partial'_\delta f(x'^\beta(x^\alpha)) \partial_\alpha x'^\delta(x^\gamma)$$

using Einstein summation in the δ indices. Often, this rule translates as

$$\partial_\alpha = \frac{\partial x'^\beta}{\partial x^\alpha} \partial'_\beta$$

leading to the formula

$$\frac{\partial x'^\beta}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial x'^\gamma} = \delta^\beta_\gamma$$

and likewise so for x^α, x'^β exchanged. δ^β_γ is a so called $(1, 1)$ tensor defined by means of $\delta^\beta_\gamma = 1$ if $\alpha = \gamma$ and 0 otherwise. At the vector space level, it is possible to identify e_α with ∂_α such that basis vectors acquire an operational significance. Insisting upon

$$dx^\alpha(\partial_\beta) = \delta^\alpha_\beta$$

to hold, one derives that dx^α has a status in $(\mathbb{R}^n)^\star$ the topological dual of $\mathbb{R}^n$. Application of a diffeomorphism results in

$$dx'^\alpha = \frac{\partial x'^\alpha}{\partial x^\beta} dx^\beta$$

which is normal given that

$$dx'^\alpha(\partial'_\beta) = \delta^\alpha_\beta$$

constitutes an invariant under local diffeomorphisms of $\mathbb{R}^n$. We now define vector-fields as differential operators

$$\mathbf{V}(x) = V^\alpha(x) \partial_\alpha$$

such that under a local diffeomorphism of $\mathbb{R}^n$ the following transformation law holds

$$\mathbf{V}'(x') = V'^\alpha(x'(x)) \partial'_\alpha = V^\alpha(x) \partial_\alpha$$

implying that

$$V'^{\alpha}(x'(x))\frac{\partial x^{\beta}}{\partial x'^{\alpha}} = V^{\beta}(x).$$

Likewise, we define dual fields

$$\omega = \omega_{\alpha} dx^{\alpha}$$

transforming as

$$\omega'_{\alpha}\frac{\partial x'^{\alpha}}{\partial x^{\beta}} = \omega_{\beta}.$$

The vector-fields $\mathbf{V}, \mathbf{W}$ constitute a Lie-algebra with defining property

$$[\mathbf{V}, \mathbf{W}] = \mathbf{VW} - \mathbf{WV} = \left(V^{\alpha}\partial_{\alpha}W^{\beta} - W^{\alpha}\partial_{\alpha}V^{\beta}\right)\partial_{\beta}$$

which transforms as a vector-field. We are now in a position to define tensor products of vector-fields and one forms where we put all vectors to the left and all one forms to the right. This leads to mathematical objects such as

$$T^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_s}(x)\partial_{\alpha_1}\otimes\ldots\otimes\partial_{\alpha_r}\otimes dx^{\beta_1}\otimes\ldots\otimes dx^{\beta_s}$$

with as transformation law

$$T'^{\alpha_1\ldots\alpha_r}_{\beta_1\ldots\beta_s}(x') = \frac{\partial x'^{\alpha_1}}{\partial x^{\gamma_1}}\cdots\frac{\partial x'^{\alpha_r}}{\partial x^{\gamma_r}}\frac{\partial x^{\delta_1}}{\partial x'^{\beta_1}}\cdots\frac{\partial x^{\delta_s}}{\partial x'^{\beta_s}}T^{\gamma_1\ldots\gamma_r}_{\delta_1\ldots\delta_s}(x).$$

This object is called an (r, s) tensor with r contra-variant and s covariant indices.

In the definition of a determinant of a matrix we encountered the procedure of anti-symmetrization and associated this with the correct fashion to compute volumes of parallelipids. We proceed now by defining lower order "determinants" having a corresponding significance for lower dimensional surfaces by using this same procedure again. Especially, we define the "wedge" product of one forms dx^{α}, the latter has the properties of associativity and is anti-symmetric as well

$$dx^{\alpha}\wedge dx^{\beta} = -dx^{\beta}\wedge dx^{\alpha}$$

and finally $dx^{\alpha_1}\wedge dx^{\alpha_2}\wedge\ldots\wedge dx^{\alpha_k}$ defines a $(0, k)$ covariant tensor. From these definitions, it is clear that

$$dx^{\alpha_1}\wedge dx^{\alpha_2}\wedge\ldots\wedge dx^{\alpha_k} = \frac{1}{k!}\sum_{\rho\in S_k}\text{sign}(\rho)dx^{\alpha_{\rho(1)}}\otimes\ldots\otimes dx^{\alpha_{\rho(k)}}$$

and given that the dimension of $\mathbb{R}^n$ equals n, we have that the space of k-forms is associated to the number $\left(\begin{array}{c} n \\ k \end{array}\right) = \frac{n!}{(n-k)!k!}$. Given a function $f : \mathbb{R}^n \to \mathbb{R}$ we define the exterior derivative

$$df = \partial_{\alpha} f dx^{\alpha}$$

which clearly constitutes a coordinate invariant. Regarding a k form

$$\mathbf{A} = A_{\mu_1\ldots\mu_k}dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_k}$$

one has the definition

$$d\mathbf{A} = \partial_{\mu}A_{\mu_1\ldots\mu_k}dx^{\mu}\wedge dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_k}$$

an expression which is invariant under coordinate transformations given that second partial derivatives are symmetric, an expression which vanishes by means of the anti-symmetric $\wedge$-product. Consequently, the entire expression behaves as a total anti-symmetric $(0, k+1)$ tensor. Note that the boundary operator ∂ in simplicial homology and the exterior derivative both satisfy

$$d^2 = 0 = \partial^2$$

which can easily be verified by the reader. Henceforth, d allows for the same construction as ∂ resulting in a de-Rahm cohomology theory. One immediately understands where this is leading to: if Ω_k is a measure for k dimensional surfaces, then $d\Omega_k$ constitutes a measure for $k+1$ dimensional surfaces depending only upon the k-dimensional boundary. This is precisely the content of the Stokes theorem which we shall study later on at the level of Lebesgue integrals. Stokes theorem can be proved however at a meta level in the following fashion: given a duality relation $\langle\Omega_k, S_k\rangle_k$ between k-forms and k-surfaces satisfying

- $\langle \Omega_k, S_k \cup T_k \rangle_k = \langle \Omega_k, S_k \rangle_k + \langle \Omega_k, T_k \rangle_k - \langle \Omega_k, S_k \cap T_k \rangle_k$ where $S_k \cap T_k$ is put equal to $\emptyset$ in case it regards a lower dimensional surface,
- $\langle \Omega_k, \emptyset \rangle_k = 0$ and $\langle a\Omega_k + b\Omega'_k, S_k \rangle_k = a\langle \Omega_k, S_k \rangle_k + b\langle \Omega'_k, S_k \rangle_k$,
- there exists a linear operator D_k mapping a k-form to a $k+1$-form such that $\langle \Omega_k, \partial S_{k+1} \rangle_k = \langle D_k(\Omega_k), S_{k+1} \rangle_{k+1}$.

The third condition is merely non-trivial to the extend that the domain of the adjoint operator must be equal to the full measure space; the latter is ensured in case

$$|\langle \Omega_k, \partial S_{k+1} \rangle_k - \langle D_k \Omega_k, S_{k+1} \rangle_{k+1}| \leq M |\langle \Omega_k, \partial S_{k+1} \rangle_k|^{\frac{k+2}{k}}$$

for a certain $M > 0$, independent of S_{k+1}, in the limit for $S_{k+1} \to \emptyset$. This condition is called the micro boundary condition. From the definition of D_k and $\partial^2 = 0$ one can derive that $D_{k+1} D_k = 0$, and provided that d constitutes the only covariant operator existent in differential geometry, it holds that $D_k = c_k d$ with c_k a real constant. Therefore, an adequate notion of an integral must satisfy the micro boundary condition with $D_k = d$. As an aside, there exist many candidate integral procedures which all coincide in the appropriate class of functions such as the Riemann, Stieltjes and Lebesgue integral. The philosophy behind the d operator is that it allows for a k-form to extend in the $k+1$-dimensional world in a way such that the $k+1$ volume only depends upon the k dimensional boundary.

In the remainder of this chapter, we introduce the Lebesgue integral on a wide class of functions encompassing the differentiable ones. We treat this subject in the most general fashion for topological spaces X equipped with a Hausdorff topology $\tau(X)$. We recall the reader that the Borel-Sigma algebra $\mathcal{B}(X)$ defined by $\tau(X)$ is generated by the open sets $A \in \tau(X)$ by taking complements as well as at most countable unions and intersections. Given $\mathcal{B}(X)$, we define a measure μ by means of the properties

- $\mu(A) \geq 0$ for all $A \in \mathcal{B}(X)$,
- $\mu(\cup_{n \in \mathbb{N}_0} A_n) = \sum_n \mu(A_n)$ if for all $n \neq m$ holds that $A_n \cap A_m = \emptyset$,
- $\mu(\emptyset) = 0$.

We say that the measure is non degenerate if $\mu(B) > 0$ for each $B \in \tau(X)$; therefore, open sets have a nonzero volume. The construction of the Lebesgue integral is rather elaborate and depends upon rather strong convergence criteria I have criticized in the past. A function $f : X \to \mathbb{R}$ is measurable if and only if $f^{-1}(C)$, with $C \in \mathcal{B}(\mathbb{R})$, belongs to $\mathcal{B}(X)$; hence, the inverse of a measurable set is measurable. One shows that a continuous function is measurable by using the construction that every measurable set can be construed out of open sets. The Lebesgue construction employs a splitting of into a positive and negative part $f = f^+ - f^-$ where $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ and the integral for measurable, positive functions is given by:

$$\int f^{\pm}(x) d\mu(x) = \sup_{\text{partitions } (A_n)_{n \in \mathbb{N}}} \sum_n \left(\inf_{x \in A_n} f^{\pm}(x) \right) \mu(A_n).$$

For a positive continuous function $f : \mathbb{R} \to \mathbb{R}^+$ one has that the integral has the geometrical significance of the surface enclosed by the graph of f, the x-axis as well as the vertical line segments through the initial and final points if those are relevant. This surface is computed by subdivision of the x-axis into small intervals and by multiplication of the infimum of the function over this interval with the length of it. Finally, summation over all these products is taken and the integral constitutes the upper limit of such sums by means of further subdivision of those intervals such that the infimum increases. The connection between k-forms $\Omega_k = \Omega_{\mu_1 \ldots \mu_k} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_k}$ and measures is then provided by subdivision of an open set by means of k-dimensional little cubes $v_1 \ldots v_k$ and write the expression $\int_{S_k} \Omega_k$ as follows

$$\sum_{\text{cubes}} \Omega_{\mu_1 \ldots \Omega_{\mu_k}}(x_0) v_1^{\mu_1} \ldots v_k^{\mu_k}$$

where x_0 is the corner of the cube on which the vectors v_j act. Clearly, this expression is invariant under coordinate transformations. Show now that the following holds for differentiable one forms $\Omega_1 = f dx$:

- $\int$ defines a duality between one forms and one dimensional simplicial complexes satisfying three conditions (in particular the micro-boundary condition); it holds that $\int_{[a,b]} df = f(b) - f(a)$,
- the one dimensional version of Stokes theorem suggests that d and $\int$ are inverses; more detailed, show that $d_x \int_{[a,x]} f(s)ds = f(x)dx$ and therefore $\int$ constitutes the right inverse of d when limited to one forms; the kernel of d equals all constant functions and henceforth $\int$ is not a left inverse of the Stokes operator,
- from $d(fg) = dfg + fdg$ it follows that $\int_{[a,b]} (g(x)\partial_x f(x) + f(x)\partial_x g(x))\, dx = f(b)g(b) - f(a)g(a)$ known as the rule of partial integration.

The law of partial integration is very important regarding the calculation of many one dimensional integrals. Apart from Stokes theorem, another important property of integrals holds which is provided by invariance under action of diffeomorphisms showing that the duality is intrinsic. We comment in a more precise way about this issue in the chapter on differentiable geometry.

Lebesgue supposed that the associated limits over all partitions are finite and posed that the integral of f is provided by the prescription

$$\int f(x)d\mu(x) = \int f^+(x)d\mu(x) - \int f^-(x)d\mu(x).$$

The canonical extension towards complex functions is self evident and we subsequently discuss two important theorems. The first one is Fubini's theorem which is often used while calculating integrals: let X, Y be two Hausdorff topological spaces equipped with a Borel-sigma algebra $\mathcal{B}(X), \mathcal{B}(Y)$, define $\mathcal{B}(X \times Y)$ starting from the product topology on $X \times Y$ generated by means of opens squares $A \times B$ with $A \in \tau(X)$ and $B \in \tau(Y)$. Clearly, it holds that $\mathcal{B}(X) \times \mathcal{B}(Y) \subset \mathcal{B}(X \times Y)$ allowing for the construction of product measures $\mu \times \nu$. Let $f : X \times Y \to \mathbb{C}$ be a measurable function for which holds that $\sup_{y \in Y} \left|\int_X f(x,y)d\mu(x)\right| < \infty$ and $\sup_{x \in X} \left|\int_Y f(x,y)d\nu(y)\right| < \infty$ then it holds that

$$\left|\int_{X \times Y} f(x,y)d(\mu \times \nu)(x,y)\right| < \infty$$

and moreover

$$\int_{X \times Y} f(x,y)d(\mu \times \nu)(x,y) = \int_X d\mu(x) \left(\int_Y f(x,y)d\nu(y)\right)$$
$$= \int_Y d\nu(y) \left(\int_X f(x,y)d\mu(x)\right).$$

Note that the uniform boundaries regarding the partial integration are necessary. Consider the space $\mathbb{R} \times \mathbb{R}_0^+$ and the function $f(x,y) = e^{-x^2 y}$ then it holds that

$$\int_{\mathbb{R}} e^{-x^2 y} dx = \sqrt{\frac{\pi}{y}}$$

and

$$\int_{\mathbb{R}_0^+} e^{-x^2 y} = \frac{1}{x^2}$$

which is not uniformly bounded. Then, one arrives at

$$\int_{\mathbb{R}_0^+} dy \left(\int_{\mathbb{R}} e^{-x^2 y} dx\right) = \infty = \int_{\mathbb{R}} dx \left(\int_{\mathbb{R}_0^+} e^{-x^2 y} dy\right)$$

where in the first integral the slow falloff of $\sqrt{\frac{\pi}{y}}$ towards $y = +\infty$ is responsible for the divergence and likewise so for the pole on $x = 0$ for $\frac{1}{x^2}$ in the second case.

The second theorem is Lebesgue's dominated convergence theorem which goes as follows: suppose that a sequence of measurable functions f_n converging point wise to a function f such that $|f_n| \leq g$ for all n and g is integrable. Then it holds that

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$$

and $\left|\int f_n d\mu\right| \leq \int g d\mu$ as well. The proofs of these theorems are rather boring and easy; the interested reader may consult a good book on the matter. One finally shows that the Lebesgue integral obeys the necessary duality conditions in each dimension such that the adjoint of the boundary operator ∂ equals the exterior derivative d.

Exercises: compute some derivatives and integrals

- compute $\int_0^1 dx \int_0^1 dy \sin^2(\pi(x+y))$,
- compute $\partial_x \partial_y \sin(xy + y^2)$,
- compute all solutions of $(\partial_x^2 - \partial_y^2)f(x,y) = 0$,
- compute $\int_1^x \frac{1}{y^2+ay} dy$.

Chapter 9

Complex Analysis

Complex analysis regards function theory in a single complex variable; this theory has stronger results as the one of a function in one real variable due to very special properties of the complex plane. Indeed, the latter has closed curves which do not self intersect, a topological property which reflects itself in the analysis. This has non-local consequences starting from local considerations and we shall just see how this arises by making use of Stokes theorem. This is the magic of the number i when considering the basic variable $z = x + iy$. As usual, one can perceive z as a composition of two real variables x, y which suggests the use of the pair of complementary variables $z, \overline{z}$. From the point of view of partial differential operators, this gives

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\partial_x - i\partial_y\right), \ \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\partial_x + i\partial_y\right)$$

with as properties

$$\frac{\partial}{\partial z} z = 1, \ \frac{\partial}{\partial z}\overline{z} = 0$$

and reversely when switching z by $\overline{z}$. Henceforth, a function in one complex variable z, $f(z)$ obeys the equation

$$\frac{\partial}{\partial \overline{z}} f(z) = 0$$

at least when those partial derivatives regarding x, y do exist. A complex valued function in z may be interpreted as a one form in $\mathbb{R}^2$ using

$$f(z) = \mathrm{Re}f(z) + i\mathrm{Im}f(z)$$

where the mapping needs to occur in a fixed coordinate system. The transformation properties of $f(z)$ under $z \to z(z')$ do not match the transformation laws of a one form meaning that the mapping between a complex function in a complex variable and a one form in the real plane in two real variables is not a canonical, meaning God given, one. As an example, consider $f(z) = z$ and $z = z'^2$; hence $f(z'^2) = (x'^2 - y'^2) + 2ix'y'$ where $\frac{\partial x}{\partial x'} = 2x'$, $\frac{\partial y}{\partial x'} = -2y'$, $\frac{\partial y}{\partial x'} = 2y'$ as well as $\frac{\partial y}{\partial y'} = 2x'$ such that

$$\mathrm{Re}f'(z') \neq \frac{\partial x}{\partial x'}\mathrm{Re}f(z'^2) \pm \frac{\partial y}{\partial x'}\mathrm{Im}f(z'^2)$$

where $f'(z') = f(z'^2)$. Nevertheless, employing the notation $\mathbf{F}(z) = \mathrm{Re}f(z)dx - \mathrm{Im}f(z)dy$, then one notices that the condition

$$\frac{\partial}{\partial \overline{z}} f(z) = 0$$

is equivalent to

$$d\mathbf{F}(z) = 0, \partial^\alpha F_\alpha = 0$$

meaning that the one form is closed and has zero divergence. Using $dz = dx + idy$ one arrives at

$$f(z)dz = \mathbf{F}(z) + i(F_x dy - F_y dx).$$

The imaginary part is closed due to $d(F_x dy - F_y dx) = (\partial^\alpha F_\alpha)\, dx \wedge dy = 0$ using the zero divergence condition. Stokes theorem then implies that

$$0 = \int_S d\mathbf{F}(z) + i\, d(F_x dy - F_y dx) = \int_{\partial S} f(z) dz$$

where S constitutes any surface in $\mathbb{R}^2$ whereupon $df(z)$ exists. Consequently, the integral of f over any closed curve depends merely upon the homology class within which f is analytical meaning

$$\frac{\partial}{\partial \overline{z}} f(z) = 0.$$

Now, we show that in case $f(z)$ is analytical, then the z-derivative $\frac{\partial}{\partial z} f(z)$ exists and is analytical again. Clearly, we merely have to show that the second partial derivatives exist and are continuous; in that case

$$\frac{\partial}{\partial \overline{z}} \frac{\partial}{\partial z} f(z) = \frac{\partial}{\partial z} \frac{\partial}{\partial \overline{z}} f(z) = 0.$$

The proof of the latter statement goes as follows

$$\lim_{|z-a| \to 0} \frac{f(z) - f(a)}{z - a} = \frac{\partial}{\partial a} f(a)$$

following from

$$\frac{f(z) - f(a)}{z - a} = \frac{(f(z) - f(a))(\overline{z - a})}{|z - a|^2} =$$
$$\frac{(\mathrm{Re} f(z) - \mathrm{Re} f(a))(x - b) + (\mathrm{Im} f(z) - \mathrm{Im} f(a))(y - c)}{|z - a|^2}$$
$$\frac{-i(\mathrm{Re} f(z) - \mathrm{Re} f(a))(y - c) + i(\mathrm{Im} f(z) - \mathrm{Im} f(a))(x - b)}{|z - a|^2}$$
$$\sim \frac{\partial}{\partial a} f(a) \frac{|z - a|^2}{|z - a|^2}$$

where in the last step we used $\frac{\partial}{\partial \overline{z}} f(z) = 0$. Using Stokes theorem as well as taking the limit of extremely small circles around a point a, the reader verifies by explicit computation that

$$\int_{S^1(a,\epsilon)} \frac{f(z)}{z - a} dz = 2\pi i f(a)$$

where $S^1(a, \epsilon)$ denotes the circle with radius ϵ around a. From this, it follows that $f(a)$ can be derived an infinite number of times with regard to a and that all derivatives are analytic. This result can easily be generalized to

$$\int_\gamma \frac{f(z)}{z - a} dz = 2\pi n i f(a)$$

where γ constitutes a closed curve in a neighborhood of a winding n times around a.

One of the most important properties of an analytic function is that the latter may be written as a converging power sequence in a neighborhood of a; more specifically,

$$f(z) = \sum_{n=0}^{\infty} a_n (z - a)^n$$

whereas the right hand side is finite for $|z - a| < \epsilon$. The proof of this assertion is pretty easy and follows from

$$\int_{S^1(a,\epsilon)} \frac{f(z)}{z - a} dz = 2\pi i f(a).$$

When differentiating n-times with regard to a one gets

$$n! \int_{S^1(a,\epsilon)} \frac{f(z)}{(z-a)^{n+1}} dz = 2\pi i \left(\frac{\partial}{\partial a} \right)^n f(a)$$

from which it follows that

$$\left| \left(\frac{\partial}{\partial a} \right)^n f(a) \right| \leq \frac{n!}{\epsilon^n} \max_{z \in S^1(a,\epsilon)} |f(z)| .$$

Subsequently one shows that

$$f(b) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\partial}{\partial a} \right)^k f(a)(b-a)^n$$

following from

$$f(b) - f(a) = (b-a) \frac{1}{2\pi i} \int_{S^1(a,\epsilon)} \frac{f(z)}{(z-b)(z-a)} =$$

$$(b-a) \frac{1}{2\pi i} \int_{S^1(a,\epsilon)} \frac{f(z)}{(z-a-(b-a))(z-a)}$$

$$= (b-a) \frac{1}{2\pi i} \int_{S^1(a,\epsilon)} \frac{f(z)}{(z-a)^2} \sum_{n=0}^{\infty} \left(\frac{b-a}{z-a} \right)^n = \sum_{n=0}^{\infty} \frac{(b-a)^{n+1} \left(\frac{\partial}{\partial a} \right)^{n+1} f(a)}{(n+1)!} .$$

We have used that

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n$$

for $|z| < 1$ as holds upon direct verification. Convergence of the sequence follows from

$$\left| \frac{(b-a)^n \left(\frac{\partial}{\partial a} \right)^n f(a)}{n!} \right| \leq M \left(\frac{|b-a|}{\epsilon} \right)^n$$

such that the series converges for

$$|b-a| < \epsilon.$$

A complex valued function $f(z)$ is called meromorphic if and only if it is analytic everywhere except at isolated points a_i such that $f(z)(z-a_i)^{n_i}$ is analytic with $\lim_{z \to a_i} f(z)(z-a_i)^{n_i} \neq 0$. We therefore have that

$$\int_{\gamma_j} f(z)(z-a_j)^{n_j - 1} dz = 2\pi i \lim_{z \to a_j} f(z)(z-a_j)^{n_j}$$

where γ_j constitutes a closed curve winding once around a_j. In the case that $n_j = 1$ we call a_j a pole and $\text{res}(a_j) := \lim_{z \to a_j} f(z)(z-a_j)$ the residue; we henceforth obtain in that case that

$$\int_{\gamma} f(z) dz = 2\pi i \sum_{\text{all poles } a_j \in S} m_j \text{res}(a_j)$$

where m_j constitutes the winding number of γ around a_j. This formula is of extreme importance for the calculation of meromorphic functions and one proceeds by making a few exercises on the matter later on.

Chapter 10

Special Functions

The chronology of presentation of the material so far has been very different from standard books. Consequently, the viewpoints as well as methods to be gained are considerably more profound. We shall start by the exponential function given that it constitutes the basis for a lot of material to come. e^x can be introduced in several ways and we shall treat plenty of them. First, e^{rx} constitutes an eigenfunction of the operator ∂_x; it is to say that

$$\partial_x e^{rx} = r e^{rx}.$$

The reader notices that the function $f_n(rx) = \left(1 + \frac{rx}{n}\right)^n$ obeys

$$\partial_x f_n(rx) = r f_{n-1}(rx)$$

henceforth taking the limit for n to infinity produces e^{rx}. Indeed,

$$e^{rx} = \lim_{n \to \infty} \left(1 + \frac{rx}{n}\right)^n$$

and one shows that this expression converges in two steps (a) for a positive rx one has that $f_n(rx)$ defines an increasing sequence and (b) the supremum is provided by

$$e^{rx} = \sum_{n=0}^{\infty} \frac{(rx)^n}{n!}$$

which is finite for any rx. Show that, from the power expansion,

$$\partial_x e^{rx} = r e^{rx}.$$

Moreover, one has that

$$e^{r(x+y)} = \lim_{n \to \infty} \left(1 + \frac{r(x+y)}{n}\right)^n = \lim_{n \to \infty} \left(\left(1 + \frac{rx}{n}\right)\left(1 + \frac{ry}{n}\right) - \frac{r^2 xy}{n^2}\right)^n = e^{rx} e^{ry}$$

something which easily follows from $\partial_x e^{r(x+y)} = r e^{r(x+y)}, \partial_x e^{rx} e^{ry} = r e^{rx} e^{ry}$ and $e^{r(0+y)} = e^{ry} = e^{r0} e^{ry}$. Finally, one may prove this property from the series expansion which suggests that the exponential function is well defined for any complex variable z en obeys

- $e^{z+w} = e^z e^w$,
- $e^0 = 1$,
- $|e^{ix}| = 1$ for all $x \in \mathbb{R}$.

This last property says that the mapping $x \to e^{ix}$ from the real axis on the unit sphere preserves length and wraps a infinite number of times around. Henceforth

$$e^{ix} = \cos(x) + i \sin(x)$$

where

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

and

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

One verifies immediately that $\partial_x \sin(x) = \cos(x), \partial_x \cos(x) = -\sin(x)$ and therefore $\partial_x^2 \sin(x) = -\sin(x)$ and likewise so for $\cos(x)$. From the definition of the modulus, it follows that

$$\cos^2(x) + \sin^2(x) = 1$$

and the Simpson rules emerge from $e^{ix}e^{iy} = e^{i(x+y)}$. It is to say that

- $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$,
- $\sin(x+y) = \cos(x)\sin(y) + \sin(x)\cos(y)$.

We have represented the complex number in polar decomposition $z = re^{i\theta}$ where $r \geq 0$ and $\theta = 0 \ldots 2\pi$ with 2π is equal to the length of the unit circle. The polar decomposition may be generalized to normal operators:

$$A = |A|U$$

where $|A| = \sqrt{A^\dagger A}$ en U constitutes a partial isometry. The exponential function has an inverse given by the natural logarithm $\ln(x)$; this obeys by definition

$$x + y = \ln(e^{x+y}) = \ln(e^x e^y) = \ln(e^x) + \ln(e^y)$$

and henceforth it holds for a general complex number that

$$\ln(z) = \ln(re^{i\theta}) = \ln(r) + i\theta.$$

One notices that this definition has a discontinuity on the positive real axis because the limit from above equals $\ln(x)$ whereas from below $\ln(x) + i2\pi$. This phenomenon is called a singular cut meaning that the function has a discontinuity there preventing for a unique continuous extension to exist. Depending upon the choice of reference point on the circle, this cut may be put anywhere the reader wants it to. Away from the cut, ln is analytical as the reader may verify.

Starting from the exponential function, one may define special functions such as the alpha, beta and gamma function. These are defined in a standard wy by means of the special integrals. As such is the gamma function $\Gamma(x)$ for $x \geq 0$ defined by means of

$$\Gamma(x) = \int_0^\infty e^{-s} s^x ds.$$

Prove that

$$\Gamma(x+1) = (x+1)\Gamma(x)$$

for $x > 0$ and $\Gamma(1) = 1 = \Gamma(0)$. Especially, it holds that $\Gamma(n) = n!$ for any natural number $n > 0$. Show that the gamma function has a unique meromorphic extension to $\mathbb{C}$ with poles in $-n$, $n \in \mathbb{N}_0$. It is not the intention to study all detailed properties of those functions and the reader is invited to look them up.

There exist many special (approximate) eigen-functions of plenty of differential operators of interest; those are given by Hermitian extensions of densely defined, symmetrical operators on specific function spaces defined by existence and bounded-ness conditions on partial derivatives of the functions of interest. More in particular, we considers sums operators such as

$$W^{i_1 \ldots i_k}(x)\partial_{i_1} \ldots \partial_{i_k}$$

where $k \in \mathbb{N}$. As and example in one real variable, one has the so called Laguerre polynomials which are eigenfunctions of

$$-\partial_x^2 + \omega^2 x^2$$

with $\omega > 0$. One can easily find those by means of the ladder operator $a = i\partial_x + i\omega x$ defined on the Hilbert space with inproduct

$$\langle f|g\rangle = \int \overline{f(x)}g(x)dx.$$

Partial integration gives $a^\dagger = i\partial_x - i\omega x$ such that

$$-\partial_x^2 + \omega^2 x^2 = a^\dagger a + \omega$$

and as such the eigenfunctions of the original operator equal those of $a^\dagger a$. There exists precisely one Ψ such that $a(\Psi) = 0$ and consequently all eigenvectors are provided by $(a^\dagger)^n(\Psi)$ which follows from $\left[a, a^\dagger\right] = 2\omega$. Calculate also the corresponding eigenvalues. The reader is invited to get a grip on such techniques which one often encounters in theoretical physics!

Clifford exponential function.
Define the Clifford exponential function from different viewpoints, for example as a Clifford generalization of the power series associated to e^x or an "eigenvector" of the operator ∂_{A_1} with Clifford valued eigenvalues.

Chapter 11

Cohomology and the de-Rahm Isomorphism

This small chapter deals with a deep insight connecting closed volume forms as well as homology classes of simplicial complexes. More accurately, we dispose of the d operator and an integral satisfying the micro-boundary condition such that Stokes theorem is obeyed by. We are interested in the $\mathbb{R}$ module C_k of closed k-forms Ω_k defined by $d\Omega_k = 0$ and ambiguous up to a k form $d\Omega_{k-1}$; the main distinction with standard homology theory consists in the fact that d increases the dimension whereas ∂ decreases it. This suggests a duality which is $H^\star_k = C_k$; indeed, consider the action of an element $\Omega_k \in C_k$ on a closed surface $S_k \in H_k$ defined by means of

$$\widehat{\Omega_k}(S_k) = \int_{S_k} \Omega_k$$

then we must first show this definition is fine. In case S_k is equivalent to S'_k given that both constitute a boundary of T_{k+1}, then we obtain

$$\widehat{\Omega_k}(S_k) - \widehat{\Omega_k}(S'_k) = \int_{T_{k+1}} d\Omega_k = 0$$

where we used Stokes theorem as well as the closed character of Ω_k. Also, we have that

$$\Omega_k \widehat{+ d\Omega}_{k-1}(S_k) = \widehat{\Omega_k}(S_k) + \int_{\partial S_k} \Omega_{k-1} = \widehat{\Omega_k}(S_k)$$

where we have used Stokes theorem as well as the assumption that S_k has no boundary. This proves that everything is well defined given that the expression does not depend upon the choice of representatives of the homology as well as cohomology classes. First, we show that the mapping is injective; this is equivalent to proving that for any non-trivial closed k-form $\Omega_k \in C_k$ there exists a non-contractible closed S_k so that $\widehat{\Omega_k}(S_k) \neq 0$. In case this is not so, we obtain that the integral over a k-surface R_k *with boundary* is completely determined by the boundary meaning that

$$\int_{R'_k} \Omega_k = \int_{R_k} \Omega_k$$

in case that $\partial R_k = \partial R'_k$. Moreover, the dependency on the boundary is local and additive such as is the case for integrals. Consequently, there exists a $k-1$ form Ω_{k-1} such that $d\Omega_{k-1} = \Omega_k$ by means of Stokes theorem. This proves injectivity and finally one proves surjectivity by posing that for any nontrivial $S_k \in H_k$ there exists a closed, but not exact, Ω_k such that $\widehat{\Omega_k}(S_k) = 1$; this is easily proved by showing that any k-form Ω_k defined on S_k may be extended in a way such that $d\Omega_k = 0$ is satisfied (in general, an infinite number of differential forms obey this property). Henceforth, we have proved that the de Betti numbers could have been defined from cohomology instead of simplicial homology. This ends the topic of this chapter.

Chapter 12

Riemannian and Lorentzian geometry

The goal of this chapter constitutes in extending the calculus of chapter eight towards so called manifolds; these are spaces which locally look as $\mathbb{R}^n$. In particular, we define a C^n manifold $\mathcal{M}$ as a topological space locally looking as $\mathbb{R}^n$; more in detail, there exists a cover with open sets $\mathcal{O}_\alpha$ homeomorphically mapped onto $\phi_\alpha : \mathcal{O}_\alpha \to \mathcal{V}_\alpha \subseteq \mathbb{R}^n$ with homeomorphisms ϕ_α onto the image $\mathcal{V}_\alpha$ which is given by an open neighborhood of the origin $\mathcal{W}_\alpha$ or an intersection of the latter with the half space $\{x | x_n \geq 0\}$. In the last case, one has that $W_\alpha \cap \{x | x_n = 0\}$ captures a piece of the boundary. In case that $\mathcal{O}_\alpha \cap \mathcal{O}_\beta \neq \emptyset$ one constructs the mapping $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(\mathcal{O}_\alpha \cap \mathcal{O}_\beta) \to \phi_\beta(\mathcal{O}_\alpha \cap \mathcal{O}_\beta)$ which is demanded to be a C^n diffeomorphism. We shall always suppose that the manifold is para-compact, which means it may be covered by means of a countable number of compact sets, and not necessarily connected although this supposition constitutes a standard request in the classic work or Hawking and Ellis. $\mathbf{V}$ constitutes a vector-field on $\mathcal{M}$ if and only if for any function $f : \mathcal{M} \to \mathbb{R}$ and chart $(\phi_\alpha, \mathcal{O}_\alpha)$ on $\mathcal{M}$ holds that there exists a $\mathbf{V}_\alpha$ on $\mathcal{V}_\alpha$ such that

$$\mathbf{V}(f) = \mathbf{V}_\alpha(f \circ \phi_\alpha).$$

$\mathbf{V}_\alpha$ is called a local representation of $\mathbf{V}$ with respect to the chart $(\phi_\alpha, \mathcal{O}_\alpha)$. From this, one may define dual fields by means of their action on vector-fields as well as general tensor-fields. Likewise, we may generalize the notion of a k-form as well as that of an exterior derivative; as an intermediate step, we define the push forward, pull back and Lie derivative. A diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ is defined as a homeomorphism such that for any two charts $(\phi_\alpha, \mathcal{O}_\alpha)$ and $(\phi_\beta, \mathcal{O}_\beta)$ such that $\psi(\mathcal{O}_\alpha) \cap \mathcal{O}_\beta \neq \emptyset$ holds that

$$\phi_\beta \circ \psi \circ \phi_\alpha^{-1} : \mathcal{V}_\alpha \to \phi_\beta \circ \psi(\mathcal{O}_\alpha) \subseteq \mathcal{V}_\beta$$

a C^n differentiable mapping and likewise so for ψ^{-1}. This is a matter of agreement, that the level at which the diffeomorphism may be derived is the same as the one of the "chart transformations". Given a diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ and function $f : \mathcal{M} \to \mathbb{R}$ we define the push forward of f by means of ψ as $(\psi^\star f)(x) = f(\psi^{-1}(x))$. The pull back is then defined as the push forward using ψ^{-1} instead of ψ and is noted as $\psi_\star f$. One generalizes the above definitions for injective, differentiable mappings $\psi : \mathcal{M} \to \mathcal{N}$. Given a vector-field $\mathbf{V}$ on $\mathcal{M}$, the push forward $\psi^\star \mathbf{V}$ is defined by means of $\psi^\star \mathbf{V}(\psi^\star f)(\psi(x)) = \mathbf{V}(f)(x)$ and likewise so for the pull back. In the same fashion, we may, by means of duality, define the push forward and pull back of one forms as well as general tensor-fields. In a coordinate representation, this reads:

$$(\psi^\star \mathbf{V})^\alpha(\psi(x)) = \frac{\partial y^\alpha(\psi(x))}{\partial x^\beta} V^\beta(x)$$

and the reader makes the obvious generalization towards one-forms and tensor-fields. We shall now measure the "difference" of $\phi^\star \mathbf{V}$ with $\mathbf{V}$; provided that such a definition must be coordinate independent, we can merely allow for infinitesimal differences. Given that one parameter family of diffeomorphisms ψ_t such that $\psi_{t+s} = \psi_t \circ \psi_s$ for t, s sufficiently small and $\psi_0 = \text{id}$ the identity transformation. Consequently, we may define the differential

$$\frac{d\psi_s^\star f}{ds}|_{s=0}(x) = \mathbf{V}(f)$$

where $V^\alpha(x) = \frac{dx^\alpha(\psi_{-s}(x))}{ds}_{|s=0}$. The ψ_s are defined by means of

$$\frac{dy^\alpha(\psi_s(x))}{ds} = -V^\alpha(\psi_s(x)).$$

Henceforth, we have obtained a connection between vector-fields and one parameter families of diffeomorphisms; this allows us to define the Lie derivative of a general tensor-field

$$\mathcal{L}_{\mathbf{V}}(T)(x) = \lim_{s \to 0} \frac{\psi_s^\star T(x) - T(x)}{s}$$

and in the context of function, vector-field and one form this procures $\mathcal{L}_{\mathbf{V}}(f)(x) = \frac{d}{ds}|_{s=0} \psi_s^\star f(x) = \mathbf{V}(f)(x)$ as well as

$$(\mathcal{L}_{\mathbf{V}}\mathbf{W})(f)(x) = \frac{d}{ds}|_{s=0}(\psi_s^\star \mathbf{W})(f)(x) = \frac{d}{ds}|_{s=0}\mathbf{W}(\psi_{-s}^\star f)(\psi_{-s}(x)) =$$
$$-\mathbf{W}\mathbf{V}(f)(x) + \mathbf{V}\mathbf{W}(f) = [\mathbf{V}, \mathbf{W}]\,(f)(x)$$

and similarly so for the one-forms. One notices that

$$d(\psi^\star \Omega) = \psi^\star(d\Omega)$$

for each k-form Ω. Henceforth,

$$\mathcal{L}_{\mathbf{V}} d = d \mathcal{L}_{\mathbf{V}}$$

which can also be shown by means of the formula

$$\mathcal{L}_{\mathbf{V}} = d i_{\mathbf{V}} + i_{\mathbf{V}} d$$

which is valid on the space of k-forms, where $i_{\mathbf{V}}$ is the contraction associated to the vectorfield $\mathbf{V}$,

$$i_{\mathbf{V}} \Omega_{\mu_1 \ldots \mu_k} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_k} = V^{\mu_1} \Omega_{\mu_1 \ldots \mu_k} dx^{\mu_2} \wedge \ldots \wedge dx^{\mu_k}.$$

From the point of view of measure theory, one obtains by definition that $(\psi^\star \mu)(\psi(A)) := \mu(A)$ for any measure μ as well as measurable set A; therefore,

$$\int_{\mathcal{B}} f d\mu = \int_{\psi(\mathcal{B})} (\psi^\star f)(\psi^\star d\mu)$$

what we call the covariant property of the integral. Show that the definitions of $\psi^\star \mu$ and $\psi^\star \Omega$ where Ω is a k-form coincide and that the integral is provided by

$$\int_{\mathcal{M}} f \Omega.$$

So far, we have not said anything in particular about differential geometry and concentrated on manifolds. In other words, we have to define a local in-product depending upon the manifold coordinates. Henceforth, we are interested in $(0, 2)$ covariant tensors $h_{\alpha\beta}$ as well as $g_{\alpha\beta}$ defining local orthonormal bases v_a as well as e_a such that

$$h_{\alpha\beta} v_a^\alpha v_b^\beta = \delta_{ab}$$

and

$$g_{\alpha\beta} e_a^\alpha e_b^\beta = \eta_{ab}$$

where η_{ab} is the so called Minkowski metric defined by means of

$$\eta_{11} = 1, \eta_{ij} = -\delta_{ij}; i, j = 2 \ldots n$$

and all other components vanish as well. In case $n = 4$, e_a is called a tetrad or vierbein; metrics such as δ_{ab} are called Riemannian and constitute a generalization of finite dimensional Hilbert spaces whereas η_{ab} is dubbed Lorentzian and defines non-compact null sets

$$\eta_{ab} v^a v^b = 0.$$

Our first task consists in showing that these geometries constitute specifications of the standard path metric and Lorentzian path metric spaces defined previously. In that vein, we define the metric distance between two points as the length of the shortest curve connecting them or the longest time-like curve in the Lorentzian case respectively. Prior to presenting the characterization of these curves, we must say something about the causal structure defined by η_{ab}. A vector v^a is called (a) causal if and only if $\eta_{ab} v^a v^b \geq 0$ (b) zero or light-like if and only if $\eta_{ab}v^a v^b = 0$ and (c) spacelike if and only if $\eta_{ab}v^a v^b < 0$. Moreover, we call a causal vector v^a future pointing with respect to the tetrad e_a if and only if $v^0 > 0$. In what follows, we presume that one disposes of a globally well defined tetrad which precludes certain topologies such as the one given by the Mobius strip. This last one is constructed by considering a rectangle, take the short edges and rotate one component for 180 degrees and subsequently identify them. The supposition on $(\mathcal{M}, g)$ may be simplified by stating that the latter needs to be time orientable as well as orientable.

Given a differentiable curve $\gamma : [a, b] \to \mathcal{M}$, we say that it is future pointing and causal if and only if the tangent vector to any point is. The length of such a curve is provided by

$$L(\gamma) = \int_a^b \sqrt{g(\dot{\gamma}(s), \dot{\gamma}(s))} ds$$

and likewise so for the length of any curve in Riemannian geometry. We define $J^+(x)$ as the set of all points y which are connected to x by means of a future pointing causal curve starting at x. Likewise, we define $J^-(x)$ as the set of all y which may be connected to x by means of a future pointing causal curve starting at y. We study now extremal curves connecting two points x and y which are considered to be fixed. For that purpose, we have to consider a "differential" with regards to the curve and put it equal to zero; the result is

$$0 = \delta L(\gamma) = \int_a^b \frac{1}{2\sqrt{g(\dot{\gamma}(s), \dot{\gamma}(s))}} \left(2g_{\mu\nu}\delta\dot{\gamma}^\mu(s)\dot{\gamma}^\nu(s) + \partial_\alpha g_{\mu\nu}\dot{\gamma}^\mu(s)\dot{\gamma}^\nu(s)\delta\gamma^\alpha(s)\right)$$

and one has to further rewrite this equation by making use of $\delta\dot{\gamma}^\mu(s) = (\delta\dot{\gamma^\mu}(s))$ until it reduces to the form

$$\int_a^b F(\gamma(s), \dot{\gamma}(s), \ddot{\gamma}(s))_\mu \delta\gamma^\mu(s) ds.$$

By noticing that this expression has to vanish for all $\delta\gamma^\mu(s)$ one arrives at the following equations

$$\ddot{\gamma}^\mu(s) + \Gamma^\mu_{\nu\alpha}\dot{\gamma}^\alpha(s)\dot{\gamma}^\nu(s) = 0$$

where $\Gamma^\mu_{\nu\alpha} = \frac{1}{2}g^{\mu\kappa}\left(\partial_\nu g_{\kappa\alpha} + \partial_\alpha g_{\nu\kappa} - \partial_\kappa g_{\nu\alpha}\right)$ is the so called Christoffel symbol. The parametrization s is called the affine parametrization given that the above equation remains invariant under the transformations $s \to bs + a$. Given that the integral is reparametrization invariant as well as invariant under coordinate transformations, this set of equations has to transform as a vector under coordinate transformations and the reader may now verify the effect of a reparametrization. We compactify this equation as

$$\dot{\gamma}^\nu \nabla_\nu \dot{\gamma}^\mu = 0$$

where

$$\nabla_\nu V^\mu = \partial_\nu V^\mu + \Gamma^\mu_{\nu\kappa}V^\kappa$$

for each vector V^μ. The reader verifies that $\dot{\gamma}^\mu(s)\partial_\mu = \frac{d}{ds}$ as well as that $\nabla_\mu V^\nu$ transforms as a $(1, 1)$ tensor. Likewise, we have that $V^\nu\nabla_\nu W^\mu - W^\nu\nabla_\nu V^\mu - [V, W]^\mu = 0$ for all vectors $\mathbf{V}, \mathbf{W}$. The first property signifies that the covariant derivative has a geometrical significance whereas the last one asserts that it is torsion-free which is equivalent to

$$\Gamma^\mu_{\nu\kappa} = \Gamma^\mu_{\kappa\nu}.$$

The equation $\dot{\gamma}^\nu\nabla_\nu\dot{\gamma}^\mu = 0$ is known as the geodesic equation and the associated solutions, being curves, are geodesics. We extend now the definition of the covariant derivative as follows

$$\nabla_\mu f = \partial_\mu f$$

or in coordinate free notation,

$$\nabla f = df, \nabla \mathbf{V} = \nabla_\mu V^\nu dx^\mu \otimes \partial_\nu.$$

Using

$$\nabla_{\mathbf{V}} = V^\mu \nabla_\mu$$

we extend the definition as

$$\begin{aligned} \nabla_{\mathbf{W}}(\omega(\mathbf{V})) &= (\nabla_{\mathbf{W}}\omega)(\mathbf{V}) + \omega(\nabla_{\mathbf{W}}\mathbf{V}) \\ \nabla_{\mathbf{W}}(S \otimes T) &= (\nabla_{\mathbf{W}}S) \otimes T + S \otimes (\nabla_{\mathbf{W}}T) \end{aligned}$$

where ω is a one form and S, T general tensors. From the specific form of the Christoffel symbols, the reader derives that $\nabla g = 0$ meaning that the metric is covariantly. It is clearly so that geodesics maximize the Lorentzian distance in case they are causal whereas a Riemannian metric involves a minimization procedure.

We study now a few tensors which one may construct from the covariant derivative as well as contractions thereof which are of primordial importance in the geometric analysis and theory of general relativity. The first case regards a $(1,2)$ tensorfield

$$\mathbf{T}(\mathbf{V},\mathbf{W}) = \nabla_{\mathbf{V}}\mathbf{W} - \nabla_{\mathbf{W}}\mathbf{V} - [\mathbf{V},\mathbf{W}]$$

and the reader verifies indeed that $\mathbf{T}(\mathbf{V},\mathbf{W}) = -\mathbf{T}(\mathbf{W},\mathbf{V})$ and

$$\mathbf{T}(f\mathbf{V}+\mathbf{Z},\mathbf{W}) = f\mathbf{T}(\mathbf{V},\mathbf{W}) + \mathbf{T}(\mathbf{Z},\mathbf{W}).$$

For the metric connection, defined by means of the Christoffel symbols, one has that the Torsion tensor vanishes and the reader is invited to study connections with torsion. The second case regards a $(1,3)$ tensorfield denoted by

$$\mathbf{R}(\mathbf{V},\mathbf{W})\mathbf{Z} = \nabla_{\mathbf{V}}\nabla_{\mathbf{W}}\mathbf{Z} - \nabla_{\mathbf{W}}\nabla_{\mathbf{V}}\mathbf{Z} - \nabla_{[\mathbf{V},\mathbf{W}]}\mathbf{Z}$$

an expression which is anti symmetrical in $\mathbf{V},\mathbf{W}$. In components, this reads $R_{\mu\nu\alpha}^{\ \ \ \ \beta}$ and we may raise or lower indices by means of $g^{\alpha\beta}$ en $g_{\alpha\beta}$ respectively. From the general Jacobi identity $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$ for operators, we conclude that, applied to the metric connection, it holds

$$R_{[\mu\nu\alpha]}{}^{\beta} = 0, \nabla_{[\alpha}R_{\beta\gamma]\kappa}{}^{\delta} = 0$$

where the square brackets are symbolic representations for the operation of total anti-symmetrization. These identities are called the first and second Bianchi identity; moreover, one verifies by means of explicit calculation

$$R_{\mu\nu\alpha}{}^{\kappa}g_{\kappa\beta} = R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}.$$

This tensor field is known as the Riemann tensor and is of great importance in the theory of general relativity. The first contraction

$$R_{\mu\nu} = R_{\alpha\mu\nu}{}^{\alpha}$$

is a symmetric tensor called the Ricci tensor and its second contraction

$$R = R_{\mu\nu}g^{\mu\nu}$$

is the Ricci scalar. The metric tensor determines a unique volume element

$$dV(x) = \sqrt{\det(g_{\mu\nu})}dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n$$

and one verifies that this expression is coordinate independent. From this, one constructs the so called Einstein-Hilbert action

$$\int dV(x)R(x)$$

the variation of which regarding the metric components produces the Einstein equations.

I encourage the reader to make computations, for example that the Torsion and Riemann tensors are really tensors and do not depend upon the derivatives of the vector-fields. The Bianchi identities are verified from the Jacobi identity without making any computation in a coordinate system, whereas the remaining symmetries do depend upon the metric and torsion-less nature of the Christoffel connection.

Exercise: Morse Theory
We have meanwhile studied the Betti numbers from the viewpoint of homology and cohomology by means of the de Rahm theorem. Now, we offer a third vision argumenting that all information is contained in the vectorfields.

- Argue that all topological information in a manifold is contained in the tangent bundle of vector-fields.

- Show that all local, topological information in a vector-field is given by the zeroes $\mathbf{V} = 0$ and some characteristics of the matrices $\partial_{(\alpha} V^{\beta)}$ and $\partial_{[\alpha} V^{\beta]}$ evaluated in the critical points. Realize that this information must correspond to some topological invariant.

- The first tensor is symmetric and has real eigenvalues; clearly, the eigenvalues are invariants and in particular the n_-, n_0, n_+ representing the number of negative, null and positive eigenvalues are too.

- The second tensor is anti-symmetric and has imaginary eigenvalues; prove that a similar topological classification holds here.

- argue why the absolute values of the eigenvalues will not contribute any topological information by re-scaling the vector fields around the critical points

- In case it holds locally that $V^\alpha = \delta^{\alpha\beta}\partial_\beta \Phi$, prove then that $\partial_{[\alpha} V^{\beta]} = 0$.

- Show that on the torus, there exists a vector-field without critical points whereas the Betti numbers are $0, 1, 1$. This shows that the Betti numbers cannot be retrieved from a single vector-field. Denote by the index N_i the number of critical points with $n_0 = 0$ and $n_- = i$; construct vector-fields on the torus with isolated zeroes only which are all non-degenerate (that is $n_0 = 0$) such that N_1 can take any value one likes and compute that $N_2 - N_1 + N_0 = 0$. This suggests that $\sum_{k=0}^{n}(-1)^k N_k$ equals the Euler number: for manifolds with boundary, it must be that the vector-fields are normal to the boundary. Note that the definition of a boundary of a manifold is different from the one we employed so far and the reader may very well restrict himself to orientable manifolds in which the latter definition coincides with the one given by means of the boundary operator. The manifold definiton says that a coordinate chart around a boundary point in an n dimensonal manifold looks like a half space in $\mathbb{R}^n$ bounded by the hyperplane $x_n = 0$. The reader must first find out that for the Mobius strip, the manifold boundary is given by a circle whereas the boundary in the other sense also contains a straight line segment connecting the two opposite sides. That the vector-field must be normal to the boundary in order provide for the correct answer can be easily seen for example by considering the 2 disc; the reader can easily construct vector-fields in $\mathbb{R}^2$ which cross the boundary and have no critical points whereas the Euler number of the 2 disc is one. However, insisting upon being outward normal to the smooth boundary the reader immediately sees that that the vectorfield must contain a critical point which is in this case a minimum $n_+ = 2$. Note also that this theorem does not hold for vector-fields which are tangent to the boundary, for the two disc for example, one may produce a rotation around the z axis in the origin and one computes that the signature is $(0, 0)$ leading to the Euler number 0 instead of 1 (however the viewpoint expressed in the next item reveals that the degree is 1 leading to the correct answer in this case). Likewise, when cutting a small disk out of the torus, the reader may easily construct a vector-field which is tangent to the boundary, for example $-(r-1)\cos\theta\partial_r + \sin\theta\partial_\theta$, and has two critical points $r = 1$, $\theta = 0, \pi$ on the boundary of signature $(+-)$ (the reader notices that this coincides with the degree) suggesting that the Euler number should be minus two instead of minus one. There is another reason why the tangent version does not hold in an odd number of dimensions which is that one may consider the mapping $V \to -V$ which causes the above sum to transform as $(-1)^n$ something which only spaces with Euler number zero don't feel in case n is odd (which includes all closed spaces due to Betti duality). Indeed, the reader may immediately prove that for a vector-field on a line segment which needs to vanish at the boundary points and has no degenerate critical points, it holds that the values of the above sum are given by $-1, 0, 1$ (and the Euler number is of course one) whereas on the circle only 0 occurs as it should! The viewpoint of the vector-field being outward normal of course breaks this argument as the mapping is simply not allowed for. Given those details, the reader may try to prove that $\sum_{k=0}^{n}(-1)^k N_k$ equals the Euler number for vector-fields with isolated, non-degenerate critical points on compact manifolds with boundaries such that the vector-field is outward normal to those. The reader might try to generalize this theorem by considering vector-fields with degenerate and non-isolated critical points and disregarding all types of critical points which occur an infinite number of times. For example, take on the torus the vector-field $\partial_\phi\left(\sin^2(\theta) + \sin^2(\phi)\right)$, then one has 4 critical points corresponding to $\theta, \phi = 0, \pi$ and the reader easily verifies that the Hessian vanishes there. Therefore, one must say that for critical points of type $(0, 0)$ the associated contribution vanishes. Now, take a different vector-field $(\sin(\phi)\partial_\phi + \sin^2(\phi - \theta)\partial_\theta$, then there are 4 critical points corresponding to $\phi = 0, \pi$ and $\theta = 0, \pi$. The reader verifies that the Hessian is diagonal in those points with eigenvalues $0, 1$ and $(0, -1)$ respectively. Hence critical points of type $(0, +)$ and $(0, -)$ should contribute opposite. By considering the vector-field $(\sin(\frac{\phi}{2})\partial_\phi + \sin^2(\phi - \theta)\partial_\theta$ the reader shows that there are only two critical points $\phi = 0$ and $\theta = 0, \pi$ and the Hessian is in both cases of type $(0, +)$ showing that all types of degenerate points should have vanishing contribution. We will now come to a contradiction showing that one cannot in general allow for a continuum of degenerate points; consider the the genus two surface which we obtain from taking the sum of two tori by cutting out small discs in both of them and identifying the boundaries. Consider a coordinate system in say the first torus around the midpoint of this disc, containing the entire disc of radius r around the origin and define the annulus as all points with coordinates (x, y) such that $0 < (r - \epsilon)^2 < x^2 + y^2 < (r + \epsilon)^2$ then the inside region of the disc may serve as a coordinate system on the second torus and the outside on the first torus. Consider now the vector-field

$$(x^2 + y^2 - r^2)\partial_y$$

and convince yourself that it can easily be extended towards both tori without any further critical points. Then, the reader notices we have a continuum of critical points for which the symmetrical part of the Hessian is of type $(+, -)$ except for $x = 0$ where two degenerate critical points of type $(0, +)$ and $(0, -)$ exist. So, we cannot ignore the continuum and the reader should see that we have precisely two connected branches of signature $(+-)$, this suggests one to consider those branches as contributing one time each leading to the correct conclusion that the Euler number is minus two. Hence, our only option being that all isolated zeroes are non-degenerate and that non-degenerate branches have to be counted as one of the same type.

- The presentation so far has been somewhat old fashioned meaning that obviously not all degenerate points are the same; it is just so that we cannot distinguish between them from the viewpoint of quasi-local analysis which caused for all above problems. To rectify this, consider an isolated critical point and draw a small ball around it such that no other critical point is contained in it. Choose any coordinate system whatsoever and take for example the flat metric h in it; then one can define a mapping from the sphere to the standard sphere S^{n-1} in $\mathbb{R}^n$ by means of

$$x \to \frac{V(x)}{\sqrt{h(V(x), V(x))}}.$$

This mapping will define a unique element of the $n - 1$'th homology group which is $\mathbb{Z}$, which upon comparison with the standard generator determines a unique integer, the degree, which says how many times it wraps around the sphere and with what orientation. Of course, for one critical point, the sphere wraps just once and the sign corresponds precisely to the index as defined above. This mapping does not depend upon the coordinate system chosen and the degree is a topological invariant. Of course, this viewpoint allows for a more accurate analysis of degenerate critical points and the reader may verify that the full theorem holds for any vector field with isolated critical points replacing the index by the degree. Similarly, an extension can be made towards vector-fields with a continuous number of critical points.

Einstein-Cartan Theory.
One might consider the introduction of a torsion tensor in the definition of the connection; it is to say, define

$$\widehat{\nabla} = \nabla + T$$

where T is a $(1, 2)$ tensorfield which is anti-symmetric in the covariant indices and ∇ is a suitable symmetric deformation of the Christoffel connection. One has in particular that

$$2\widehat{\Gamma}^{\gamma}_{[\alpha\beta]} = T^{\gamma}_{\alpha\beta}$$

as well as

$$\nabla_\alpha g_{\beta\kappa} - T^{\gamma}_{\alpha\beta} g_{\gamma\kappa} - T^{\gamma}_{\alpha\kappa} g_{\beta\gamma} = 0$$

to ensure that

$$\widehat{\nabla}_\alpha g_{\beta\kappa} = 0.$$

As an exercise: calculate $\Gamma^{\gamma}_{\alpha\beta}$ and recover the Christoffel connection by putting $T = 0$. Calculate the Riemann tensor and torsion corrections to the first and second Bianchi identity. Define the Einstein action $I(g, T)$ which depends now upon g, T, as well as the Einstein tensor $G^{\alpha\beta}$ and spin tensor $S^{\alpha\beta}_{\kappa}$ by means of the variations

$$G^{\alpha\beta}(x) = \frac{\delta I(g, T)}{\delta g_{\alpha\beta}(x)},\ S^{\alpha\beta}_{\kappa}(x) = \frac{\delta I(g, T)}{\delta T^{\kappa}_{\alpha\beta}(x)}.$$

This produces two equations measuring the energy content and rotation of the universe by means of similar quantities in the matter sector. The reader is invited to find the appropriate four conservation laws of energy-momentum *plus* spin. Determine torsion corrections on parallel transport by means of geodesics which can only feel torsion by means of symmetrical corrections in the Christoffel part. Compute the lowest order corrections to the exponential mapping defined in the following chapter.

Chapter 13

Dispersion of a bundle of Geodesics and blowing up or collapsing of Spheres

In this chapter, we proceed in the study of global geometric properties of Riemannian and Lorentzian apace times. The subject of study par excellence regards the study of neighboring geodesics with respect to one and another. Do they expand or contract; is there rotation or expansion? In other words, we shall study in detail consequences of the geodesic deviation equation on the qualitative and weakly quantitative level. It is evidently possible to obtain sharper results for Riemannian spaces as it is for Lorentzian ones given that the former do define in a natural fashion a notion of compact surfaces. Lorentzian geometry does not enjoy this property and merely allows one to obtain results regarding a time-like curve or world line. The Riemannian context is pretty well understood but the Lorentzian one however has plenty of open gaps for research. Henceforth, one may be invited to deliver an original contribution to mathematics and I shall touch a few of these problems in this chapter.

To start with, one considers a geodesic $\gamma(s)$ such that the length of the tangent vector $\mathbf{V}(s) = \frac{d}{ds}\gamma(s)$ is constant. Indeed,

$$\frac{d}{ds}g(\dot{\gamma}(s), \dot{\gamma}(s)) = (\nabla_{\mathbf{V}(s)}g)(\dot{\gamma}(s), \dot{\gamma}(s)) + 2g(\nabla_{\mathbf{V}(s)}\mathbf{V}(s), \mathbf{V}(s)) = 0$$

due to the fact that the metric is covariantly constant and where one employs the geodesic equation. In the Lorentzian case, we choose a parametrization such that the length of the tangent vector equals one for time-like ones, zero for null geodesics and minus one for space like geodesics. Denote with $T^{\star}\mathcal{M}_x$ the linear space of all vectors in x, then we define the mapping

$$\exp_x : T\mathcal{M}_x \to \mathcal{M} : v \to \exp_x(v)$$

where $\exp_x(v)$ is the endpoint of the geodesic in affine parameter length one such that the tangent vector in x is given by v. Clearly, one has that

$$\mathrm{D}\exp_x(0)(w) = w$$

where $w \in TT^{\star}\mathcal{M}_{(x,0)} \sim T^{\star}\mathcal{M}_x$ what simply means that in first order, the metric is given by a global Euclidean or Minkowskian one. Define $T^{\star}\mathcal{M} = \cup_{x\in\mathcal{M}}T^{\star}\mathcal{M}_x$, then we introduce a topology which is equal to the product topology $\mathcal{O} \times \mathcal{V}$ where $\mathcal{O} \subseteq \mathcal{M}$ and $\mathcal{V} \subseteq \mathbb{R}^n$. The exponential mapping constitutes henceforth a local diffeomorphism meaning that there exists an open neighborhood $\mathcal{V}$ of 0 in $T^{\star}\mathcal{M}_x$ such that

$$\exp_x : \mathcal{V} \to \exp_x(\mathcal{V})$$

is a diffeomorphism. Given two points x and y then it is possible for them to be connected by multiple geodesics due to a non trivial topology (think about winding on a cylinder) or the existence of focal points (think about a lens). We now derive the geodesic deviation equation: given a one parameter family of geodesics $\gamma(s,t)$ where $s : a \ldots b$ and $t \in (-\epsilon, \epsilon)$ with as tangent vectors $\mathbf{V} = \left(\frac{\partial}{\partial s}\right)^\star$ en $\mathbf{Z} = \left(\frac{\partial}{\partial t}\right)^\star$ then it holds that

$$[\mathbf{V}, \mathbf{Z}] = 0$$

and consequently

$$\nabla_V \nabla_{\mathbf{V}} \mathbf{Z} = \nabla_{\mathbf{V}} \nabla_{\mathbf{Z}} \mathbf{V} = \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}$$

where in the first equality, we have used that the connection is Torsion-less whereas the geodesic equation $\nabla_{\mathbf{V}} \mathbf{V} = 0$ has been employed in the second one. The latter equation is of the form

$$\ddot{\mathbf{Z}} + A(s)\dot{\mathbf{Z}} + B(s)\mathbf{Z} = 0$$

with A, B matrices and the dot operation indicates derivation regarding s. Such an equation is of the Newtonian type with A the friction matrix and B the oscillation frequency ($B > 0$) or expansion factor ($B < 0$) squared. Indeed, in one dimension this reads as

$$\ddot{z} + a\dot{z} + bz = 0$$

what may be rewritten by

$$\ddot{h} + (b(s) - \frac{1}{2}\dot{a}(s) - \frac{1}{4}a^2(s))h = 0$$

with $h(s) = e^{\frac{1}{2}\int_0^s a(t)dt} z(s)$. The last one is of the type

$$\ddot{h}(s) + c(s)h(s)$$

and we call $c(s)$ minus the expansion squared in case it is smaller as zero and the oscillation frequency squared in case it is larger than zero. Indeed, for c constant, the solutions are given by $xe^{\sqrt{c}s} + ye^{-\sqrt{c}s}$ and $x\cos(\sqrt{-c}s) + y\sin(\sqrt{-c}s)$ respectively where x, y are arbitrary real constants. As we shall see later on, this simple computation allows us to bound a few geometrical properties.

Along a geodesic, one may define the notion of Fermi-Walker transport of a tensor by means of

$$\nabla_{\mathbf{V}} \mathbf{T} = 0$$

where $\mathbf{V}$ constitutes the tangent vector along the geodesic. One can easily verify that Fermi-Walker transport keeps the expression

$$g(\mathbf{Z}, \mathbf{W})$$

constant for any $\mathbf{Z}, \mathbf{W}$ along some geodesic. One notices here that transport of a vector is the mathematical concept expressing the operation of dragging an infinitely small measure stick along the world line of the observer; this pulling happens without force and is exclusively determined by means of relationships inherited from the space time metric. This constitutes therefore a very important physical concept and a geodesic may be interpreted as a force-less propulsion through space time; Einstein interpreted this as the motion of a *free* observer in a gravitational field which gave rise to the birth of the theory of general relativity.

An orthonormal basis is a set of n orthonormal vector-fields e_a which may be interpreted as a local reference system which is generically not associated to a special coordinate system given that the vectors e_a do *not* commute,

$$[e_a, e_b] \neq 0.$$

A Lie algebra is defined by demanding that

$$[e_a, e_b] = f^c_{ab} e_c$$

for real *constants* f^c_{ab}. Because parallel transport of an orthonormal basis produces an orthonormal basis, it defines a Lorentz transformation by writing out the transported reference system in terms of the local one. It is to say, we have a transformation $\Lambda(x, w)^{a'}_b$ where $w \in T^\star\mathcal{M}_x$ and b is a (Lorentz) index related to $e_b(x)$ and a' a (Lorentz) index with regard to $e^{a'}(\exp_x(w))$. A Lorentz transformation constitutes a symmetry of the Minkowski metric meaning that

$$\Lambda(x, w)^{a'}_b \Lambda(x, w)^{c'}_d \eta_{a'c'} = \eta_{bd}.$$

One verifies that those form a continuous group of dimension $\frac{n(n-1)}{2}$ and that it has 4 disjoint components in case $n = 4$ which are determined through the matrices $1, T, S, ST$ where T indicates time reversion and S space reversion. The Euclidean case, determined by means of the rotation group, has been studied previously.

Another useful tool to study geodesics is provided by means of Synge's function, a mathematical object generalizing the quadratic form

$$\sigma(x,y) = \frac{1}{2}(y-x)^{\mu}(y-x)^{\nu}\eta_{\mu\nu}$$

on Minkowski, where x, y represent points in $\mathbb{R}^4$. The definition then implies that $\sigma(x,y)$ equals *some* extremal value of

$$I(x,y) = \frac{1}{2}(t_1 - t_0)\int_{t_0}^{t_1} g_{\mu\nu}\frac{dx^{\mu}(s)}{ds}\frac{dx^{\nu}(s)}{ds}ds$$

where $x^{\mu}(s)$ is a curve connecting x with y. One calculates that the extremal values are provided by the geodesic curves $x^{\mu}(s)$; moreover, the latter expression is invariant under affine reparametrizations $s \to as + b$ implying one may consider only variations $\delta x^{\mu}(s)$ such that the end points t_0, t_1 remain fixed. The reader should complete the following computation

$$\delta I(x,y) = (t_1 - t_0)\int_{t_0}^{t_1} g_{\mu\nu}\left(\frac{D}{ds}\frac{dx^{\mu}(s)}{ds}\right)\delta x^{\nu}(s)ds$$

where

$$\frac{D}{ds} = \nabla_{\frac{d}{ds}}$$

and $\delta x^{\mu}(t_0) = \delta x^{\mu}(t_1) = 0$. This variation disappears if and only if

$$\frac{D}{ds}\frac{dx^{\mu}(s)}{ds} = 0$$

which is the geodesic equation indeed. Henceforth, Syge's function equals

$$\sigma(x,y,w) = \frac{1}{2}g(w,w) = \frac{1}{2}\epsilon L^2(x,y,w)$$

where $w \in T^{\star}\mathcal{M}$, $L(x,y,w)$ equals the length of the geodesic emanating from x with tangent vector w and endpoint $\exp_x(w) = y$. Here, $\epsilon = 1$ for time-like geodesics and -1 for space like ones. Supposing that w varies continuously when x and y do, we decide that w only serves to indicate that several geodesics between x and y may exist. The reader now verifies that

$$\sigma(x,y,w)_{,\mu} := \partial^x_{\mu}\sigma(x,y,w) = -w_{\mu} = g_{\mu\nu}w^{\nu}, \sigma(x,y,w)_{,\mu'} := \partial^y_{\mu'}\sigma(x,y,w) =$$

$$-g_{\mu'\kappa'}\Lambda^{\kappa'}_{\nu}(x,w)w^{\nu}.$$

From this it follows that

$$2g^{\mu\nu}\sigma(x,y,w)_{,\mu}\sigma(x,y,w)_{,\nu} = \sigma(x,y,w)$$

and likewise so for derivatives with respect to y.

We shall finish this chapter by giving away further details regarding Riemannian geometry and the exercises concern original extensions of these theorems towards Lorentzian geometry. Let us begin with the usual definition: given two vectors v, w, then the surface of the parallelipid spanned by v, w is given by means of $g(v,v)g(w,w) - (g(v,w))^2$ and we define sectional curvature $s(v,w)$ as

$$s(v,w) = -\frac{g(R(v,w)v,w)}{g(v,v)g(w,w) - (g(v,w))^2}.$$

A metric is of constant sectional curvature if and only if

$$R_{\mu\nu\alpha\beta} = \frac{R}{n(n-1)}(g_{\mu\beta}g_{\nu\alpha} - g_{\mu\alpha}g_{\nu\beta}).$$

Consequently, the Ricci tensor reads

$$R_{\mu\nu} = \frac{R}{n} g_{\mu\nu}$$

implying any space of constant sectional curvature is an Einstein space. One notices that the mapping $g \to -g$ leaves the Ricci tensor invariant whereas the Ricci scalar and sectional curvature are mapped to their opposite. In particular, this is also true for the Einstein action in an even number of dimensions. It may be clear that in case only sectional curvature matters, we must focus on the study of homogeneous and isotropic model spaces of constant sectional curvature. These are evidently unique; these model spaces are maximally symmetrical where a symmetry is represented by means of a diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ such that $\psi^\star g = g$. In case we speak about a one parameter family ψ_t of diffeomorphisms this leads to

$$\mathcal{L}_{\mathbf{V}} g = 0$$

where $\mathbf{V}$ constitutes the generating vector-field. The reader verifies that this equation may be rewritten as

$$\nabla_{(\alpha} V_{\beta)} = 0$$

where $V_\beta = g_{\beta\alpha} V^\alpha$ and the round brackets represent the operation of symmetrization; it is to say

$$Z_{(\alpha_1 \ldots \alpha_n)} = \frac{1}{n!} \sum_{\sigma \in S_n} Z_{\alpha_{\sigma(1)} \cdots \alpha_{\sigma(n)}}.$$

This equation is known as Killing's equation and the reader easily sees that there are exactly $\frac{n(n+1)}{2}$ of them leaving $\frac{n(n-1)}{2} + n$ parameters where n is the number of free coordinates of a point. Therefore, any space with $\frac{n(n+1)}{2}$ lineairly independent Killing fields is called maximally symmetric. Using the Killing equation in the definition of the Riemann tensor gives

$$\nabla_\alpha \nabla_\beta V^\alpha = R_{\alpha\beta\gamma}{}^{\alpha} V^\gamma.$$

More in general, from

$$\nabla_\alpha \nabla_\beta V_\gamma - \nabla_\beta \nabla_\alpha V_\gamma = -R_{\alpha\beta\gamma}^{\kappa} V_\kappa$$

follows that

$$\nabla_\alpha \nabla_\beta V_\gamma + \nabla_\beta \nabla_\gamma V_\alpha = -R_{\alpha\beta\gamma}^{\kappa} V_\kappa$$

and consequently, using the symmetries of the Riemann tensor, one obtains

$$\nabla_\alpha \nabla_\beta V_\gamma = R_{\beta\gamma\alpha}^{\kappa} V_\kappa.$$

We now study the prototype *Riemannian* maximally symmetric spaces with null, positive and negative sectional curvature respectively. These geometries are not unique and may differ topologically from one and another: for example, in two dimensions, one disposes of a flat cylinder, torus and plane. Clearly, any cylinder may be isometrically embedded into $\mathbb{R}^2/T_{x,a}$ and likewise for the torus in $\mathbb{R}^2/\{T_{x,a}, T_{y,b}\}$ where $T_{x,a}$ defines a translation over a distance $a > 0$ in the x direction. In that vein, the plane is maximally symmetric and one can show that any flat metrical space can be construed by means of elementary cutting and pasting of such quotients. The maximally symmetric ones are called flat, spherical and hyperbolic respectively.

The prototype *maximal* flat spaces are provided by the Euclidean $(\mathbb{R}^n, \delta_{\alpha\beta})$ where $\delta_{\alpha\beta}$ refers to a canonical flat coordinate system. The reader may construct distinct flat spaces by gluing two of them together by means of a so called wormhole; specifically, cut a $n-1$ dimensional torus out of $\mathbb{R}^n$ and connect those by means of a n dimensional cylinder. One knows that the $\frac{n(n+1)}{2}$-dimensional symmetry group of $(\mathbb{R}^n, \delta_{\alpha\beta})$ is provided by $SO(n) \times \mathbb{R}^n$ where $SO(n)$ constitutes the special orthogonal group with as elements matrices O^μ_ν such that

$$O^\alpha_\mu O^\beta_\nu \delta_{\alpha\beta} = \delta_{\mu\nu}$$

and with determinant one. Moreover, one has a n dimensional translation group $\mathbb{R}^n$ with as total action

$$((O,a)x)^\alpha = O^\alpha_\beta x^\beta + a^\alpha.$$

$SO(n)$ is generated by means of the anti-Hermitian matrices A meaning that $A^\dagger = -A$ which implies that the group is $\frac{n(n-1)}{2}$ dimensional. Regarding global properties, it holds that between any two points x and y

one can find exactly one geodesic which is the straight line segment connecting both points. The canonical volume measure is given by $dx^1 \wedge \ldots \wedge dx^n$ which may be rewritten as

$$r^{n-1}dr \wedge \Omega_{S^{n-1}}$$

where $r = \sqrt{\sum_{i=1}^{n}(x^i)^2}$ and $\Omega_{S^{n-1}}$ constitutes the canonical volume measure on the $n-1$ dimensional sphere which is closed but not exact. This fact can easily be shown using the property that ∂_r is perpendicular to $S^{n-1}(r)$ as well as the scaling formula $d(rx, ry) = rd(x, y)$ for any $r > 0$. Henceforth, balls of radius r around a point x always have a volume equal to

$$\frac{r^n}{n}\mathrm{Vol}(S^{n-1}).$$

Another important property which the reader may show is that the sum of interior angles in a triangle equals π.

Now we shall describe maximal model spaces of positive sectional curvature which equals the n dimensional sphere of radius r in $(\mathbb{R}^{n+1}, \delta_{\alpha\beta})$; the symmetry group equals $SO(n + 1)$ which provides for the correct dimension. The distance between two points x, y is provided by the arc-length θ of the segment of the circle determined by x, y and the origin, where

$$\cos(\theta) = \frac{x^\alpha y_\alpha}{r^2}$$

and x, y $n + 1$-dimensional vectors. In order to understand the geometry, one does not need any explicit computation; it is fully determined by means of the parameter r and in particular this implies that the sphere has a constant sectional curvature of the form $\frac{a}{r^2}$ on dimensional grounds. A small computation in two dimensions reveals that $a = 1$ and the reader verifies that

$$\int_{S^2} R\sqrt{h} = 4\pi\chi(S^2) = 8\pi$$

where we bring to the recollection that χ is the Euler number. This result holds for any Riemannian metric on the sphere a result which is called the Gauss Bonnet theorem. The reader is invited to prove this as a non-trivial exercise (hint: show that the variation of the density to the metric produces a total derivative in two dimensions). One immediately appreciates that the geodesic expansion equation reduces to

$$\frac{d^2}{ds^2}Z = -\frac{1}{r^2}Z$$

providing one with solutions of the form $Z(s) = b\sin\left(\frac{s}{r}\right)$ in case $Z(0) = 0$. This result implies that the sum over all angles in a geodesic triangle is larger as π given re-convergence of geodesics.

Finally, we arrive at the model of the hyperbolic space $\mathbb{H}^n(r)$, $r > 0$, which we may derive out of $n + 1$-dimensional Minkowski as the hyperbola

$$\mathbb{H}^n(r) = \{x | x^\alpha x^\beta \eta_{\alpha\beta} = r\}.$$

This space is maximally symmetric with as symmetry group $SO(n, 1)$ which constitutes the $n+1$-dimensional Lorentz transformations; this Riemannian space has again constant sectional curvature which is equal to $-\frac{1}{r^2}$ as an explicit calculation in $n = 2$ shows. Geodesics expand henceforth according to $b\sinh\left(\frac{s}{r}\right)$ and therefore the shorter r te faster geodesics diverge from one and another. This has as a ramification that the volume of a ball of radius s grows faster as is the case in Euclidean space; concretely

$$V(s) = \int_0^s dtr^{n-1}\left(\sinh\left(\frac{t}{r}\right)\right)^{n-1} V(S^{n-1}).$$

The volume in the spherical case computes

$$V(s) = \int_0^s dtr^{n-1}\left(\sin\left(\frac{t}{r}\right)\right)^{n-1} V(S^{n-1})$$

and the reader is invited to calculate those integrals explicitly.

Taking all considerations into account, it is clear that the following result holds: if $(\mathcal{M}, h)$ constitutes an

n-dimensional Riemannian manifold with sectional curvature greater or equal to $R \in \mathbb{R}$ then one has that the volume of any ball with radius r is smaller or equal to that of the ball with identical radius in the model space $\mathbb{H}^n\left(\frac{1}{\sqrt{-R}}\right)$ or $S^n\left(\frac{1}{\sqrt{R}}\right)$ depending of whether $R < 0$ or $R > 0$ respectively. The reason is evident: consider any $x \in \mathcal{M}$ and take the Euclidean sphere with radius r on the tangent space and wharp that by means of the exponential map, then small angular cones $\delta\Omega$ which constitute a bundle of radial geodesics are deformed in such a way that the volume equals at most the one int the reference space, given that in general more compression as well as rotation takes place. One can write this decently down by means of the geodesic expansion equation.

One can try to achieve a similar result in Lorentzian geometry and study growth patterns of the so called Alexandrov sets $A(x, y)$ where $y \in J^+(x)$ and

$$A(x, y) = J^+(x) \cap J^-(y).$$

$A(x, y)$ is therefore the set of all points located on causal future oriented curves from x to y. One notices henceforth that there appears to be a connection between global propertiees of the Riemann tensor on one hand and topology on the other. This is indeed as such and more profound results may be obtained (see Milnor, Perelman). Topology from the differentiable viewpoint leads to rich results such as the Brouwer and Kakutani fix point theorems as well as Morse theory which we have treated already.

Chapter 14

Curved spacetime and bending of light rays: a taste of Relativity Theory.

This small chapter is a bit a loner in this book given that it regards applications in gravitational physics of the geometrical and analytical ideas studied previously. As mentioned, the Einstein field equations may be derived from the action principle

$$\int_{\mathcal{M}} R\sqrt{-g}$$

for a Lorentzian metric g and manifold without boundary $\mathcal{M}$. In case $\mathcal{M}$ has a boundary, it is mandatory to include the so called Hawking-York-Gibbons boundary term (integral) to compensate for the exterior derivatives appearing in the variation of the (bulk) Einstein-Hilbert action. One computes that

$$\frac{\delta\sqrt{-g}}{\delta g^{\alpha\beta}} = -\frac{1}{2}\sqrt{-g}g_{\alpha\beta}$$

as well as

$$\frac{\delta R}{\delta g^{\alpha\beta}} = R_{\alpha\beta} + g^{\gamma\kappa}\frac{\delta R_{\gamma\kappa}}{\delta g^{\alpha\beta}}.$$

The last tensor $\frac{\delta R_{\gamma\kappa}}{\delta g^{\alpha\beta}}$ is symmetric under exchange of α, β and γ, κ separately and therefore cannot contain the Riemann tensor. Closer inspection reveals that

$$g^{\gamma\kappa}\delta R_{\gamma\kappa} = \nabla_\alpha V^\alpha(\delta g_{\kappa\gamma})$$

where the vectorfield $V^\alpha(\delta g_{\kappa\gamma})$ depends in a linear fashion upon the covariant derivatives of $\delta g_{\kappa\gamma}$. The reader should try to see that this *has* to be true without making any computation (hint: the scalar needs to be of second order in the derivatives and of first order in the variation of the metric field). Consequently, one has that

$$\int (R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta})\delta g^{\alpha\beta}\sqrt{-g} + \nabla_\alpha V^\alpha(\delta g_{\kappa\gamma})\sqrt{-g}$$

and the last term may be rewritten as

$$\int \partial_\alpha(\sqrt{-g}V^\alpha)dx^1 \wedge \ldots \wedge dx^n$$

which equals the exterior derivative of some $n-1$ form. In case the manifold has no boundary $\partial\mathcal{M} = \emptyset$, this term vanishes due to Stokes theorem. The *vacuum* Einstein equations are henceforth given by

$$R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta} = 0$$

and the reader is invited to determine the Hawking-York-Gibbons boundary term (hint: study the notion of exterior curvature provided that infinitesimal information from the bulk around the boundary is of importance). The Einstein (symmetrical) tensor

$$G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta}$$

obeys

$$g^{\gamma\alpha}\nabla_\gamma G_{\alpha\beta} = 0$$

a conservation law which follows straight from the second Bianchi identity. In physics, this law has the significance of conservation of energy and momentum which is provided by a four current. It is important to understand that this conservation law is a geometrical identiy and holds for any Einstein tensor irrespective of the equations with or without matter. One immediately understands the proof of the Gauss Bonnet theorem which is that the Einstein tensor vanishes in two dimensions: indeed $G_{\alpha\beta}g^{\alpha\beta} = (1 - \frac{n}{2})R$ together with the identity $R_{\alpha\beta} = \frac{R}{2}g_{\alpha\beta}$ which only holds in $n = 2$, we arrive at $G_{\alpha\beta} = 0$. This implies that, apart from the boundary terms, Einstein's theory is *topological* in $n = 2$ meaning no metrics are distinguished. Physicists interpret this result by positing that no gravitational laws exist in $1+1$ dimensions; evidently, this is not the case in a higher number of dimensions. The vacuum Einstein equations may be simplified to

$$R_{\alpha\beta} = 0$$

implying that the Ricci tensor vanishes. Proof that the conservation law for the Einstein tensor is a direct consequence of the invariance under the action of a one parameter group of diffeomorphisms; in particular, it holds that in such a case

$$\delta g_{\alpha\beta} = \epsilon\mathcal{L}_{\mathbf{V}} g_{\alpha\beta} = 2\epsilon\nabla_{(\alpha}V_{\beta)}.$$

Therefore, the conservation law is a direct consequence of a continuous symmetry, a result which is known in general as Emmy Noether's theorem.

The Einstein equations are very hard to solve explicitly in general; under very general conditions it holds that there exists a unique solution up to a bulk diffeomorphism. Much more easy and straightforward is to engage into a post Newtonian expansion (this allows you to get rid of the nasty bulk diffeomorphisms); that is, write $g_{\alpha\beta} = \eta_{\alpha\beta} + \epsilon h_{\alpha\beta}$ on a manifold whose topology is that of $\mathbb{R}^n$ and write out the Einstein equations up to second order in ϵ. The resulting differential equations are known as the graviton equations.

In Einstein's theory, another central concept is geodesy which is explicitated by means of the geodesic equations; pointlike objects are completely characterized by means of their mass and move on geodesics. One could imagine different equations such as

$$\nabla_{\mathbf{V}}\mathbf{V} + \alpha\mathbf{R}(\nabla_{\mathbf{V}}\mathbf{V}, \mathbf{V})\nabla_{\mathbf{V}}\mathbf{V} = 0.$$

In case the object has an extension, one could make the latter dynamical by introducing the *expansion tensor*

$$\nabla_{(\alpha}V_{\beta)}$$

or *rotation tensor*

$$\nabla_{[\alpha}V_{\beta]}.$$

Light rays travel by good approximation on null geodesics, at least in the optical approximation (excluding quantum effects). One computes henceforth things such as change of frequency by parallel transport of the wave vector over the null geodesic and consequently projecting them on the tangent vectors of the world-lines of the distinct observers. Specifically, let k be a wave-vector tangent in a point x and y a point in the future connected by means of a null geodesic with x. Let n and m be the normalized tangent vectors to the world lines of two apparatus in x and y respectively, then it holds that the frequency of the wave as measured in x by the apparatus equals $g(k,n)/\hbar$ with $\hbar$ a universal constant. The frequency measured in y is given then by $g(k',m)/\hbar$ where k' equals the Fermi-Walker transport of k along the null geodesic in y. In that way, one computes red-shifts (meaning a lower frequency) due to interaction with the gravitational field.

Exercises : differential equations, existence and unicity of solutions
We have used in this chapter equations of the kind

$$\frac{df(x)}{dx} + P(f(x)) = 0$$

where $P(f)$ equals a polynomial of the kind

$$\sum_{i=0}^{n} g_i f^n$$

with g_i real valued functions in the real variable x. By choosing an $\epsilon > 0$ and positing that the differential $\frac{df(x)}{dx}$ in the ϵ approximation is given by $\frac{f(x+\epsilon)-f(x)}{\epsilon}$ one easily shows that all $f(n\epsilon)$ are fixed given $f(0)$. Take the limit for ϵ to zero and prove that the continuously differentiable closure exists and is unique in case all g_i are differentiable.

- Apply the same tactic to prove that equations of the form $\partial_t f(t, x^i) = P(f, \partial_i f, \partial_i \partial_j f, \ldots)$ have a unique solution given $f(t_0, x^i)$ for all x^i and chosen t_0.
- Generalize this theorem to equations of the form

$$g^{ij}(f,x)\partial_i\partial_j f + A^i(f,x)\partial_i f + P(f) = 0$$

and argue that a substantial difference exists regarding the qualitative behavior of the solutions in case g is Lorentzian or Riemannian. In the first case, the complex numbers are not important whereas in second case they are. This is because surfaces of the form $x^2 - y^2$ always have real zeroes whereas for any $y \in \mathbb{R}$ this is not the case for $x^2 + y^2$.

- To understand this last point, we study the nucleus of the operators $\partial_x^2 - \partial_y^2$ and $\partial_x^2 + \partial_y^2$. In the first case, we obtain the eigen-fuctions $e^{ik(x \pm y)}$ whereas in the second case they are $e^{kx \pm iky}$ with $k \in \mathbb{C}$. Usually, one considers real k due to bounded-ness properties of e^{iky} which constitutes the basis for the Fourier transformation. In the Lorentzian case, those functions remain bounded for all x whereas they blow up for Riemannian geometries. Stability of nature henceforth requires a Lorentzian metric.

Exercises: the Fourier transformation on Lorentzian spaces.
Fourier waves on Minkowski space time are of the form $e^{ik_a(x^a - x^a_0)}$; the huge advantage being that the wave vector k_a does not depend upon the space time coordinates and therefore has an interpretation as a conserved physical quantity called the momentum. It is therefore not appropriate to search for generalizations of those waves by considering the so called generalized d'Alembertian

$$g^{\alpha\beta}\nabla_\alpha\nabla_\beta$$

due to time dependency of the metric coefficients and henceforth, Fourier waves with distinct spatial characteristics intertwine in time. Fourier waves should therefore be defined starting from a space time point of view and not one of space.

- Consider a base point (source) x and a Fourier wave $\phi(x, y, k)$ where y is the point of detection and $k \in T\mathcal{M}_x$ the wave vector defined in x. Because k must define a pure wave vector in y there exists a linear mapping $\star(x, y) : T\mathcal{M}_x \to \mathcal{M}_y$ satisfying $\star(yx) \circ \star(xy) = \text{id}$. Arithmetically, it is reasonable to presume the existence of an operation $\Box$ and associated involution $\dagger$ such that

$$\phi(x,y,k)^\dagger = \phi(y,x,k_{\star(xy)}),\ \phi(y,x,k_{\star(xy)})\Box\phi(x,y,k) = \phi(x,x,k).$$

Moreover, one normalizes that $\phi(x, x, k) = 1_\Box$.

- For a normal Fourier wave, one has that $1_\Box = 1$ and as such is $\Box$ the multiplication of complex numbers and $\dagger$ the complex conjugate.
- Consider now the idea of propagation, that is that the reception of the wave at y is the consequence of information travelling over a path $\gamma : [0, 1] \to \mathcal{M}$ from $\gamma(0) = x$ to $\gamma(1) = y$. Show then that the most simple first order differential equation is of the form

$$\frac{d}{ds}\phi(x,\gamma(s),k) = i\mu g(k(s),\dot{\gamma}(s))\phi(x,\gamma(s),k)$$

with $\frac{D}{ds}k = 0$. μ is a constant here.

- Prove that on Minkowski, solutions are given by

$$\phi(x,y,k) = e^{i\mu k_a(y^a - x^a)}$$

and therefore independent of the chosen path. This is a topological property given that it is Riemann flat and as such has no local gravitational degrees of freedom.

- Let γ be a geodesic from x to y, then it holds that

$$\phi(x,y,k) = e^{-i\mu g(k,n)}$$

with $\exp_x(n) = y$.

Chapter 15

Further abstraction: Exercises

This chapter is entirely new and contains novel means of reflection about geometry. To summarize again, we started from the edifice of set theory, followed by number theory, theory of linear spaces and operators, differential geometry based upon the notion of a perfect exterior derivative d obeying $d^2 = 0$. We were somewhat extravagant is using coordinate dependent methods associated to local linear properties. We shall now forget about perfect exterior derivatives as well as local arithmetic properties inherited from function theory on $\mathbb{R}^n$. This implies we have to abandon a naive definition of the Lie bracket and we shall try to restitute it later on from the viewpoint of generalized connection theory.

Topological differentials.
Let X be any topological space (we do not insist upon it being metrical yet) and consider an equivalence relation $R \subset X \times X$ which is topologically open. R defines vectors, that is $(x, y) \in R$ is a vector connecting x with y; the correspondence to the usual vectors on a manifold being that (x, y) has to be thought of as the vector at x such that the image of the exponential map equals y, so they defined in a way relative to a metric and not a coordinate system. As said in the introduction, the notion of transport can easily be generalized and is defined by means of the following

$$\nabla_X : \{(x, y, z) : y, z \in R(x, \cdot)\} \to X \times X : (x, y, z) \to \nabla_{(x,y)}(x, z) = (y, w)$$

is called the transported relation regarding (x, z) over (x, y) from x to y and as such it indicates a preferred path or geodesic at least locally. ∇_X should obey the following further properties: (a) for any x, there exists an open O around it, such that $\{x\} \times O \subset R$ and such that for any $y, z \in O$ holds that $\nabla_{(x,y)}(x, z) \in R$, allowing one to define the composition of two transporters (b) ∇_X is continuous in the product topology (c) $\nabla_X(x, x, z) = (x, z)$, $\nabla_X(x, y, x) = (y, y)$ indicating that transport over the zero vector is the identity map and the zero vector gets transported into the zero vector. Before we proceed, it is useful to define two projections $\pi_1 : R \to X : (x, y) \to x$ and $\pi_2 : R \to X : (x, y) \to y$. We shall impose a further condition on R which is that for any x and sufficiently small neighborhood O around it, that for any y, z, p, q it holds that $(\pi_2(\nabla_{(x,y)}(x, z)), \pi_2(\nabla_{(x,p)}(x, q))) \in R$ meaning that for sufficiently small vectors sufficiently small vectors around a point, the resulting endpoints of the parallel transport again constitute a vector. Another, useful operation is the reversion P which maps (x, y) into (y, x), something which has to do with the linear structure of vectors. To localize, the reversion, we define $\tilde{P}(x, y)$ as $\nabla_{P(x,y)}(P(x, y)) \in R(x)$, so again, taking the minus sign is a geometrical operation. On R, it is now possible to define two kinds of (non-commutative) sums; the first one is mere composition, that is

$$(x, y) \circ (y, z) = (x, z)$$

being non local operation and the second one

$$(x, y) \oplus (x, z) = \nabla_{(x,y)}(x, z) \circ (x, y)$$

being a local operation. The reader notices that the reversion also defines a minus operation

$$(x, y) \ominus (x, z) = (x, y) \oplus \tilde{P}(x, z).$$

So, the reader understands that the local notion of a sum is a geometrical one and not one which merely originates from the manifold structure. Now, we can easily define the torsion functor

$$T : X \times R(x) \times R(x) \to R(x) : (x, y, z) \to ((x, y) \oplus (x, z)) \ominus ((x, z) \oplus (x, y))$$

and we shall prove that in a way this coincides with the usual definition in case y, z converge to x at the same rate. The Riemann function may be defined in a sufficiently small neighborhood of x as

$$R(x,p,q,r) = ((x,p) \oplus ((x,q) \oplus (x,r))) \ominus ((x,q) \oplus ((x,p) \oplus (x,r))).$$

The reader notices here that we did not include the commutator in this definition as we have no natural substitute for a vector-field, neither commutator and all drags are supposed to define commuting vector-fields anyway. We shall investigate these two definitions in further detail in the next section. There is no meaningful topological way to define this, you need a metric for that. Finally, we may consider functions between two metrical spaces $(X, d_X), (Y, d_Y)$ with vector structures R, T and transporters $\nabla_X,\ \nabla_Y$ defined upon it: we then say that $F : X \to Y$ is differentiable in a surrounding of $x \in X$ in case for any open $\mathcal{V} \subset T(F(x))$ there exists an open neighborhood $\mathcal{O} \subset R(x)$ such that the canonical bi-continuous mapping $DF(w,v) : (w,v) \in \mathcal{O}^2 \to \mathcal{V}^2,\ v,w \in \mathcal{O}$ defined by $(F(v), F(w)) = DF(v,w)$ satisfying

$$DF(((x,y) \oplus (x,w)) = DF(\nabla_{(x,y)}(x,w)) \circ DF(x,y)$$

also obeys

$$\frac{d_2(DF\,((x,y) \oplus (x,w)) \ominus (DF(x,y) \oplus DF(x,w)))}{\epsilon} \to 0$$

in case $d_1(x,y) = \epsilon a,\ d_1(x,w) = \epsilon b$, where $a, b > 0$ constants, which is the linearity condition. To define the torsion and Riemann "tensor", we need additional information. A connection is called weakly metric compatible if and only if

$$d(\nabla_{(xy)}(xz)) = d((xz))$$

which is, by itself insufficient to select for an "integrable" class of connections; for example, consider $\mathbb{R}^2$ with the standard Euclidean metric and define the connection $\nabla_{(x,y)}(x,z) = (y, y + R(z-x))$ where R is the rotation over the minimum of the angle θ between the vector $y-x$ and $z-x$ and $\pi - \theta$ in opposite orientation to the one defined by $z - x$ and $y - x$. Then the reader convinces himself that the angle is not preserved and that the torsion function vanishes identically. So, we must insist upon a stronger metric compatibility which says that the angles are preserved. For doing this, we need a path metric defined by the property that for any $x, y \in X$ it holds that there exists a $z \in X$ such that

$$d(x,z) = d(y,z) = \frac{d(x,y)}{2}.$$

The latter is equivalent to stating that there exists a curve, called a geodesic, $\gamma : [0,1] \to X$ which minimizes the length functional L for paths with endpoints x, y and, moreover, $L(\gamma) = d(x,y)$. The latter is defined by

$$L(\gamma) = \sup_{0 = t_0 < t_1 \ldots < t_n = 1, n > 0} \sum_{j=0}^{n-1} d(\gamma(t_j), \gamma(t_{j+1}))$$

and γ can be parametrized in arc-length parametrization by means of the Radon Nikodym derivative. Furthermore, this only makes sense if the geodesic connecting two points x, y close enough to one and another exists and is unique so that we can associate vectors to geodesics. Consider a point $x \in X$ and take a sequence of points y_n, z_n placed on two half geodesics emanating from x converging in the limit for n to infinity towards x. In case the limit

$$\lim_{n \to \infty} \frac{d(x,y_n)^2 + d(x,z_n)^2 - d(y_n,z_n)^2}{2d(x,y_n)d(x,z_n)}$$

exists, we define the angle $\theta_x(y,z)$ between both geodesics by equating the latter expression to $\cos(\theta_x(y,z))$. So, we must also require that ∇_X preserves angles; in short, $\theta_x(y,z) = \theta_p(\pi_2(\nabla_{(x,p)}(x,y)), \pi_2(\nabla_{(x,p)}(x,z)))$ for x, p, y, z sufficiently close to one and another. Obviously, this is still not enough given that one may consider the connection $\nabla_{(x,y)}(x,z) = (y, y - (z - x))$ and notice that $(x,y) \oplus (x,y) = (x,x) = 0$. The reader sees immediately that angles as well as distances are preserved and that the torsion vanishes since $(x,y) \oplus (x,z) = (x, y - (z - x))$ and $(x,z) \oplus (x,y) = (x, z - (y - x))$ so that

$$((x,y) \oplus (x,z)) \ominus ((x,y) \oplus (x,z)) = (x, y - (z - x)) \oplus \tilde{P}(x, z - (y - x)) =$$

$$(x, y - (z - x)) \oplus (x, y - (z - x)) = (x,x) = 0$$

since $\tilde{P}(x, z-(y-x)) = \nabla_{(z-(y-x),x)}(z-(y-x),x) = (x, x-(z-y)) = (x, y-(z-x))$. So, therefore we need to impose the strongest form, which amounts to an integrability condition which is that the d geodesics are auto-parallel curves meaning that for any geodesic γ from x to y in arc-length parametrization, it holds that

$$\nabla_{(\gamma(t),\gamma(s))}(\gamma(t),\gamma(s)) = (\gamma(s),\gamma(2s-t))$$

for $s > t$ sufficiently small. In that case, we find back the ordinary Levi-Civita connection with vanishing torsion in case for metrics on a manifold. To allow for torsion, one may impose that for any vector x, y sufficiently small, there exists a unique curve γ from x to y in arc-length parametrization such that for $t < s$ sufficiently small, the above condition holds. We shall henceforth insist upon the last integrability condition. To give a nontrivial example of our construction, take two manifolds glued together at a point p, with identified induced metrics on both meaning there exist two orthonormal basis at p which are identified by means of a linear mapping $T : T\mathcal{M}_p \to T\mathcal{N}_p : v \to T(v)$ and T^{-1} of course for the opposite directions. Then, for general vectors $a \in \mathcal{M}_p$ corresponding to a unique vector (p, x) and $b \in T\mathcal{N}_p$ corresponding to a unique vector (p, y), one can define $a \oplus b \equiv a \oplus_{\mathcal{M}} T^{-1}(b)$ in $\mathcal{M}$ resultng in a vector (p, z) and vice versa for $b \oplus a \equiv b \oplus_{\mathcal{N}} T(a)$. So, usually, the torsion function does not vanish, but it does so for infinitesimal vectors $a = \epsilon a'$, $b = \epsilon b'$ keeping a' and b' fixed. In the limit for ϵ to zero (as we shall show in full detail below) will $a \oplus T^{-1}(b)$ reduce to $\epsilon(a' + T^{-1}b') + O(\epsilon^2)$ so that in first order of ϵ, we have that

$$(a \oplus T^{-1}(b)) \ominus (b \oplus T(a)) = \epsilon(a' + T^{-1}(b') - T^{-1}(b' + T(a')) + O(\epsilon^2) = O(\epsilon^2)$$

and we will show below that even the second order term in ϵ vanishes in case the torsion tensors are anti-podal. Notice that differentiability is a priori a metric dependent concept but as the reader may verify, this is not the case for smooth metrics and general metric compatible connections defined by scalar products on a manifold. Here, the metric locally trivializes and the connection gives subleading corrections so that the sum reduces to the ordinary one. Let us work this out in full detail here so that the reader understands that the usual manifold definitions follow from ours. Given a metric tensor, $g_{\mu\nu}$ the reader verifies that the general connection is given by

$$\widehat{\Gamma}^{\delta}{}_{\mu\nu} = \Gamma^{\delta}{}_{\mu\nu} - \frac{1}{2}\left(T_{\mu\nu}{}^{\delta} + T_{\nu\mu}{}^{\delta} - T^{\delta}{}_{\mu\nu}\right)$$

where

$$T^{\delta}{}_{\mu\nu}$$

is the Torsion tensor which is anti-symmetric in $\mu\nu$ and in the previous expression, lowering and raising of indices has been done by means of the metric tensor. Now, take two vectors V, W at x, take $\epsilon > 0$ and consider the exponential map defined by ϵV, equivalent to (x, y) and ϵW, equivlent to (x, z) respectively. Up to second order in ϵ those are given by

$$y = x + \epsilon V - \frac{\epsilon^2}{2}\widehat{\Gamma}(V, V)$$

and likewise for W. Parallel transport of ϵW along ϵV gives

$$W(y) = \epsilon W - \epsilon^2 \widehat{\Gamma}(V, W)$$

and likewise for V, W interchanged. Hence,

$$\nabla_{(x,y)}(x, z) = \left(y, x + \epsilon V - \frac{\epsilon^2}{2}\widehat{\Gamma}(V, V) + \epsilon W - \epsilon^2\widehat{\Gamma}(V, W) - \frac{\epsilon^2}{2}\widehat{\Gamma}(W, W)\right)$$

and likewise for V, W interchanged. The reader notices that $\widehat{\Gamma}(V, V)$ can be retrieved from the geodesic equation and therefore $\widehat{\Gamma}(V, W)$ from the transport equation, both in order ϵ^2. We shall make this now precise. One sees now that

$$(x, y) \oplus (x, z) = \left(x, x + \epsilon(V + W) - \frac{\epsilon^2}{2}\left(\widehat{\Gamma}(V + W, V + W) + T(V, W)\right)\right)$$

implying that

$$\pi_2\left((x, z) \oplus (x, y)\right) = x + \epsilon\left(W + V - \frac{\epsilon}{2}T(W, V)\right)$$

$$-\frac{\epsilon^2}{2}\widehat{\Gamma}\left(W+V+\frac{\epsilon}{2}T(W,V),W+V+\frac{\epsilon}{2}T(W,V)\right).$$

Hence,

$$((x,y)\oplus(x,z))\ominus((x,z)\oplus(x,y))=(x,x+\epsilon^2T(W,V)+O(\epsilon^3))$$

so, as promised, the torsion tensor emerges in leading order ϵ^2. To make this precise in our setting, consider the generalized geodesics γ_y, γ_z in arclength parametrization representing with $\gamma_y(0)=x$, $\gamma_y(1)=y$ and likewise for γ_z. Furthermore, choose any reference direction γ_q then we have that with

$$\widehat{T}:=T(s):=\pi_2(T(x,\gamma_y(s),\gamma_z(s)))$$

that

$$\theta(\widehat{T},\gamma_q),\ \ \lim_{s\to 0}\frac{d(x,T(s))}{s^2}$$

are well defined and fully capture the Torsion tensor without coordinates. In order to find the Riemann tensor, we need to be a bit more careful and expand terms up to the third power of ϵ; more in particular,

$$(x,y):=\left(x,x+\epsilon V-\frac{\epsilon^2}{2}\widehat{\Gamma}(V,V)-\frac{\epsilon^3}{6}\left((V\widehat{\Gamma})(V,V)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),V)-\widehat{\Gamma}(V,\widehat{\Gamma}(V,V))\right)\right)$$

and

$$W(y)=W-\epsilon\widehat{\Gamma}(V,W)-\frac{\epsilon^2}{2}\left((V(\widehat{\Gamma}))(V,W)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),W)-\widehat{\Gamma}(V,\widehat{\Gamma}(V,W))\right)$$

so that

$$\pi_2\left((x,y)\oplus(x,z)\right)=x+\epsilon V-\frac{\epsilon^2}{2}\widehat{\Gamma}(V,V)-\frac{\epsilon^3}{6}\left((V\widehat{\Gamma})(V,V)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),V)-\widehat{\Gamma}(V,\widehat{\Gamma}(V,V))\right)+$$

$$\epsilon\left(W-\epsilon\widehat{\Gamma}(V,W)-\frac{\epsilon^2}{2}\left((V\widehat{\Gamma})(V,W)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),W)-\widehat{\Gamma}(V,\widehat{\Gamma}(V,W))\right)\right)$$

$$-\frac{\epsilon^2}{2}\left(\widehat{\Gamma}(W,W)-\epsilon\left(\widehat{\Gamma}(\widehat{\Gamma}(V,W),W)+\widehat{\Gamma}(W,\widehat{\Gamma}(V,W))-(V\widehat{\Gamma})(W,W)\right)\right)$$

$$-\frac{\epsilon^3}{6}\left((W\widehat{\Gamma})(W,W)-\widehat{\Gamma}(\widehat{\Gamma}(W,W),W)-\widehat{\Gamma}(W,\widehat{\Gamma}(W,W))\right).$$

We seek now for the associated geodesic of time ϵ which maps to this endpoint; that is we have to solve for

$$Z(V,W,\epsilon)=V+W-\frac{\epsilon}{2}T(V,W)+\frac{\epsilon^2}{6}K(V,W)$$

such that

$$x+\epsilon Z-\frac{\epsilon^2}{2}\widehat{\Gamma}(Z,Z)-\frac{\epsilon^3}{6}\left((Z\widehat{\Gamma})(Z,Z)-\widehat{\Gamma}(\widehat{\Gamma}(Z,Z),Z)-\widehat{\Gamma}(Z,\widehat{\Gamma}(Z,Z))\right)$$

equals the previous expression up to third order in ϵ. This leads to

$$(W\widehat{\Gamma})(W,W)-\widehat{\Gamma}(\widehat{\Gamma}(W,W),W)-\widehat{\Gamma}(W,\widehat{\Gamma}(W,W)+(V\widehat{\Gamma})(V,V)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),V)-$$

$$\widehat{\Gamma}(V,\widehat{\Gamma}(V,V))+3(V\widehat{\Gamma})(W,W)-3\left(\widehat{\Gamma}(\widehat{\Gamma}(V,W),W)+\widehat{\Gamma}(W,\widehat{\Gamma}(V,W))\right)+$$

$$3\left((V\widehat{\Gamma})(V,W)-\widehat{\Gamma}(\widehat{\Gamma}(V,V),W)-\widehat{\Gamma}(V,\widehat{\Gamma}(V,W))\right)$$

must be equal to

$$-K(V,W)-\frac{3}{2}\left(\widehat{\Gamma}(V+W,T(V,W))+\widehat{\Gamma}(T(V,W),V+W)\right)+((V+W)\widehat{\Gamma})(V+W,V+W)-$$

$$\widehat{\Gamma}(\widehat{\Gamma}(V+W,V+W),V+W)-\widehat{\Gamma}(V+W,\widehat{\Gamma}(V+W,V+W))$$

which leads to

$$K(V,W)=(W\widehat{\Gamma})(V,V)+(W\widehat{\Gamma})(V,W)+(W\widehat{\Gamma})(W,V)+(V\widehat{\Gamma})(W,V)-2(V\widehat{\Gamma})(V,W)-2(V\widehat{\Gamma})(W,W)$$

$$-\frac{5}{2}\widehat{\Gamma}(\widehat{\Gamma}(V,W),V)+2\widehat{\Gamma}(\widehat{\Gamma}(V,V),W)+\frac{1}{2}\widehat{\Gamma}(\widehat{\Gamma}(W,V),V)+$$
$$\frac{1}{2}\widehat{\Gamma}(\widehat{\Gamma}(W,V),W)+\frac{1}{2}\widehat{\Gamma}(\widehat{\Gamma}(V,W),W)-\widehat{\Gamma}(\widehat{\Gamma}(W,W),V)+\frac{1}{2}\widehat{\Gamma}(V,\widehat{\Gamma}(V,W))+$$
$$\frac{1}{2}\widehat{\Gamma}(W,\widehat{\Gamma}(V,W))+\frac{1}{2}\widehat{\Gamma}(V,\widehat{\Gamma}(W,V))+$$
$$\frac{1}{2}\widehat{\Gamma}(W,\widehat{\Gamma}(W,V))-\widehat{\Gamma}(V,\widehat{\Gamma}(W,W))-\widehat{\Gamma}(W,\widehat{\Gamma}(V,V)).$$

The kinetic term can be rewritten as

$$2\left((W\widehat{\Gamma})(V,W)-(V\widehat{\Gamma})(W,W)\right)+\left((W\widehat{\Gamma})(V,V)-(V\widehat{\Gamma})(W,V)\right)+2\left((V\widehat{\Gamma})(W,V)-(V\widehat{\Gamma})(V,W)\right)$$
$$+\left((W\widehat{\Gamma})(W,V)-(W\widehat{\Gamma})(V,W)\right)$$

which suggests for two distinct Riemann tensors and two derivatives of torsion tensors. Further computation yields that

$$K(V,W)=2\widehat{R}(W,V)W+\widehat{R}(W,V)V+2\widehat{\nabla}_V T(W,V)+\widehat{\nabla}_W T(W,V)+\frac{1}{2}T(V,T(V,W))+\frac{1}{2}T(W,T(W,V)).$$

The reader must note here that we used the following definition of the Riemann tensor

$$\widehat{R}(X,Y)Z=\widehat{\nabla}_X\widehat{\nabla}_Y Z-\widehat{\nabla}_Y\widehat{\nabla}_X Z-\widehat{\nabla}_{[X,Y]}Z;$$

Note also that $K(V,\lambda V)=0$ and the reader immediately calculates that

$$Z(S,Z(V,W,\epsilon),\epsilon)=S+V+W-\frac{\epsilon}{2}\left(T(V,W)+T(S,V)+T(S,W)\right)+$$
$$\frac{\epsilon^2}{6}(K(S,V+W)+K(V,W)+3T(S,T(V,W)))$$

and therefore

$$D(S,V,W,\epsilon):=Z(Z(S,Z(V,W,\epsilon),\epsilon),-Z(V,Z(S,W,\epsilon),\epsilon))=-\epsilon T(S,V)+$$
$$\frac{\epsilon^2}{6}\left(K(V,W)+K(S,V+W)-K(V,S+W)-K(S,W)\right)+$$
$$\frac{\epsilon^2}{6}\left(3T(S,T(V,W))-3T(V,T(S,W))+3T(S+V+W,T(S,V))\right)$$

and the expression of order $\frac{\epsilon^2}{6}$ reduces to

$$2\left(\widehat{R}(S,V)W+\widehat{R}(W,S)V+\widehat{R}(V,W)S\right)+6\widehat{R}(V,S)W+3\widehat{R}(V,S)V+3\widehat{R}(V,S)S+3\widehat{\nabla}_S T(V,S)+3\widehat{\nabla}_V T(V,S)$$
$$+\widehat{\nabla}_V T(W,S)+2\widehat{\nabla}_W T(V,S)-\widehat{\nabla}_S T(W,V)+3T(S,T(V,W))-3T(V,T(S,W))+3T(S+V+W,T(S,V))$$

In the absence of torsion, our vector-field reduces to

$$\frac{\epsilon^2}{2}(2\widehat{R}(V,S)W+\widehat{R}(V,S)V+\widehat{R}(V,S)S).$$

In general, the reader may enjoy observing that $D(S,V,W,\epsilon)=-D(V,S,W,\epsilon)$; in order to eliminate the quadratic terms in the above expression, it is useful to consider

$$E(S,V,W,\epsilon):=D(S,V,W,\epsilon)-D(S,V,-W,\epsilon)=$$
$$\epsilon^2\left(\frac{2}{3}\left(\widehat{R}(S,V)W+\widehat{R}(W,S)V+\widehat{R}(V,W)S\right)-2\widehat{R}(S,V)W+\frac{1}{3}\widehat{\nabla}_V T(W,S)\right)$$
$$+\epsilon^2\left(-\frac{2}{3}\widehat{\nabla}_W T(S,V)+\frac{1}{3}\widehat{\nabla}_S T(V,W)+T(S,T(V,W))+T(V,T(W,S))+T(W,T(S,V))\right)$$

so that we now have a tensor! The reader immediately notices that in the absence of torsion this expression reduces to

$$-2\epsilon^2 \widehat{R}(S,V)W$$

by means of the first Bianchi identity, so we would have isolated the Riemann curvature. In general, the first Bianchi identity reads

$$\widehat{R}(S,V)W + \widehat{R}(W,S)V + \widehat{R}(V,W)S =$$
$$T(T(S,V),W) + T(T(W,S),V) + T(T(V,W),S) + \widehat{\nabla}_S T(V,W) + \widehat{\nabla}_W T(S,V) + \widehat{\nabla}_V T(W,S)$$

so that the above expression reduces to

$$\epsilon^2 \left(-2\widehat{R}(S,V)W + \widehat{\nabla}_S T(V,W) + \widehat{\nabla}_V T(W,S) + \frac{1}{3}\left(T(S,T(V,W)) + T(V,T(W,S)) + T(W,T(S,V))\right)\right).$$

In order to get rid of the torsion terms, the reader may verify that

$$\frac{1}{3}\left(E(S,V,W,\epsilon) + E(W,S,V,\epsilon) + E(V,W,S,\epsilon)\right) =$$
$$\epsilon^2 \left(T(S,T(V,W)) + T(V,T(W,S)) + T(W,T(S,V))\right)$$

using the first Bianchi identity again. So, therefore

$$\frac{8}{9}E(S,V,W,\epsilon) - \frac{1}{9}E(V,W,S,\epsilon) - \frac{1}{9}E(W,S,V,\epsilon) = \epsilon^2 \left(-2\widehat{R}(S,V)W + \widehat{\nabla}_S T(V,W) + \widehat{\nabla}_V T(W,S)\right)$$

There is no way to further reduce this and eliminate the remaining derivatives of the Torsion tensor and the reader is invited to play a bit around and consider different sum operations in order to extract those. Finally, we return to the case without torsion, which is considerably easier and we now turn the prescription into our novel language; the reader may verify that to third order in ϵ our definition of $E(S,V,W,\epsilon)$ coincides with

$$E(x,p,q,r) := [((x,p) \oplus ((x,q) \oplus (x,r))) \ominus ((x,q) \oplus ((x,p) \oplus (x,r)))] \ominus$$
$$\left[\left((x,p) \oplus ((x,q) \oplus \tilde{P}(x,r))\right) \ominus \left((x,q) \oplus ((x,p) \oplus \tilde{P}(x,r))\right)\right]$$

and we have applied the same limiting procedure as we did tor the torsion tensor previously. The reader may repeat that exercise and define $E(x,p,q,r)(s)$ with $s \in \mathbb{R}_+$ and show that

$$d(E(x,p,q,r)(s)) \sim 2s^3 ||\widehat{R}(S,V)W||.$$

Considering the angle with a reference direction, the entire Riemann tensor may be retrieved in a coordinate independent way. Note also that we have a very nice "arithmetic" interpretation of torsion and curvature; that is, they express the failure of $\oplus$ to be commutative and perhaps associative to some extend. In the next section, we shall abandon the case with torsion and give an entirely different prescription for the Riemann tensor. This treatment shall be more basic and rough, which may not be a bad thing given the connections constructed so far are extremely subtle. We now finish this section by some comments upon differentiability and how the usual bundle apparatus of differential geometry may be generalized to our setting.

Given that we dispose of a local notion of a (non-commutative) sum whose infinitesimal version may very well become commutative and associative as explained previously and moreover, we have a natural notion of scalar multiplication by means of our generalized exponential map which associates to a vector (x,y) a unique geodesic γ in arc-length parametrization such that $\gamma(0) = x$ and $\gamma(s) = y$, then we define for any sufficiently small positive real number λ,

$$\lambda(x,y) = (x, \gamma(\lambda s))$$

and in case λ is negative we suggest

$$\lambda(x,y) := (-\lambda)\tilde{P}(x,y)$$

and the reader immediately verifies that these definitions induce the usual ones on the tangent bundle of a manifold. The reader should understand therefore, that it is natural to speak of directions at x defined by means of the geodesics (with respect tot the connection, so they don't need to be the geodesics of the metric) and that also in our general context of a non-commutative and non-linear sum meaning that

$$\lambda((x,y) \oplus (x,z)) \neq (\lambda(x,y)) \oplus (\lambda(x,z))$$

the very concept of a linearly independent and generating set of directions at x is still a well defined concept albeit I believe this does not imply that each vector can be written in a unique way by means of $\oplus$ and scalar multiplication. So, the concept of a basis is somewhat less restrictive but it is still well defined as a minimal set of independent and generating directions. The dimension is then an ordinary integer defined by the number of basis directions; these observations allow one to transport the entire cnstruction of tangent and cotangent spaces to our setting. But beware, we work very differently here as in the case of the ordinary theory; here it are the connections which determine the tangent bundle as well as its dimension, a much more intrinsic approach as the usual one where the backbone differential structure defines the connections. So, a linear functional, or covector, is defined by means of a continuous functional ω_X on the displacements (x,y) satisfying

$$\frac{1}{\epsilon}\left(\omega_X((x,z) \oplus (x,y)) - \omega_X((x,y)) - \omega((x,z))\right) = 0$$

and

$$\frac{1}{\epsilon}\left(\omega_X(\lambda(x,y)) - \lambda\omega_X((x,y))\right) = 0 \in \mathbb{R}$$

in the limit for $d(x,z) = \epsilon a$, $d(x,y) = \epsilon b$ for $a,b > 0$ constant and $\epsilon \to 0$. Note that we cannot request $\omega_X((y,z) \oplus (x,y)) = \omega_X((x,y)) + \omega((y,z))$ for finite displacements given that the sum operation allows for ambiguities non-locally. Furthermore, if ω_X were a field, then we could define it to constant meaning that

$$\omega_X(\nabla_{(x,y)}(x,z)) = \omega_X((x,z)).$$

Just as in ordinary functional analysis, we can define the weaker notions of continuity and differentiability of functions regarding convergence properties with respect to linear functions which all define semi-norms when suitably re-scaled in the infinitesimal limit given by

$$||(x,y)|| := |\omega_X((x,y))|.$$

All proceeds now in a fairly trivial way: given our geodesics (with or without torsion), we have, as mentioned before directions which are endowed with a natural notion of length and angles between them. You can consider generalizations of tensors in those directions which upon suitable re-scaling in the infinitesimal limit might become ordinary linear objects. We leave such developments to the reader.

15.1 Riemannian Geometry

In this section, we shall take a very different point of view as in the previous one; the latter was delicate and subtle and very much in line with the standard manifold treatment. Note that we have sidestepped the issue of existence of connections something which seems not totally obvious to prove and might be too delicate for practical purposes. For example, regarding hyperbolic spaces with conical singularities, it is rather obvious that no connection exists at the singular points. To give away the detail, take a couple of equilateral flat triangles (all angles having 60 degrees) and glue them together along their edges such that one has the situation where an interior vertex meets $n > 6$ triangles; in either the internal angle measure exceeds 360 degrees. Take now any half line starting from the vertex, then it will have an angle of π with all other half lines in a range of $(n-6)2\pi$. Obviously, it is impossible for any mapping to preserve angles when it returns to a normal region where the measure of the circle equals 2π. The situation is the reverse for conical spherical spaces where no mapping towards such points exist. Nevertheless, our coarse grained notion of curvature is still able to capture the curvature around such vertex whereas local curvature fails. I invite the reader to think about this; after all, the integrability condition was together with preservation of distances by far the most important criterion. But it is not sufficient either, so maybe we should be clever enough to find a weaker condition as the preservation of angles which amounts in the manifold case to precisely that. For example, a weaker criterion would be that the angles with the direction of propagation need to be preserved as well as the angles amongst themselves as long as both angles with respect to the direction of propagation are less than π. This definition would certainly fit all path metrical spaces and coincide with the usual lore of differential geometry. This does not change anything to what we have said in the previous section, but merely generalizes the setting to which it be applied. Nevertheless, the downside of the connection theory is that in general it is impossible to give a concrete prescription something which made the Christoffel connection so powerful. There are people who think you should give an easy prescription to calculate curvature even without constructing geodesics which might be a very daunting if not impossible task for a general path metric. Now, I am someone who is very found of geodesics, which are barely manageable in a general curved Riemannian space but I also sympathize with such an idea. The least you should know, I believe are distances and the work done in this section does precisely that. The price to pay is that we cannot speak any longer of vectors, but we have to directly calculate the scalar invariants.

With this in mind, we work now on general path metric spaces (X, d). We have the following definitions:

- Alexandrov curvature: in flat Euclidean geometry, the midpoint r of a line segment $[ab]$ satisfies

$$\vec{xr} = \frac{1}{2}(\vec{xa} + \vec{xb})$$

for any x. Hence, one arrives at

$$d(x,r)^2 = \frac{1}{4}(d(x,a)^2 + d(x,b)^2 + 2d(x,a)d(x,b)\cos(\theta_x(a,b))).$$

We define the non-local Alexandrov curvature as

$$R(x,y,z) = \frac{-2d(x,y)^2 - 2d(x,z)^2 + d(y,z)^2 + 4d(x,r)^2}{d(x,y)^2 d(x,z)^2 \sin^2(\theta_x(y,z))}.$$

Taking again geodesic segments between (x, y) and (x, z) parametrized by ϵ and corresponding to the vectors V, W respectively then, as before

$$y = x + \epsilon V - \frac{\epsilon^2}{2}\Gamma(V,V) - \frac{\epsilon^3}{6}\left((V\Gamma)(V,V) - \Gamma(\Gamma(V,V),V) - \Gamma(V,\Gamma(V,V))\right)$$

and

$$d(x,y)^2 = \epsilon^2 h(V,V)$$

by the very property of the exponential map. To find the midpoint between y and z we solve for

$$x + \epsilon V - \frac{\epsilon^2}{2}\Gamma(V,V) - \frac{\epsilon^3}{6}\left((V\Gamma)(V,V) - \Gamma(\Gamma(V,V),V) - \Gamma(V,\Gamma(V,V))\right)$$

$$+\epsilon Z - \frac{\epsilon^2}{2}\Gamma_y(Z,Z) - \frac{\epsilon^3}{6}\left((Z\Gamma_y)(Z,Z) - \Gamma_y(\Gamma_y(Z,Z),Z) + \Gamma_y(Z,\Gamma_y(Z,Z))\right) =$$

$$x + \epsilon W - \frac{\epsilon^2}{2}\Gamma(W,W) - \frac{\epsilon^3}{6}\left((W\Gamma)(W,W) - \Gamma(\Gamma(W,W),W) - \Gamma(W,\Gamma(W,W))\right)$$

leading to

$$Z := W - V + \epsilon\left(\Gamma(V,V) - \Gamma(W,V)\right) +$$

$$\epsilon^2\left(\frac{1}{2}(V\Gamma)(V,V) - \frac{2}{3}(V\Gamma)(W,V) + \frac{1}{3}(V\Gamma)(W,W) + \frac{1}{6}(W\Gamma)(V,V) - \frac{1}{3}(W\Gamma)(W,V)\right)$$

$$+\epsilon^2\left(\frac{2}{3}\Gamma(W,\Gamma(V,V)) - \frac{1}{3}\Gamma(W,\Gamma(W,V)) - \Gamma(V,\Gamma(V,V)) + \frac{1}{3}\Gamma(V,\Gamma(W,W)) + \frac{1}{3}\Gamma(V,\Gamma(V,W))\right).$$

This implies that the midpoint has coordinates, up to third order in ϵ given by

$$r = x + \epsilon\left(\frac{V+W}{2}\right) - \frac{\epsilon^2}{2}\Gamma\left(\frac{V+W}{2},\frac{V+W}{2}\right) - \frac{\epsilon^3}{6}\left(\frac{V+W}{2}\Gamma\right)\left(\frac{V+W}{2},\frac{V+W}{2}\right)$$

$$-\frac{\epsilon^3}{6}\left(\frac{1}{2}R(V,W)V + \frac{1}{2}R(W,V)W\right)$$

$$-\frac{\epsilon^3}{6}\left(-2\Gamma\left(\Gamma\left(\frac{V+W}{2},\frac{V+W}{2}\right),\frac{V+W}{2}\right)\right)$$

This shows that

$$d(x,r)^2 = \frac{\epsilon^2}{4}\left(h(V,V)+h(W,W)+2h(V,W)\right) - \frac{\epsilon^4}{6}h(R(V,W)V,W) + O(\epsilon)^6$$

and because

$$d(y,z)^2 = \epsilon^2\left(h(V,V)+h(W,W)-2h(V,W)\right) + \frac{\epsilon^4}{3}h(R(V,W)V,W)$$

the Alexandrov curvature equals

$$-\frac{h(R(V,W)V,W)\epsilon^4+\ldots}{3\epsilon^4(h(V,V)h(W,W)-h(V,W)^2)+\ldots}$$

which in the limit for ϵ to zero provides for $\frac{1}{3}$ times the sectional curvature. The reader might have guessed this result apart from the front factor based upon the symmetries of the Alexandrov curvature and the Riemann tensor.

- We now arrive to the notion of Riemann curvature; here, we shall have to take midpoints of midpoints. To understand why this is the case, consider the following expression

$$h\left(R\left(\frac{V+X}{2},\frac{W+Y}{2}\right)\frac{V+X}{2},\frac{W+Y}{2}\right) =$$

$$-\frac{1}{16}\left(h(R(V,W)V,W)+h(R(V,Y)V,Y)+h(R(X,W)X,W)+h(R(X,Y)X,Y)\right)+$$
$$\frac{1}{4}\left(h\left(R\left(\frac{V+X}{2},W\right)\frac{V+X}{2},W\right)+h\left(R\left(V,\frac{W+Y}{2}\right)V,\frac{W+Y}{2}\right)\right)$$
$$+\frac{1}{4}\left(h\left(R\left(X,\frac{W+Y}{2}\right)X,\frac{W+Y}{2}\right)+h\left(R\left(\frac{V+X}{2},Y\right)\frac{V+X}{2},Y\right)\right)$$
$$+\frac{1}{8}h(R(V,Y)X,W)+\frac{1}{8}h(R(X,Y)V,W)$$

Now, to undo the symmetrization in the curvature terms

$$\frac{1}{8}h(R(V,Y)X,W)+\frac{1}{8}h(R(X,Y)V,W)$$

note that by means of the Bianchi identity, this can be rewritten as

$$-\frac{1}{4}h(R(Y,X)V,W)+\frac{1}{8}h(R(V,X)Y,W)$$

so that we have broken the coefficient symmetry. Considering therefore the expression

$$h\left(R\left(\frac{V+X}{2},\frac{W+Y}{2}\right)\frac{V+X}{2},\frac{W+Y}{2}\right) - h\left(R\left(\frac{V+Y}{2},\frac{W+X}{2}\right)\frac{V+Y}{2},\frac{W+X}{2}\right) =$$

$$-\frac{1}{16}\left(h(R(V,Y)V,Y)+h(R(X,W)X,W)-h(R(V,X)V,X)-h(R(Y,W)Y,W)\right)+$$
$$\frac{1}{4}\left(h\left(R\left(\frac{V+X}{2},W\right)\frac{V+X}{2},W\right)+h\left(R\left(V,\frac{W+Y}{2}\right)V,\frac{W+Y}{2}\right)\right)$$
$$-\frac{1}{4}\left(h\left(R\left(\frac{V+Y}{2},W\right)\frac{V+Y}{2},W\right)+h\left(R\left(V,\frac{W+X}{2}\right)V,\frac{W+X}{2}\right)\right)+$$
$$\frac{1}{4}\left(h\left(R\left(X,\frac{W+Y}{2}\right)X,\frac{W+Y}{2}\right)+h\left(R\left(\frac{V+X}{2},Y\right)\frac{V+X}{2},Y\right)\right)$$

$$-\frac{1}{4}\left(h\left(R\left(Y,\frac{W+X}{2}\right)Y,\frac{W+X}{2}\right)+h\left(R\left(\frac{V+Y}{2},X\right)\frac{V+Y}{2},X\right)\right)$$
$$+\frac{3}{8}h(R(X,Y)V,W)$$

which is the result we needed. Denoting by $\widehat{(y,z)}$ the midpoint between y,z, we arrive at the following prescription for the curvature

$$S(x,y,z,p,q)=-8\left(S(x,\widehat{(y,p)},\widehat{(z,q)})-S(x,\widehat{(p,z)},\widehat{(y,q)})\right)$$
$$-\frac{1}{2}\left(S(x,p,z)+S(x,y,q)-S(x,p,y)-S(x,z,q)\right)$$
$$+2\left(S(x,\widehat{(p,y)},q)+S(x,\widehat{(q,z)},p)-S(x,\widehat{(p,z)},q)-S(x,\widehat{(y,q)},p)\right)$$
$$+2\left(S(x,\widehat{(z,q)},y)+S(x,\widehat{(p,y)},z)-S(x,\widehat{(q,y)},z)-S(x,\widehat{(p,z)},y)\right)$$

where

$$S(x,y,z)=-2d(x,y)^2-2d(x,z)^2+d(y,z)^2+4d(x,\widehat{(y,z)})^2.$$

The reader verifies that all symmetries of the Riemann tensor hold, meaning

$$S(x,y,z,p,q)=-S(x,z,y,p,q)=-S(x,y,z,q,p)=S(x,p,q,y,z)$$

and

$$S(x,y,z,p,q)+S(x,p,y,z,q)+S(x,z,p,y,q)=0.$$

This concludes our definition of the Riemann tensor.

- We shall now first define a notion of measure attached to any metric very much like the canonical volume element attached to a Riemannian metric tensor; there are several ways to proceed here. Define for any subset $S\subset X$, the outer measure of scale $\delta>0$ and dimension d as

$$\mu^d_\delta(S)=\inf\{\sum_i r_i^d|B(x_i,r_i)\text{ is a countable cover of open balls of radius }r_i<\delta\text{ around }x_i\text{ of }S\}.$$

Obviously, the $\mu^d_\delta(S)$ increase as δ decreases so we define

$$\mu^d(S)=\lim_{\delta\to 0}\mu^d_\delta(S).$$

The reader verifies that this defines a measure on the Borel sets of X and moreover $\mu^d(S)$ is a decreasing function of d which is infinity for $d=0$, in case X does not consist out of a finite number of points, and 0 for $d=\infty$. Upon defining α as

$$\alpha=\inf\{d|\mu^d(X)=0\}=\sup\{d|\mu^d(X)=\infty\}$$

an equality which holds as the reader should prove and it is $\mu^\alpha(S)$ which is of interest. α is called the Hausdorff dimension of X. I invite the reader to "localize" this concept such that one can speak of the local dimension of a space at a point and not just a global one.

- We define now a one parameter family of "scalar products" by means of

$$g^\epsilon(x,a,b)=\frac{d(x,a)d(x,b)\cos(\theta_x(a,b))}{\epsilon^2}.$$

The reader notices the scaling here as we shall be interested in taking the limit for ϵ to zero in a well defined way. Note that we could replace the metric compatibility of our connections in the previous section by the single demand that $g^\epsilon(x,a,b)$ is preseved under transport meaning that

$$g^\epsilon(y,\pi_2(\nabla_{(x,y)}(x,a)),\pi_2(\nabla_{(x,y)}(x,a)))=g^\epsilon(x,a,b).$$

We want now, in full analogy with the standard treatment in differential geometry define contractions of the Riemann "tensor" in order to construct the Ricci and Einstein tensor. Note that we do not necessarily dispose of a connection here and therefore we have no addition of vectors, seen as defining a direction. Therefore, we cannot rely upon the notion of a dual tensor associated to our functionals defind in the previous section. Nevertheless, we want to construct a notion of inverse which coincides in the latter cases with the more advanced linear concept. To set the ground for this discussion, note that there exists a natural generalization of the Dirac delta function regarding the Hausdorff measure. That is, there exists a symmetric $\delta(a,b)$ such that for all continuous functions f on X, it holds that

$$\int_X d\mu^\alpha(a)\delta(a,b)f(a) = f(b).$$

Defining now the nonlinear dual $\widehat{a}$ of a as

$$\widehat{b}(a) = \delta(a,b)$$

we define inverses $g^\epsilon(x,\widehat{a},\widehat{b})$ as

$$\frac{\int_{B(x,\epsilon)} d\mu^\alpha(b)g^\epsilon(x,\widehat{a},\widehat{b})g^\epsilon(x,b,c)}{\mu^\alpha(B(x,\epsilon))} = \delta(a,c).$$

The existence of a uniqueness of the inverse follows from the fact that the former defines a Toeplitz operator with trivial kernel. It is to say, $g^\epsilon(x,\widehat{a},\widehat{b})$ is the standard Green's function of the metric regarding the Hausdorff measure. This holds of course only if the measure is well behaved and we leave such details to the reader.

Prior to defining contractions with the metric tensor, remark that

$$\int_{B(x,\epsilon)}\int_{B(x,\epsilon)} d\mu^\alpha(b)d\mu^\alpha(a)g^\epsilon(x,\widehat{a},\widehat{b})g^\epsilon(x,b,a)$$

is ill defined and requires "a point splitting" procedure to obtain a well defined answer. Concretely, we consider

$$\alpha\frac{\int_{B(x,\epsilon)}\int_{B(x,\epsilon)} d\mu^\alpha(b)d\mu^\alpha(a)\int_{B(a,\delta)} d\mu^\alpha(c)g^\epsilon(x,\widehat{c},\widehat{b})g^\epsilon(x,b,a)}{\mu^\alpha(B(x,\epsilon))^2}$$

an expression which is independent of $\delta > 0$. Note that the dimension α has been added here to restore for the correct trace.

- The reader may now define the re-scaled Riemann curvature tensors $S(x,y,z,p,q,\epsilon) := \frac{S(x,y,z,p,q)}{\epsilon^4}$ and consider contractions with $g^\epsilon(\widehat{y},\widehat{q})$ to define the Ricci tensor $S(x,z,p,\epsilon)$ and from thereon the Ricci scalar. We leave this as an exercise to the reader.

15.2 The Lorentzian theory.

The matter now is how to generalize the above theory towards spaces with a Lorentz metric. We henceforth consider spaces (X,d) with a compact topology such that $d : X \times X \to \mathbb{R}^+$ is continuous and obeys

- $d(x,y) \geq 0$ and $d(x,x) = 0$
- $d(x,y) > 0$ implies that $d(y,x) = 0$
- $d(x,y) > 0$ and $d(y,z) > 0$ implies that $d(x,z) > 0$.

From d, we construct a chronology relation $y \in I^+(x)$ if and only if $d(x,y) > 0$ where $I^+(x)$ constitutes the set of all happenings in the chronological future of x. Likewise, one defines the chronological past $I^-(x)$ consisting out of events y such that $d(y,x) > 0$. Associated to this is a partial order $\prec$ defined by means of $x \prec y$ if and only if $d(x,y) > 0$. We suppose that for ope $\mathcal{O}$ around x one finds points y,z such that $y \prec x \prec z$ and $I^-(z) \cap I^+(y) \equiv A(y,z) \subset \mathcal{O}$. The sets $A(x,y)$ called the Alexandrov sets define a basis for the space time topology. Regarding the Riemann tensor, it is required that $a,b \in I^-(c) \cap I^-(d) \cap I^+(x)$ where time-like geodesics are defined by means of a maximization instead of minimization procedure. Likewise, it may be that $c,d \in I^-(a) \cap I^-(b) \cap I^+(x)$ or two similar options with a,b,c,d in the past of x. Because Lorentzian geometries define no canonical *local* compact neighborhoods, it is impossible to define a Hausdorff measure starting from Alexandrov neighborhoods. For example, on a piece-wise linear Lorentzian manifold with conical singularities the volume of an Alexandrov set is direction dependent. Consequently, it is better to define an additional metric $\tilde{d}$ as well as a Lorentzian metric tensor $g^{\pm}_{\epsilon}(a,b)$ on pairs of points $(a,b) \in I^{\pm}(x)$ for which holds that $d(a,b) > 0$ or $d(b,a) > 0$ such that hyperbolic angles are well defined (replace cosine and sine by cosine-hyperbolic and sine-hyperbolic). Call these regions $Z^{\pm}(x)$, then we define an inverse $g^{\pm,\epsilon}(\hat{a},\hat{b})$ by means of integration over $(B(x,\epsilon) \times B(x,\epsilon)) \cap Z^{\pm}(x)$.

15.3 Fuzzy logic.

In spoken language, one has reasons of reasons and the latter are sometimes infinite and circular. Let us give an example: "an apple falls down on the earth" WHY? "because of the gravitational field" WHY? "Because the laws of physics are background independent and involve a (pre) geometry" WHY? "because you need to be able to say what a straight line is and because nothing is static in this world" WHY? "If you could not, then you cannot tell the difference between a free object and one which is acted upon and because the notion of freedom isn't absolute" WHY? ... Obviously, this sequence is either never ending or ends with a dogmatic statement we suppose to be true. In mathematics, we have allowed for such argumentation by means of the logical operator "implies" of the words "if A then B" and every situation of this kind can be written as a sequence of such sentences. Also, we have in spoken language the words "by means of" for example: a chemical substance A changes into B by means of the catalyst C. Here, a mathematician would try to say that there is an equivalence with the statement "if A and C react then B and some leftover D remain". However, sometimes it is just not known what the leftovers are and neither is the mechanism by which C serves as a catalyst for A to change. The attitude of the mathematician is that this imperfect knowledge is just due to a limitation of our knowledge and is not intrinsic by any means and he would still write

$$A + C \to B + D.$$

But what if *nature* would be such that no precise statements can be made, not even probabilistically? Then we could not write it down in this way and we would have to invent a new logical operator in order to accomodate for the meaning of this sentence; the latter is noted down as

$$A \overset{C}{\to} B$$

where we ignore the leftovers. This operation is intended to mean "A evolves into B by means of the catalyst C" or "A implies B if C amongst others holds" where we just don't know the others or perhaps don't know if others are needed in the first place. This is the principle of incomplete knowledge which I think one needs to import into mathematics because as far as I know nature operates in this way by means of our free will. The reader automatically notices that it is possible for

$$A \overset{C}{\to} B \,\text{and}\, A \overset{C}{\to} \neg B$$

to hold where $\neg$ can be interpreted in the classical or intuitionistic sense. A mistake which is commonly made in science is that $A \to B$ is interpreted to mean that A is a cause for B, or that A is a reason for B to hold. Such interpretations however are wrong since it is not (experimentally) possible to verify a reason for something to occur; the only thing we can measure are coincidences. For example, it is equally possible for angels to move the planets the way they do than it is for gravity to do the job; the laws of gravity merely establish the way in which the motion of the earth around the sun occurs but it provides no reason or cause for it. One expects the following rule to hold

$$\left(A \overset{C,D}{\to} B\right) \to \left(A \overset{C}{\to} B\right)$$

in either further specification narrows down the implication. One does not have that

$$(A \to B) \to \left(A \xrightarrow{C} B\right)$$

is true since for $C = \neg B$, the right hand side is always false. We call C *compatible* or of no influence if this sentence is true. In case also the reverse implication holds

$$\left(A \xrightarrow{C} B\right) \to (A \to B)$$

we call C *redundant* or unnecessary. For example, A is H_2O and B liquid water and C is 50 degrees centigrade; since the molecule H_2O is always liquid water at this temperature, this information is redundant. In case C is an influence, we call the latter maximal or complete if

$$\left(A \xrightarrow{C} B\right) \to (A \wedge C \to B)$$

is true while the implication

$$(A \wedge C \to B) \to \left(A \xrightarrow{C} B\right)$$

is true by definition. While the sentence

$$(A \wedge C \to B) \to (A \wedge \neg C \to B)$$

is not always true, it can be that

$$\left(A \xrightarrow{C} B\right) \to \left(A \xrightarrow{\neg C} B\right)$$

is true and the reader should give an example of this (for example when C and $\neg C$ are redundant).

Chapter 16

Afterword.

The author of this book may be consulted by email on johan.noldus@gmail.com and is permanently prepared to revise as to expand the material. Suggestions of readers as to include new topics are gratefully considered and further discussed. I hope that in this work, which contains many new results of top level, the critical mind of the youngster is enthused. A reference work, written in English about fundamental physics is in production and published soon.

Postal address:
Dr. J. Noldus
Tel: 0032485338384
Email: johan.noldus@gmail.com