Generally covariant quantum theory.

Johan Noldus[1]

July 3, 2018

[1]Johan.Noldus@gmail.com, Relativity group, department of mathematical analysis, university of Gent, Belgium.

Contents

Some words upfront.

This book is written for future students as a guide into the physical world and I intend this in the broadest sense possible. No generation should share the confusion of the previous one and, at a certain age, it is our duty to succinctly present our findings for sake of the future of humanity. We should do this from the standpoint of a childish, curious, spirit who tries to conquer the world around him or her. Therefore, I have made the effort to keep the presentation as complete and elementary as possible, to introduce the mathematical language employed to a large extend and to present the arguments at a high level of detail. In principle, every bachelor in the sciences should be able to read this book but a very bright and ambitious student in his last year in high school might understand a large portion of the material also. Indeed, chapters one and two are mainly on the level of an intelligent 16 year old who is willing to work for knowledge whereas chapter three starts from a high school level of mathematics in a science program and brings you in some 50 pages to a pretty decent level of knowledge in mathematics. Every bachelor in science should be able to make that step but maybe an ambitious and hard-working 18-year old student could do that too. Chapters four till eleven present the main theme of this book and the reader should be able to follow all details having gone through the previous chapters.

Obviously, I will make some comments which require a higher level of mathematics but these should not disturb the reader and I shall try to make them in a way that they are intuitively evident. They should stimulate the reader to go beyond this book and try to find out things for him or herself. In the long run, it is my intention to rephrase the material in this book in a way which is still more accessible so that more people can enjoy its content and reflect in a deeper way about nature than I did. For now, this is the best I can do regarding aspects of the presentation as well as the content; its purpose is to convey the spirit of scientific research as well as its importance for humanity.

Chapter 1

Introduction.

The intention of this book is to dig deep in the first principles of physics and develop, what is known as, a relativistic quantum theory from scratch. By this, I mean a theory for the so-called elementary particles which constitute building blocks of nature which cannot, by hypothesis, be subdivided themselves into smaller parts. You might want to suggest that everything can be broken into smaller pieces if you try hard enough and that the Greek philosophy of "atoms" is merely an illusion. This may be so, but nobody managed yet into breaking the electron since its discovery around 1900 and the reader will figure out that we can do very fine by making the assumption that the electron can be described by a mathematical point. This brings us immediately to one of the main lessons of the Copenhagen philosophy which is that a theory is just a theory and it may be impossible to figure out the reality behind it all, if there exists one which we can comprehend with the human mind. No prior knowledge of non-relativistic quantum mechanics or quantum field theory is assumed although it wouldn't hurt in understanding some of the comments in the fourth chapter regarding the non-covariance of quantum mechanics and "quantum field theory". I often get the comment, even from the so-called specialists, that my work is hard to read and understand given that it is interdisciplinary and uses advanced techniques or ideas from different fields such as mathematical relativity, operational quantum field theory, global geometrical analysis and so on. I am of course aware that those mathematical fields are *not* on the standard curriculum at university and I know of a good deal of university professors who do not master at least one of these subjects; this includes the so-called specialists who are knowledgeable of one subject but entirely incapable of arguing in a different field in a different language. It takes courage and a good deal of patience to write up a comprehensible account of a collection of research papers which span in total some 170 pages; a size which is just enough to obtain some results of importance and not too much in order not to overload the reader with the gigantic task of reading through more than 500 pages in order to gain some insight into the matter. This book will be written in a unique personal brand which involves a mixture of historical comments and modern approaches from *my* point of view:

it is by no means so that the topics which I mention and which fall outside the main line of this book are dealt with in a way as envisioned by one of the historically important founding fathers. Rather, it is my personal view on these matters, which has been formed after long years of reflection, that is presented. Readers interested in historical development should consult the original papers or read a conventional textbook on these matters. This book is the result of recent research and is therefore at the front line of science: it is obviously my hope that it will become a standard reference work one day.

Particle physicists come in different guises, but most of them still have little knowledge about the geometrical aspect of gravitation. This is expressed in the worshipping of Richard Feynman, a rather inventive physicist, but completely unfamiliar with "modern" notions of geometry even in the 1960 ties. Einstein, in 1915 knew more about geometry, through communication with Elie Cartan, than Feynman did: as is always the case, praising ignorance is on the boundary of stupidity and it is with great regret that I see some people taking on a similar attitude in the last years. Now, things have evolved a little over the years of course, but it remains a fact that most people's comprehension of geometry is limited to local tensor calculus and some hear-say about global topological considerations. Very few physicists have tackled modern mathematical work on global geometrical aspects of physical problems and most are therefore completely unaware of the appropriate language to deal with these issues. This is pretty bad as the language one speaks often shapes the ground for ones ideas to form.

The ideas expressed in this book are of such fundamental importance to physics, and therefore to society, that in my wildest dreams, I would turn it into a book for the intelligent 16 year old high-school student. I remember very well that at the age of 13, I was interested in the genesis of things, why is the world and we humans in particular, the way it is? What is the purpose behind it all and in what language can we adequately express these things? In those days, I was in my philosophical mood and an ardent listener of the Germanic classical work: Wagner, Bruckner and the magical eye of Wilhelm Furtwangler were high on my list. Likewise were Freud and his evil student Jung: the human mind has fascinated me ever since and if I am lucky enough, I will write another book about that. Sad is ones destiny given the fact that one has to choose a particular thing to study: luckily enough, I took physics, the science to which every other science comes second. I could write an entire book about the importance of physics for society, for its structure, its policies, its comfort, glory and everything which is associated to human progress, but this is neither the place nor the time to do that. Being a physicist is a responsible job, it is hard to speak about the unintelligent things and try to figure out laws these "objects" or maybe subjects have to satisfy. Take an electron for example, now, there is nothing in the world which would tell me that an electron is not as complicated as a human being, perhaps electrons are sexual too and have babies! This old idea of "scale invariance", meaning that physics is the same on a wide range of scales is quite far reaching indeed and there is no a-priori reason why we should have success in figuring out the behaviour of electrons. Perhaps, when they are close together in some sense, they might show intelligent and socializing capacities but we should start from investigating them when they are "isolated" in a reasonable sense. At least, this is the classical viewpoint: the problem resides in the word *to be*. Now, at this point, you might want to turn on the alcohol level in your bloodstream and ask yourself whether you can a-priori demand that some particle is at a fully specified event in spacetime when no other material subject in the universe sees it as being so. In other words, in the absence of any "connection" to another subject, *is* an electron at a well specified place? We, humans, would say it is rather obvious that we are always "somewhere" and "connected" to something else such as a chair, a bed, the floor or at least some object but electrons have more "potential" to be free than we are, so for them the answer may be different. At the beginning of the 20'th century, a most dramatic experiment has been performed shedding *some* light on this issue: one took an electron gun pointed at a distant screen, separated from the gun by means of a plate with two distant slits, and found that by shooting one electron at a time, the arrival density on the screen was one corresponding to the energy density of a wave instead of what they expected from a point-like particle. In either, they found an interference pattern! Now, you can assume that the surrounding through which the electron "travels", provided we can even speak in such terms, is a kind of medium, where at some distant point, the particle has "information" about the environment and eventually the previous particles passing there as well as both slits in the plate in particular. You see that such scenario becomes pretty complicated soon and what should happen if we *cannot* assume that a particle *travels* and perhaps does not even exist after it leaves the gun? These are hard questions indeed and frankly speaking, we do not have a unique answer to these issues. Now, things become even more complicated: if I would like to determine whether a particle goes through the right slit and not the left by means of some light bulb, then it turns out that this knowledge destroys the interference pattern what should make one conclude that the medium also remembers the particles *history* given that a particle that goes unobserved through the left slit does not take into account the data associated to the previous detected particle(s) through the right slit. So, a particle *knows* if another one has been detected or not or, at least, the presence of the detector changes its behaviour even though the particle and detector might not "connect".

Let me take a sip of wine, we are still talking here about electrons, remember? As I said, the issues associated to this experiment have not been settled yet but physicists were more than happy to find *one* theory which would also work in the case electrons get bounded, for example to a nucleus. Indeed, the double slit experiment by itself does not really show in which direction we should think, but it might help if we find an equation which would lead to predictions for other circumstances which would then, again, be confirmed by experiment. This would still not reveal the true mechanism behind what is really happening, neither would it answer the question if the particle travels or not, but at least it would make *predictions* which we can test experimentally. Of course, the mathematical form of the theory evoques some distinct interpretations, but none of those have ever been checked experimentally. More radical, *assuming* completeness of the theory regarding observable predictions, shows that we will never know which interpretation is right and which not: it is all a matter of metaphysics. At least, this was the attitude of eminent pre-war physicists who were convinced that the best a theory can do is to predict probabilities of a certain outcome of a given experiment, which begs the question who or what can perform an experiment. Can our electron not perform an experiment and measure if it arrives on the screen? The split these physicists introduced between observer and observed has been a major obstacle to form a quantum theory of the universe in which observer *and* observed should be described in the same language. An idea which has been rejected because it is rather complicated and counterintuitive, since many of these double slit experiments have been performed in vacuum, is that of a medium which carries the imprint of the history of particles having passed there. Such a theory of a spacetime with memory and intelligent "classical" electrons is of course still possible, but not very efficient.

There was another set of experiments, made with atoms which indicated that, at small time and distance scales, the behaviour of a single electron had to be very different from the one predicted by the classical physics known at that time. They measured that atoms come in layers: more specifically, the atom emits and absorbs radiation energy in discrete packages. Now, not only did the classical physicists of that time compute that the atom would be unstable on very short time scales, given that the accelerated electron would radiate out energy and therefore eventually bump on the nucleus, but there was no reason to expect why energy exchanges with a radiation bath would come in discrete packages. Such thing was mathematically known to occur only if something "nonlocal" is going on, meaning that the electron "knows" or "feels" the environment of and around the nucleus and reacts holistically. Now, it may be that the mathematical language at the time was not fit to deliver the proper physical ideas and indeed, recent computer simulations show that it is enough to assume a stochastic radiation background in order for the classical atom to have a nontrivial ground state. So, there are still people working in the field who have not given up yet upon "advanced classical ideas" in order to "explain" those phenomena regarding the double slit experiment and the atomic structure. I think this is a good thing and it is valuable to humanity to know several different points of view on the same matter. Now, we just anticipated that the mathematical language known at the beginning of the twentieth century suggested that something "nonlocal" or "holistic" had to be going on and I added the wisdom that this may be a "wrong" idea after all. "Wrong" is then to be used in the metaphysical sense but not in a practical one, given that our holistic viewpoint gives rise to predictions which accurately match observations; at least, this is so in weak gravitational fields on small time and distance scales where "small" means a couple of kilometers or so. The "brilliant" idea theoretical physicists had at the beginning of the 20'th century was that the wavelike behaviour exhibited by repeated single events in an electron double slit experiment was not due to an intelligent medium and alike electrons, but was present for any single electron! So, mathematically, they argued, we can replace the electron by the wave as long as it is not measured; such an holistic approach would also be capable of explaining the atomic spectra as was soon shown afterwards. So, the picture they worked towards was that an electron was double faced: when measured, it behaves as a particle, when it is free, it behaves as a wave but it is *unknown* what it *is*. Some researchers, who like to be metaphysical, would say that the electron really is a particle and that the wave somehow reflects "information" or knowledge available to the electron. This is a viable point of view, but the minimalists shamelessly use the argument that we will never know, given that their interpretation of the theory does not allow for such fantasies.

Indeed, the so-called Copenhagen clan under impetus of Werner Heisenberg, discovered an operational formalism which changed the wave when a measurement occurs; this change was irreversible so there is no way of tracking the electron as a particle without disturbing it. So their point of view was that if you see the electron as a particle, then it is because you measure it, not because it *is* a particle. It is unknown whether this point of view is really correct or not or even if measurements only come into one single category as the Copenhagers assume. To me, it appears that measurement might have many more fine details associated to it and we shall come back to that in chapter two on general philosophy of physics. A weak point of the Copenhagen interpretation is that they do never link a specific measurement apparatus to an observable quantity; that is, they do not define the conditions needed for some experiment to take place. This was of course realized by the founding fathers and the physicists at that time felt that this problem would be way too hard to tackle and that it had to be postponed to the future. So, we have now come that far that the very substance of an electron is questioned, but we accepted it shows intelligent and holistic behaviour. It is to me still amazing how a couple of simple experiments can lead to such different viewpoints and that we nevertheless all agree that the wave-particle picture does a good job in predicting outcomes of experiments even if the theory is incomplete. I recall that we are only dealing with electrons here, how much more complicated would it become if we would replace them by humans. Would it be possible to say something intelligent about that at all?

The reader must sense here already the excitement of what is going on; instead of sending the electron to the humanoid "electron-psychologist" in order to figure out the motives and ground rules behind its behaviour and label if the latter is good or bad, physicists largely gave up on that idea and contented themselves with statistical predictions, given a certain experimental setup. Many believe this is the best they can do; in this regard, I do not care about this issue but wouldn't mind a conversation with an electron couple or two. Striking of this all is the shear humility of physicists; *even* for electrons, they have given up on the illusion that we might be capable of figuring out the individual, indeed even *defining* "it" or maybe better "he" or "she" is no easy matter and lots of debates around these issues are going on. Comparing this with the outright arrogance and stupidity of social scientists, who have the pretense of figuring out the human behaviour, it must become clear to you why proper physicists do not want to have anything to do with such "scientists". It is already bad enough that they reside in any university council. So, mister Noldus, I hear you think, what kind of theory for elementary particles will you present us with in this book? In this regard, I must shamelessly admit I am going to follow the *philosophy* of the Copenhagers and stay far removed from the "psychology" of the individual electron, photon, Higgs boson and so on. In other words, I am going to make it as easy as possible for me: that is already hard enough as the reader will notice!

In a short while, I will present more in depth the "Copenhagen" philosophy which I deem to be different from what most people mean with the Copenhagen *interpretation*. The reader might notice that I will present my own "interpretation" or "picture" which should by no means be taken literally given that I am *not* an (accredited) electron-psychologist. This picture will nevertheless be useful in "understanding" the meaning behind certain concepts such as locality and relativistic causality and the reader should cherish it in that respect. I will use some terminology which shall be fully explained in chapter three; the reader who is unfamiliar with it should simply absorb the ideas and flesh out the details later on. The main achievement of Werner Heisenberg constituted out of a very simple mathematical observation: a free wave $\Psi_k(x)$ with four momentum k^a is given in Minkowski spacetime, which we shall define in chapter three, by

$$\Psi_k(x) \sim e^{ik_a x^a}$$

and therefore, the operation of multiplying this function with k_a is equivalent to $-i$ times deriving it with respect to x^a, which we denote by $-i\partial_a$. In other words, the momentum "operator" P_a together with the position operators x^b satisfies the following Lie-algebra

$$\left[x^b, P_a\right] = i\delta^b_a,\ \left[x^b, x^a\right] = 0,\ [P_a, P_b] = 0$$

where $[X, Y] = XY - YX$ and the Lie-bracket obeys a desirable property

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

called the Jacobi-identity. We will say more about Lie-algebra's in chapter three. Since the wave of a free particle must have a geometrical significance, we should find one and generalize that to curved spacetime. This is the path taken in chapter six, but history unfortunately took another road. Indeed, physicists concentrated on the Heisenberg Lie-algebra and found a recipe related to the classical Poisson bracket algebra $\{f, g\}$. Indeed, as has been shown by Hamilton a few centuries ago, simple classical physics can be fully described by an energy function $H(q_i, p_j, t)$, called the Hamiltonian, and a bracket $\{f, g\}$ on functions

$f(q_i, p_j, t)$. Here, the q_i are thought of as the coordinates of the theory and the p_j the canonical momenta. The bracket is defined as

$$\{f, g\} = \sum_i \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q_i}\frac{\partial f}{\partial p_i}$$

and the reader is invited to show that it satisfies the Jacobi identity. In particular,

$$\{q_i, p_j\} = \delta_{ij} 1$$

and Dirac's idea constituted in making the simple replacement between the Poisson bracket and i times the commutator. Alas, this scheme is full of ambiguities due to the non-commutative nature of the operators x^b, P_a and the commutative[1] nature of the classical variables q_i, p_j. More in particular, the quantum replacement for a classical energy like qp^2 could be, for example, PxP or $\frac{1}{2}(xP^2 + P^2 x)$. The "operational" viewpoint however, in spite of its problems, has led to an interesting principle which I shall explain now. We shall prove in chapter three that from the commutation relations

$$\left[x^b, P_a\right] = i\delta^b_a$$

and some additional assumptions it follows that that

$$\Delta_\Psi x^a \Delta_\Psi P_a \geq \frac{1}{2}.$$

This formula means that if a particle is in a state Ψ, then the product of the uncertainty on the position times the uncertainty on the momentum in the direction a is larger than one half. For a free particle with momentum k in a state $\Psi_k(x)$, $\Delta_{\Psi_k} P_a = 0$ whereas $\Delta_{\Psi_k} x^a = \infty$ so this formula does not really work well for a free state and we shall fill in the details of that later on. The reader may wish to construct a state which fixes position but with an infinite uncertainty or spread on the momentum. It is this principle which shall be the cornerstone of *my* formulation of relativistic quantum theory too albeit in a somewhat different guise.

It is easy to understand, however, *why* this formulation of physics is not a good one and tied to flat Minkowski spacetime. This argument shall be worked out in far more detail later on but it is so simple that most readers will profit from it now. There are two things which matter: (a) as it turns out, elementary particles come with internal spin degrees of freedom and therefore wave functions are not really complex scalars but spinors $\Psi_{m_1 \ldots m_k}$ for a particle of spin-$\frac{k}{2}$ (b) in a curved spacetime, we should have equivalents of the plane waves too. Hence, *suppose* now that the x^a should be replaced with coordinates x^μ where the Greek index has a different significance, as I will show to you later on, than

[1] Two variables x, y are said to commute if and only if $xy = yx$.

the Latin one; then, the correct momentum operators in a curved spacetime background should be of the form

$$P_\mu(y) = -i\nabla^s_\mu(y) + \ldots$$

where μ is a *spacetime* index, ∇^s the spin derivative and y the spacetime variable. But then, we have that

$$[P_\mu(y), P_\nu(y)] \sim -\mathbf{R}_{\mu\nu}(y) + \ldots$$

where $\mathbf{R}_{\mu\nu}(y)$ is the so-called Riemann tensor at y defined with respect to the coordinate basis ∂^y_μ at y. This shows that a geometric formulation of quantum mechanics clashes with the Heisenberg Lie-algebra given the assumption that the Minkowskian coordinates used before should have no global physical significance. This shows that the Dirac recipe cannot be sustained given that each known action principle is defined from *local* spacetime quantities and physical Minkowskian coordinate systems are *globally* defined with respect to a reference point. This argumentation will be fleshed out in full detail in chapters three and four but it shows already that locality, in some sense, has to be given up. The reader should wonder why the founding fathers of quantum mechanics did not think of this argument back in 1920; the reason is historical. Einstein's relativity theory was then just discovered and I really did not exaggerate a while ago when stating that even contemporary physicists do not know geometry beyond some elementary level which allows them to make some computations.

For now, I will explain to you a geometric vision on the plane wave $\Psi_k(x)$ which should allow the reader to generalize this idea to any curved spacetime; in other words, we *define* the free particle in a spacetime which curves and bends. I will explain you this at an intuitive level, the necessary mathematical background being presented in chapter three and worked out in full detail in chapter six. Minkowski spacetime obscures many fine points regarding the correct way to define the notion of a free particle with four momentum k^a given that it is maximally symmetric. This means that it is translation invariant, in either spacetime looks the same at all points, and moreover isotropic, meaning that at any given point, it looks the same in all directions. This implies that the notion of a free particle created at some spacetime event x does coincide with the notion of a free particle created at some event z. That is,

$$\Psi_{k,x}(y) = e^{ik_a(z^a - x^a)}\Psi_{k,z}(y)$$

as the reader may easily verify. This is no longer true in curved spacetime where a free particle has to "travel" on hills and through valleys. This is the first lesson we should remember: the notion of a free particle depends upon the spacetime point where it is *created.* Indeed, we will speak in general terms about creation and annihilation of a particle at certain spacetime points and about free propagation of the *information* that such particle was created at x. So, the "Fourier wave"

$$\Psi_{k,x}(y) = e^{ik_a(y^a - x^a)}$$

where x denotes the point where the particle has been created, indicates the way how "information" about the birth of a particle at x propagates - even backwards in time- through spacetime. We will now come to a geometric picture how this propagation takes place; logically, it is something which originates at x. Let $\gamma(s)$, $s : 0 \ldots 1$, be a curve joining x to y and k be a four dimensional vector at x. What we will do is to drag k along γ and define $k(s)$ as the dragged vector defined at $\gamma(s)$. We want to derive the principles behind the propagation mechanism which allows one to write down an equation for $\phi_\gamma(x, k, \gamma(s))$ such that

$$\phi_\gamma(x, k, y) = \Psi_{k,x}(y).$$

This principle is easily found and I call it Lorentz invariance which means (a) that the modulus squared or "probability density" of the wave does not depend upon the wave vector k and, moreover, (b) $\phi_\gamma(x, k, x)$ is independent of k. This means that the simplest equation satisfying these demands looks like

$$\frac{d}{ds}\phi_\gamma(x, k, \gamma(s)) = i\dot{\gamma}^\mu(s)k_\mu(s)\phi_\gamma(x, k, \gamma(s))$$

with initial condition $\phi_\gamma(x, k, x) = 1$. We call this a relativistic Schrodinger equation and its solution in Minkowski spacetime does indeed *not* depend upon γ and is given by

$$\phi_\gamma(x, k, y) = e^{ik_a(y^a - x^a)}$$

which is a remarkable result indeed. We just derived that the notion of a free particle created at some point x follows from the principle of Lorentz invariance and does not depend upon the path of propagation given that Minkowski spacetime does not contain any local geometrical degrees of freedom and is therefore topological in some sense. This last result is of course no longer true in a general curved spacetime and we will have to select a very special class of curves along which propagation occurs. The reader is invited to think about a physical idea giving him precisely this class.

The reader should now object and utter fine, you just defined the notion of a free particle created somewhere, but how do you now define the probability that it can be annihilated somewhere else given the uncertainty principle. Indeed, our quantum theory shall be mainly derived from the function $W_m(x, y)$ which denotes the amplitude or propagator associated to the creation of a particle of mass m at x and the annihilation thereof at y. Clearly, we should build this object from $\phi(x, k, y)$ which indicates the creation of a free particle with momentum k at x. Now the uncertainty principle dictates that given the fixed position x we are totally uncertain about the momentum which means we should integrate over all momenta $\int d^4k$. Now, it is so, that the four momentum should point towards the future, which demands the insertion of $\theta(k^0)$ and that it satisfies Einsteins energy momentum relationship meaning $k^2 = m^2$ leading to a factor $\delta(k^2 - m^2)$. In total, this gives

$$W_m(x, y) \sim \int d^4k\, \delta(k^2 - m^2)\theta(k^0)\phi(x, k, y)$$

which is the correct formula derived in quantum field theory for a spin-0 particle in Minkowski spacetime, but is here generalized to *any* spacetime.

These very few elementary remarks constitute the very core of the definition of the dynamics of the theory: what remains to be discussed are the internal degrees of freedom of a particle, a principle telling you how to define interactions and finally, the rule for a probability interpretation. These subjects are carefully introduced and physically argumented in full detail in chapters five and six but are essentially no more difficult than the argument given here. The reason for me to postpone these issues is that I feel that they are easily understood once one has the technical baggage of chapter three under the belt as well as an in depth discussion on the definition of probability. Indeed, we will go beyond the standard Kolmogorov axioms which means we extend the Hilbert space formalism to be treated in that same chapter. I want to warn the reader upfront that my treatment of the internal degrees of freedom of a particle is different from standard ideas about these matters: for example, a photon will not only have states with helicity plus minus one but also a state of helicity zero corresponding to a polarization parallel to its four momentum k, the so-called longitudinal one. In standard quantum electrodynamics on a Minkowski background, these states can be ignored which basically means that helicity of a particle is a Lorentz invariant concept in the operational sense of Weinberg which relies upon non-degenerate representation spaces. I will argue that in the presence of non-trivial gravitational degrees of freedom, this is *not* the case and a state of helicity plusminus one in one local reference frame may lead to a different amplitude than a state of helicity plus minus one in another reference frame. It is just so that we demand this not to be the case if we would ignore *all* gravitational effects and go back to the standard Minkowski theory where the longitudinal state is *dynamically* irrelevant and therefore decouples from the theory. The latter is just an unphysical idealization where one would imagine elementary particles to move in an absolute vacuum where a dynamical $SO(3)$-class of reference frames is totally absent: it is known that the theory does not exist in that limit and therefore our novel physical insight appears to be mandatory. Technically, albeit our theory is Lorentz invariant in a well defined sense[2], it is not Lorentz invariant in the operational sense of Weinberg since the operation of changing between reference frames has an impact on the "state" of the particle which is not quantified as usual. Therefore, Lorentz invariance in the ontological meaning of the word has different implications than Lorentz invariance in the *standard* operational sense as worked out by Weinberg[3].

We will start by expressing "standard" physical ideas in our language and one therefore anticipates upfront that they lead to exactly the same problems as quantum field theory on Minkowski. Thereby, the "only" virtue so far being that we uniquely generalized a troublesome framework in Minkowski to an

[2]We will go deeper into that matter in chapter six.

[3]The distinction resides in the inclusion of zero norm states or ghosts as the reader will discover in chapter six.

equally plagued theory in any curved spacetime. The reader must understand nevertheless that this is, by itself, a nontrivial thing to do and that all known standard approaches to the subject still did not reach the level of generality and maturity obtained in the discussion above. The results displayed in this book are, however, much more far reaching than that and we will cure the theory by looking for more general equations attached to the principle of Lorentz covariance defined above. Indeed, I specifically said that the simplest equation led to the standard plane waves which means that there exist other equations with non-trivial parameters leading to "approximately" the plane waves and a well defined theory. Actually, I will not only argue from the mathematical side but also give a physical analogy why such parameters should be nonzero. Hence, the failure of standard quantum field theory is explained here as being associated to ignoring very crucial physical effects which should simply be included in your theory. These effects are commensurable with our notion of Lorentz invariance and therefore do not endanger the theorist of being in conflict with recent cosmological observations. We introduce this new physics step by step in chapters six and seven.

So, from a historical point of view, our theory constitutes the very first example of a well defined framework which includes the gravitational force as well as the "quantum" laws for elementary particles. It is this reason why I have decided to reveal these results for a larger audience; that they may enjoy the new language as well as its inherent simplicity and generality. Now, of course, I do not claim this *is* the only point of view on these matters possible and the reader may enjoy a much broader perspective in chapters two and eleven. What I do think however, is that this theory is the simplest one possible and that its lessons must be present in any more complex scheme of "quantum gravity". In this sense do I believe that the novel physics required to cure old problems are not merely associated to a particular scheme but are lessons of a more universal nature which should show up in any future alternative candidate theory in one form or another.

The remainder of this introduction is devoted to the deeper underlying idea that formulating physical theories in a manifestly geometrical language is a superior strategy. This has by itself nothing to do with the fact that a dynamical theory for a curved geometry constitutes the suitable candidate for a theory of gravitation as Einstein dictated by means of his famous equivalence principle. The latter can only be true because gravitation acts on a test particle in a way which does not depend upon the mass of this particle. Einstein showed that under reasonable conditions the Newtonian laws for a particle in a gravitational field could be deduced from *free* motion in curved spacetime which is the reason why relativists often stress that gravitation is not a force but the geometry of spacetime. As I shall explain in chapter two on the general philosophy of physics, spacetime is the theater in which elementary particles reside and humans live: in order to be able to define laws for elementary particles or humans, we need relationships between the events of spacetime. Physical definitions of elementary constituents should be made by using these relationships only since they constitute the ground rules of the house defining the irreducible characteristics of its inhabitants. The reason why geometry should constitute at least a part of these relationships is to delineate a class of events which are locally to the past of some given event, meaning that the latter notion imposes a physical condition on processes of elementary particles. This physical condition states that only quantities associated to processes which do not travel into the past, in some sense, should be included. This statement is not intended to mean that there do not exist processes going backwards in time, just that to such process one associates a quantity corresponding to a process going forwards in time. This is a dynamical condition of "internal temporality" which should be satisfied by any physical theory albeit it constitutes a weaker statement than the assertion that processes never travel into the past in spacetime. Hence, some notion of pregeometry, defining a local partial order, should be the constituent of any physical theory: spacetime may be endowed with more structure than that but the latter is the theoretical minimum.

Hence, every decent formulation of physical laws is a relational one and that is precisely the picture we advocated above regarding the "quantum" laws for elementary particles. Here, elementary particles were defined in a holistic way with respect to a reference event which we associated with the place of creation. Einstein's theory of general relativity, which constrains the geometry by means of the mass distribution in the universe, constitutes the simplest example of a relational theory in the sense that it satisfies the stringent property of locality, something which should be abandoned in any quantum theory of the universe. It, moreover, is a theory which imposes conservation laws for matter *and* geometry separately whereas one would generically suspect the real world not to satisfy any conservation law and certainly not two of them. In a sense, it is a philosophically challenging statement to claim that these conservation laws are the result of a dynamical symmetry the laws have to obey but as we shall argue in this book in chapters two and three, this symmetry group is way too large and the consequence of a philosophy which prefixes certain qualities of our universe. This is certainly against Einstein's own philosophy which constituted of the desire that a physical theory should depend upon a minimal number of kinematical assumptions. Indeed, such theories are the most generic ones and it is the fate of humanity to drop such "ad-hoc" assumptions when going from one theory to another: a story which is never-ever finished. One of the assumptions which we shall dispose of in this book is the very idea of an a-priori kinematical symmetry which is elevated to a dynamical symmetry leading to conservation laws. This is connected to a "Hamiltonian" formulation of physics and indeed, those ideas are referred to history and permanently disposed of in this book. The technically trained reader will then conclude that we will dispose of local partial differential equations too as *the* language in which physical laws have to be formulated and this is indeed the case. We shall, instead appeal to the much richer framework of geometrical processes to formulate physical laws; a philosophy which allows one to obtain a true spacetime formulation of physical laws given that the initial value point of view, which required a split between space and time, is improved upon. For a relativist, this sounds a natural thing to do and the reader will notice that a lot of the quantum magic, associated to the $3 + 1$ point of view, dissapears in this four dimensional formulation of affairs.

The enormous advantage of a book, which reflects your own research, is that you are acutely aware of further generalizations of your line of thought. Whereas the world used to be in a tight jacket and you partially released it from that constraint, doing so opens a new world of possibilities *never* envisioned before. I will profit from this occasion and fill the book with remarks, preceded by the header "Extension", of that nature; this should enable the reader to *understand* better why we currently impose the constraints on the theory we do. So, this book is written in a way as to make it as organic as possible, a quality I admire in the conducting of Wilhelm Furtwangler and which I aspire to follow. The 16 year old reading this book has certainly scratched in his hair a couple of times already and must have wondered "this is way above my head", certainly when it regards the technical aspects of the matter. I would advice to keep calm and proceed; things you never heard about, let along never thought about, always come as a shock when you meet them for the first time. But later on, it becomes fun, you start to understand where they come from and you engage in naive ideas of how to make improvements; chapter two of this book tries to fill in the philosophical gap. It introduces you to a language which we will develop to even speak in appropriate terms about the world you live in; if you have never thought about it, read it, try to understand why this primitive language is most likely mandatory and if you are really ambitious, look for improvements or generalizations you can think of. Also I talk about some things which I do not reflect upon since they appear natural or evident to me, in case you would see potential troubles with some way of phrasing things the way I do; write down your own improvements and motivate the fine details. I must warn you that I speak already in a way which is far more general than most physicists do, so you will have a hard but enjoyable time.

After chapter two, I will try to hang the 18 year old on a mathematical infuse, pouring ideas and results directly into ones brain. I will do this in a playful way but, nevertheless, not shy away from giving you the exact definitions. These are important to learn to accurately speak about the things you want to say: for the 16 year old, it is more important to understand the idea behind every definition rather than being able to work directly with it in a logically consistent way. You may try to do that when reading this book for a second time, but I formally promise that I will write chapter three in such a way that anybody reaching that level of intuitive understanding can associate the correct meaning to every mathematical statement I make in this book. To be able to verify these things by yourself requires further skills which one obtains after some formal training or playing with them on a sheet of paper in the solitude of a quiet chamber for a sufficient amount of time. This is as ambitious as I can reasonably be: one book cannot constitute the substitute for an entire five year training at a reasonable level but it may just give you the most important ideas and concepts you are going to learn about in those five years. Being someone who cherishes some aspects of the modern generation, I advice anyone who wants to dig deeper in some of the mathematical material to first consult the internet prior to buying a good book about it. The huge advantage of Wikipedia is that it provides you with the crucial insights to understand what is going on while a textbook, all too easily, dwells on details which are not that important. The 16 year old should use this resource when feeling that he or she has not quite reached a sufficient level of understanding from my presentation, whereas some parts, but certainly not all of them, will be rather standard material for the trained physicist. He or she may decide to skip chapter three and consult it only in case it is really needed, I have included chapter three for the general reader who is no so mathematically trained as a scientist is.

We start to work towards the real theory in chapter four which every reader should carefully digest; albeit I have gone through some effort in this introduction to argue for the geometrization of physical laws, the reader should understand what kind of things can happen when you do not do that. The non-covariance of standard quantum mechanics is argued from some different points of view which basically imply that the theory becomes unphysical and leads to all kinds of strange effects, some of which have become standard lore in the physics community. Well, if you sleep with a dog, then strange things happen and general covariance is precisely the kind of recipe which turns the dog into a beautiful woman, if you are a man at least. Chapter five is like a first taste of something new: the idea of how to geometrically construct free waves, satisfying a positive energy condition, is explored and words such as propagation and probability of detection are carefully introduced. It also explains, to some extend, the gap with the standard view and how problems in the latter are solved within our philosophy. The reader should really embrace this relatively short chapter with care as it constitutes the basis for everything to follow. Chapter six expands upon the ideas of chapter five and contains an entirely new theory of spin having some subtle, but not unimportant, differences with the standard lore: since our viewpoint is much more geometrical, we deem it to be the correct one. The results obtained in chapter six are truly wonderful in the sense that it normally takes an entire book on quantum theory and a few chapters on quantum field theory to reach an understanding on these matters which is inferior to the one obtained in those twenty pages. To illustrate the progress made in this totality of thirty pages, a standard quantum field theory textbook only treats relativistic quantum theory in Minkowski, it has no clear interpretation and is by no means rigorously defined. On the contrary, it is not clear what the theory really *is* and it fills typically around five hundred pages of unmotivated, unrigorous calculations to obtain some "results" of interest. This brings us to chapter seven where we carefully examine the implications of the gems of our theory: we introduce some typical integrals corresponding to elementary scattering processes and show that, as happens in the standard theory, (a) the propagator is not well defined in the Lebesgue sense and exists only in some very weak dual sense and (b) that the latter does not suffice to *define* the general integrations of interest. Standard renormalization theory is then some ad-hoc game to provide for a "physically motivated" way to *define* these integrals by substracting a finite part of them and to absorb the *infinite* remainder in some free parameters of the theory. Also, it leads to the wrong physical insight that a graviton theory cannot be meaningfully defined and it is completely uncertain how the "theory" should be extended to include gravitational effects. We will dispose of such "philosophy" here and proceed in a proper way by recognizing that our viewpoint, unlike that of standard QFT, allows for the introduction of extra terms in the defining equation of a free particle so that the propagator exists in the Lebesgue sense and that all interaction processes are well defined. The way of how to achieve such result is first introduced in chapter seven and, at that point, we have definetly moved beyond all existing programmes. There remain now three things to do: (a) define the general interacting theory, a task which is done in chapter eight, where we introduce the "usual" Dyson series as a summation over geometric processes (b) show that every geometric process is finite, a task accomplished in chapter ten where we start by obtaining bounds on the particle propagators and finally (c) investigate whether the series is finite for a certain parameter range. This last task is assumed for the standard, known, theories in chapter eleven where we will gather some interesting results. Hence, the main theory of this book spills around 110 pages of ink, it is rigorously defined and proven to be sound and works in a "generic" gravitational field. Moreover, the issues regarding the role of the observer have been fully cleared out and very little mysterious elements in the theory remain albeit it could be more complete. These facts should suffice for any reader to carry on: I wish him or her a lot of fun and courage in trying to digest my effort to bring one at the forefront of physics in as little as 250 pages.

Obviously, this theory is a work in progress albeit it constitutes a considerable step forwards regarding the programme of unification. The way in which it is incomplete should be clear after the reader has carefully digested the next chapter; indeed, the measurement axiom is somewhat put in a more profound way as it is in standard quantum theory but not all issues regarding measurement have been solved. He or she will get a sense regarding the difficulty of such project when going through chapters two, five and six. There remain plenty of other interesting topics to be investigated regarding global properties of the background space times and such issues are the subject of future research work.

Chapter 2

General philosophy of particle physics.

This chapter is a shameless, literal copy of a section I have written in a long philosophical paper about the foundations of physics. I did not modify any word in it given that it took me years to be able to present its content in the way I did; all symbols are thoroughly introduced and motivated and should give the reader an idea about the general enterprise we are going to embark upon in this book. Central to the entire enterprise of physics is the concept of spacetime and to really have an insight into the assumptions which creep in regarding the structure of spacetime, we must focus on our senses. Our mind tells us that each individual undergoes a *process* of renewal meaning we have a relationship of the form

$$A \to B$$

where we call A the initial state and B the final state, the arrow denotes the "process". One fundamental issue here is that if A and B determine the state of the entire "universe" then one has that B always has to be different from A otherwise we would not speak about a process, that is if there is no change in the state of the *total* system, there is nothing that happens (note that we do not speak about "time" here which I intend to *ban* from all discussions). This already delineates a very strong distinction between so called open and closed systems where I will call a system closed if observer and observed belong to it and open otherwise; this is a very rough definition indeed and one might opt for calling a system closed if and only if all physical questions asked come from within the system and not from outside. For example, a closed box (thermically isolated or not) can be seen as a closed system as long as one does not ask questions like "if I open the box, do I find a cat in it?". If the box is not isolated from the rest of the universe, a creature within the box will resort to a theory in which no conserved energy and momentum exist, probably he or she will expect there is something beyond the universe due to the different nature of the boundary region. However, when asked from the outside of the box, speaking as God, "where do you think this voice comes from?", the cat will probably answer that she came from the ferocious boundary region. When finally revealed the truth by opening the box, the cat will recognize God as another living creature and destroy all her boundary theories. The same trick can be played with other animals such as ants making them believe that the world is two dimensional instead of three dimensional, and who knows, perhaps we are fooled in the same way. This situation already appears for singularities in general relativity, where one can glue together different universes to form a larger inextensible universe. For an open system, it is possible for

$$A \to A$$

as long as it is understood that the process denoted by $\to$ comes with a change in the state of the observer but this is usually ignored in the description of the theory. Now, this begs the question already, what do we, as observers *and* observed perceive? The process or the initial and final states, both or something else? One should distinguish between the processes we observe and those which happen: indeed, as mentioned already, we can observe a static open system but that process is *not* the one which is happening since *nothing* is static for a closed system. This philosophical stance alone reveals that in the language of quantum mechanics one would have to say that the discrete spectrum of the Hamiltonian operator for a closed system cannot be observed, whatever time evolution would mean in such context. So, we have to rephrase our question as

$$A_o \to_o B_o$$

where the subscript stands for the *derived* notion of states and processes the observer holds. A quantum physicist would say that we measure *properties* of a state, but then, shouldn't properties not evolve too or is it something eternal? In the Schrodinger picture of quantum mechanics, one would indeed say that properties are eternal, but we will argue for a very different view on these matters later on. So, with my states A_o, B_o I really mean "state *and* operators" when referring to quantum mechanics; they are what evolves[1] according to $\to_o$ which is something entirely different from $\to$. It has been stressed several times by the founding fathers of quantum mechanics that the classical world is needed for the very existence of the quantum world and we shall argue that *therefore* the picture $A \to B$ is much richer than the picture $A_o \to_o B_o$ where the observer evolutions may be the set theoretical union of the observer's classical and quantum world. So a fundamental quantum universe does not exist and neither does an intrinsically classical one, as we shall posit in the last section of this work every object composed out of a sufficient number of elementary particles *is* classical as well as quantum. As a general comment, notice that the distinction between a state and process only arises when one writes down finite sequences, for infinite sequences the role of process and state might be

[1]So, in the traditional viewpoint, I would say that a state A_o equals $(\Psi_o, \mathcal{L}(\mathcal{H})_o)$ where $\mathcal{L}(\mathcal{H})_o$ is some star algebra of operators acting on Hilbert space $\mathcal{H}_o$.

interchanged and it is all a matter where one begins to write down the first entity. One notices that observation itself is a process, and one may ask oneself *what* one is observing : (a) what we observe are processes and we infer states from processes, this is identical to the idea that only work (the process) can be measured and that energy is the state which is arbitrary up to a constant (b) what we observe are (properties of) states now and we infer processes from measurements at different times (c) both. Note that all take (a personal) "now" as a fundamental notion and assume another process, the one the observer is consciously *undergoing*, as primordial although a theory from the observer's perspective cannot qualify, but nevertheless quantify this process, it is just called the variable "time" in non general relativistic theories. In general relativity, Einstein extrapolated, as did all scientists which came before him, the personal notion of time as a quantifier to (mathematically) integrate the processes, to a meta time which exists and flows. Note that the notion of process is even more fundamental than the notion of time and that one can always rescale the time duration of a process. What the continuum hypothesis of Einstein's meta time says is that no process is irreducible, meaning it can always be split up as the composition of two or more processes; the discreteness assumption on the other hand goes to the other extreme and says that, very much like the natural numbers, every process can be written as a finite composite of irreducible processes with that distinction that this decomposition does not need to be unique. There is of course the more generic attitude that irreducible processes do exist but not every process can be (finitely or infinitely) decomposed into them; nevertheless, as mentioned already, this situation does not need a quantifier for its description and therefore we think time is not a sane metaphysical concept to rely upon but process and state are. For sake of convention, nevertheless, I will continue to use the word spacetime but as the reader will discover it gets a somewhat different meaning than envisioned before. Before we proceed, we must make a perhaps unnecessary distinction between thought processes and actual processes in the sense that for the former, nothing is happening but for the latter something is; this evoques spiritual discussions such as, does there exist a mind irrespective of the body? These thought processes reveal a *reason* for a particular transition $A \to B$ to take place; one could argue that these thought processes do exist but from the point of view of our universe, they are pointless and we like to capture them into physical principles (note that there has to exist an infinite hierarchy of thought processes and universes in this line of thought). This is how one *has* to close that discussion if one wants to get anywhere in reasoning regarding nature. Other processes are *expected* processes, one might again put forward that these do not really exist but are a synthesis of real processes; expected processes are processes anticipated in the state of the universe which does not only reveal its being but also its potential becomings as well as their associated potentialities, prospects of the future if one likes. Mathematically, this would boil down to the fact that the *state* of the universe A contains symbols such as

$$B \to C,\ \lambda_{B \to C}$$

where in quantum mechanics $\lambda_{B \to C}$ is a complex number. There is something to say for this viewpoint if you want to make a distinction between *expected* potentialities and *realized* potentialities of *processes*. I believe, we humans do make that distinction in the way we move in a very simple way, we estimate the possible outcomes of a future process and their probabilities and make an actual move (with a certain probability) based upon this estimate and this move depends upon how far we can calculate ahead, such as is the case in a chess game. Again, as said before, one might argue that these calculations are the result of more elementary processes in our brain and that it takes a while to make that specific move; physicists in any case do seem to prefer this explanation for the very elementary reason, I think, that it is the simplest one. This is a sane attitude as it would be hard, at this point, to falsify this simple premise; on the other hand physicists have also ignored the direct influence of processes which have happened in one's person's past, not just the remote past, as playing a role in his actions, this is so called physics where the dynamical law has no memory. As this does not imply at all that there does not exist such thing like memory, also this attitude is not stupid and motivated by the principle of simplicity.

This discussion already reveals that instances of perception of processes is not the same concept as instances between processes, nevertheless we *might* call such instances "now" and (eigen)time the lapse between succession of instances. Coming back to the question what we observe, I think it is pretty clear which philosophy is the correct one and that is (c), we observe both processes and properties of states: for example, one can observe oneself breathing but one never knows precisely how far ones chest is extended at that instant but nevertheless one measures that ones chest is extended. In physics so far, we speak about a *generator* of processes, something which has to do with determinism and the assumption of an a-priori existence of spacetime and other eternal structures. Quantum physics is so far our only theory where the generator has a different mathematical prescription than the concept state which is good since both have philosophically nothing to do with one and another. So fine, we haven't said anything about space and time yet, nor about dynamics, neither did we clarify the word state and process, we just wrote down a diagram which clarifies the notion of a process related to the word state. Before we enter into the discussion of spacetime let us show by a very careful, but simple, reasoning how a generalization of the kinematics of classical and quantum mechanics comes around - this example is preliminary and ignores certain important points but it will do for now. Suppose that *space* (and not spacetime, we will explain later on why) is related to distinct generating properties (also called atoms in the lattice of propositions) $x, y, z \ldots$ all of which might have been created from nothing by a sequence of processes, then one can make (not logical, as we will see later on) *propositions* about these properties and the properties themselves have to be regarded as potentialities for making propositions. For example, the proposition $\star(x, \lambda)$, where $\star$ indicates the fact that we are making a proposition, would mean "the potentiality that a particle is found with property x is λ"; from now on, we will drop the star and λ and simply denote a one property proposition by the property x itself. This may be read as "a particle exists with property x" but one should always keep in mind that the potentiality matters too; we leave it completely open as how this potentiality is quantified. How space is related to atomistic properties depends upon one's point of view and we will come back to it later on but one may propose for now that space is "common" to all atomistic properties without really knowing what it means. Generating means that every property of a *single* particle is constructed from them by means of the operations $\wedge$ and $\vee$ *and* a quantifier potentiality λ associated to such word while distinct indicates that every combination (word, potentiality) is distinct (meaning there is no relationship between different words and potentialities). From the definition, it follows they are not mutually exclusive in having, meaning that one particle can have the property of two distinct atoms but in quantum mechanics some of them are mutually exclusive in the sense of measuring, meaning that one particle cannot be measured to have more of some specific properties (for example, a particle cannot be measured in two different positions). Before we proceed, we must decide whether the above sentence is a meaningful one; that is, wouldn't it be better to say *which* particle has property x? We do for sure theorize about one and the same particle having those and those properties and in particle experiments, it is *necessary* to assume that properties of a single particle propagate and under certain conditions, we pertain that "reason" obliges us to acknowledge that we observe the same particle once again. In the macroword, this happens all the time; for example if I were to commit a crime and later the police captures a person looking exactly like me using the record of the criminal act, he can sustain as much as he wants to that he did not do it but nevertheless he would get convicted. Supposing that we are identical in all other aspects too, that is our voice sounds the same, we both live on the streets and are beggars and so on, it would be impossible for the judge to separate us in case they would find me too; in a super advanced society, they might read your hard disk (memory stored in brain) and in this way they finally could judge me! This somewhat funny example reveals the following elementary facts: being identical of me and my twin drifter is in the eyes of a third party, I *know* I am me and not him and likewise does he, second the matter of being identical or not might be a matter of perception of the beholder and if he were to improve his or her perception he or she might discover the truth after all. We cannot ask to elementary particles or even cats where they come from, in the case of cats we have the possibility to trace back their steps based upon characteristics of their smell, skin and so on, but we have no such chance with elementary "particles". Nevertheless, *nature* might care which properties a particle had before it decides about its future properties; since *we* cannot decide about the "history" of a single particle, it is hard (but not impossible) for us to develop a dynamics where this should matter; in the theory of quantum mechanics, this limitation is elevated to a virtue since it is declared by fiat that nature just works like that - end of discussion. Heisenberg was a proponent of this principle which he revealed by the cryptic sentence that our theories of nature should not contain elements which *we* cannot measure, where the act of measurement refers to the object under study. Obviously, I cannot measure anybodies identity but I know

I have one; in psychiatry one can even speak about appearant "multiple" identities when a person undergoing a psychosis hears voices in his head and wonders if they are really him or not. That is, there one could speak about the reflexive relationship that one hears oneself thinking and depending upon the sensation this comes with, one imagines that one hears oneself speaking or someone else. Of course, the predominant opinion is that these voices are just processes in the brain fooling us, but nevertheless it is necessary to speak about *at least* one I outside the brain, otherwise this sentence wouldn't make any sense. This I doesn't observe, but is the awareness of the observation; for example I could say "I observe the computer screen" and imagine what this sentence really means. Well, they would explain me that photons coming from the computer screen are hitting my eye, which translates this in an electric signal which is transmitted to my brain, there it is decoded into an image. This may all very well be so, but logically speaking I should still add that I see the image in my brain. This I cannot be a property one could measure but it is ultimately that what observes. This is the *least* one could say, I have met religious types who would claim that our "intelligence" is not measurable either, by this of course they do not mean that one cannot perform IQ tests, for what that may serve, but that it does not correspond to any physical process. That is, the "I" should be supplemented with our thought process which are stored on a spiritual "hard disk", in other words they are processes happing outside the framework of spacetime (note that an identity falls outside spacetime too); as I have said before, these are processes for which nothing happens, since happening is tied to spacetime and visa versa. Note that I don't say what spacetime is, nor what a happening is, all I say is that one is tied to another; as I have mentioned before, science *has* to close that door and elevate it into physical principles. Coming back to our discussion, the I exists (even quantum physicists don't deny that) but they claim it to be irrelevant for physical processes since we cannot measure it; what I will do, is to turn the role around and *define* measurement from certain changes concerning the properties of "I". Indeed, quantum physicists are confronted with a supplementary definition of measurement process which they take for fundamental; the measurement process distinguishes itself from the "evolution" process in certain characteristics but in my opinion, they are just two extremes of something much more reasonable. Let me elaborate upon why this would be necessary in a closed system: in quantum mechanics of a single particle, the measurement apparatus is "symbolized" by an operator, so that what measures gets a fundamentally different status but also falls out of the quantum system. If one would take quantum physics of the universe seriously, then one arrives at the contradiction that the operator cannot be applied since the observer also belongs to the system. This has lead to speculations about a universal consciousness making the observation, but that doesn't really make much sense and I believe most of the founding fathers of quantum mechanics would have repelled that attitude. What I will do instead, is to restore the identity matter, leaving it entirely open if "most of time" the dynamics doesn't care about this issue, but making it of primordial importance when it boils down to defining the measurement process.

Now, it happens all too often that one symbol gets different meanings and we shall disentangle those from the beginning by using another symbol: for example, the couple (a, b) can stand for the ordered pair (a, b) where a has meaning with respect to A and b with respect to B and as such a and b have nothing to do with one and another, or the sentence a is in relation to b which we shall denote by aRb. Likewise, we have the distinction of "union" and "joining" where the union of two identities just means that they happily live next to one and another (they still can send messages, but they need a third party for this) while two identities join if they *behave as one identity*, that is, their description is *generically* larger than the union of the individual descriptions. Let me say here that words are just words and one could say, as we will do in just a few moments, that a union is a trivial join is we decide that the dynamics should leave the properties of a single particle rather trivial. This relates to the eternal interplay between kinematics and dynamics where *both* decide what is possible: the kinematics paves the ground for what is *potentially* possible while it is the dynamics which is the ultimate arbiter of what possibilities *happen* and which don't. Obviously, if a possibility does not happen, then one might wish to make ones structure tighter and eliminate the potentiality theirof. So when we say that two particles behave as a single particle, we intend to mean this in the *broadest* sense possible (which is broader than quantum mechanics indeed), so from our viewpoint, the distinction between classical and quantum physics might *not* be one of kinematics *and* dynamics but one of dynamics only. Let me illustrate this point of view which I realized only when I wanted to say things very accurately: in a classical first order system (we will see how second order systems can be treated later on) for a particle given by an equation of the form

$$\dot{x} = f(x)$$

one could describe the kinematics by x and λ where λ indicates the potentiality for x to happen. In classical physics, this potentiality does not change over time and remains 1. That is why classical physicists do not even speak about it, because it is trivial; in quantum physics, we have a so called wavefunction

$$\Psi(x)$$

which is equivalent to writing

$$\wedge_{x \text{ realized in space}} x$$

and λ where λ attaches a potentiality (complex number) $\Psi(x) \neq 0$ to each property x in the conjunction of properties a single particle possesses. So, if we take the proposition of properties and their potentialities as fundamental: it is the dynamics that decides whether the proposition contains just one property and likewise whether λ remains trivial or not. Now comes the catch one must understand very carefully: *generically* the dynamics of quantum theory will prefer to include as many properties in the conjunction as possible while the classical theory remains at one single property. So the quantum theory is more "generic" from this point of view but it is not at all for sure that the quantum dynamics is more likely to happen as the classical dynamics; it might be something in between or even something far more general than quantum dynamics. What physicists unconsciously do when dealing with this matter is to call an electron a quantum particle and a closet a classical object so in our language one might interpret this as additional properties of an entity. Now, why would a fundamental particle be exclusively quantum and not classical also or something in between? In physics, David Bohm and Louis de Broglie foresaw this possibility and attributed as well classical as quantum properties to the electron: so this is a new idea, that an unmeasurable property of a single particle might limit the kinematics and dynamics, we will illustrate that point of view later on too. So the reader must appreciate that what I am going to say below can be said in many different ways, from different angles and one needs supplementary philosophical prejudices to eliminate one way of speaking about something or to elevate one above another. Since this chapter was about general principles of physics, I will refrain so from doing that as much as possible but on the other hand I must give a balanced account of what possibilities have been entertained so far in physics when dealing with these issues. In other words, when we say that two quantum particles join in a quantum way, it might be that the dynamics can make them behave as if we would take the union of them, such dynamics would keep both particles effectively separated from one and another and might be called classical. Differently, we might take the classical join of two quantum particles and therefore pave the ground for a mixed classical quantum dynamics where the classical part relates to the dynamics between them and the quantum part refers to the individual dynamics. Such situation can, as said before, also be described by the union of two quantum particles. However, if we take the union, *then* we forbid the dynamics of a join unless the union becomes a join of course. I will come back to this point later on, it is largely a matter of language but it is far from trivial to even just imagine writing a classical dynamics in Hilbert space, although this is very much possible as exemplified above, it is just so that Hilbert space is not the "natural" habitat for classical (first order) dynamics. So, all of what follows is largely a matter of language and one may shift between different points of view; the reader should realize this and keep in mind that one cannot speak about *the* kinematics and that the framework presented below is a flexible one.

As stated already, one could "join" in different ways, for example in a classical and quantum way (we will define precisely what it means later on); so far in physics, the joining is minimal: that is the properties of the join are composites of properties of the individuals[2], there is no room for fundamentally new elements to arise from the joining. One could argue against this idea by relying

[2]Mathematically, this expresses itself in quantum mechanics by the idea that the two particle Hilbert space is the tensor product of one particle Hilbert spaces.

upon an old argument of Leinaas and Myrheim. So what mathematical operations could correspond to those ideas of union and (minimal) joining? The proposition of the union of particle one with property x and particle two with property y is denoted by $x_1 \cup y_2$ while the joining of particle one with property x and particle two with property y is denoted by $x_1 \otimes y_2$ (or we might equivalently have used the logical symbol $x_1 \wedge y_2$). Actually, we meet here already a very point of reflection; when we say that 1 and 2 have nothing with one and another to do, it is fine that we use $\cup$ as a symbol since it is symmetric, that is $x_1 \cup y_2 = y_2 \cup x_1$; we alternatively could also have written $\{x_1, y_2\}$ but when we "join" 1 and 2 it is maybe of importance in which order we join them (that would be new in physics) or the "join" depends upon the properties of the particles; indeed, so far we have not said that x would just correspond to a point in space, it might include other properties which do not "commute" meaning that one cannot just interchange them between particles (so it might matter for the "join" that particle one has property x and particle two has property y and not the other way around, depending upon the properties x and y). Therefore, we will use the symbol $x_1 \otimes_\alpha y_2$ where the α reminds us that the joining can happen in many different ways. This is an important fact as the questions we can ask about the "join" or "marriage" depend on the way it has been constructed. We are not there yet since so far, the only reason to make a distinction between a join and union was that the join might depend upon particle properties and the union not. Since our original definition was that a join behaves as a single element, we must first specify what operations can be done on single elements; since we identified $(\star)x_1$ with the proposition "particle one has property x" we simply have to take over all well known operations one can do with (logical?) propositions, that is $\wedge, \vee, \neg$. For example $x \wedge y$ means that a single particle has properties x and y - we will define a single particle to be classical if none of these operations apply, stochastic-classical if only $\vee$ applies, quantum if $\wedge$ applies and stochastic-quantum if both $\vee, \wedge$ apply. For example, in classical mechanics, a single particle cannot have two positions which in quantum mechanics they can have; also, in quantum mechanics, a particle can have spin $\frac{1}{2}$ and spin 1, but these properties can only exclusively be measured which we said before. Let me stress so far that by distinct properties, I really mean properties which have nothing do with one and another - *appearantly*, this may differ from theory to theory, in classical mechanics the position and momentum of a particle are distinct properties while in quantum mechanics they are not; I do not need to comment so far on the specific mathematical implementation of this concept, sufficient to know is that these properties can be specified exactly (for example they can correspond to a real number or a word). I will argue against this viewpoint that the momentum is a property of a particle, something which is grounded in the philosophy of continuous time and "eternalism"[3], I will present a different interpretation later on. So, for a single particle, one could suggest that $(x \wedge y \wedge z) \vee (w \wedge v) \vee (\neg z)$ has to be interpreted in the usual way (and corre-

[3]By eternalism, I mean that space and time exist a-priori, are uniform in (global and local) structure and are of vital importance in defining the dynamical laws.

sponds in quantum mechanics to a "union" of states, and can be interpreted in the same way as a density matrix, as we will see later on) albeit the $\neg$ operation is *never* used in classical nor quantum mechanics. There, one only states what *is*, or the property one possesses and not, what *not is* (or what one does not possess) since the absence of being (or possessing) of something automatically follows if it does not belong to the list of what is (or what one possesses); while, on the other hand, the above notation refers to what is true and what is not true and the absence of a statement about the truth of something does not reveal that it is false. So there is a *real* distinction in declaring that "an entity has those properties" or by "the sentence that "this entity has those properties" is true"; the former is just an object "sentence" while the latter corresponds to a process

$$\text{sentence} \stackrel{\text{logical}}{\to} 1$$

where 1 stands for "true". The reader must notice that I did not talk about the word "implies" symbolized by $\Rightarrow$ as a logical operation since there is no a-priori *logic* in spacetime; it is pointless to say that one event "implies" the other event without saying something about the dynamics. So, the above operations are *not* the logical ones and the interpretation of the property $x \wedge y$ is that the single entity has both properties x and y and no other which is not the same as the proposition that particle one has properties x and y is true. Similarly, the property $w \wedge x \wedge y \ldots \wedge z$ means that a single entity has properties w, x, y, $\ldots$ and z. The $\vee$ operation used here, is the exclusive "or" where the property $x \vee y \ldots \vee z$ means that the entity has exactly one of the properties $x, y \ldots z$ and no other where the latter can be in general composed properties by $\wedge$. The de Morgan rules hold exactly in the same way as they do for the ordinary logical or/and operations.

Let us now come back to the definition of a join which was that two identities join if and only if they behave as one identity: this already means there exist at least (and indeed there exist more) four joins: a classical, stochastic-classical, quantum and stochastic-quantum one. According to the definition of a stochastic-quantum join, we can write down sentences like

$$((x_1 \otimes_\alpha x_2) \wedge (y_1 \otimes_\beta y_2) \wedge (p_1 \otimes_\gamma p_2)) \vee ((r_1 \otimes_\delta r_2) \wedge (s_1 \otimes_\kappa s_2)) \vee (t_1 \otimes_\lambda t_2)$$

where x_i stands for property x of particle i and these operations are already generalizations of what in quantum mechanics is called the two particle Schrodinger theory albeit the latter needs an extra, trivial ingredient. We are far from being done and as the reader will appreciate, the current status of theoretical physics is just at the second stage in a sequence of infinite stages one can write down in this way. Before we proceed, we will slightly change our notation for the better (as the reader will see):

$$(x_1, y_2)_{\otimes_\alpha} \equiv x_1 \otimes_\alpha y_2$$

this will allow us to speak about the "join" of n particles with properties x_i as

$$(x_{1\star}, x_{2\star}, \ldots, x_{n\star})_{\otimes_\alpha}$$

where $x_{j\star}$ stands for property x_j of particle j. In the physics literature, this tensor product is "deduced" from the lower tensor products by means of a strictly quantum mechanical argument (the cluster decomposition principle); as we will argue later on, the cluster decomposition principle is irrelevant in our more general setting and therefore such reduction should not take place. The catch is of course that one must be able to speak about the union of quantum mechanical systems, or in our language, the union of sentences in "joined" particles. This constitutes a part of the question of how our operations should be extended on composite objects. One usually regards it as a virtue of relativistic quantum field theory to recognize the following simple fact, which is here merely a question of completion of operators, which is that $\wedge, \vee$ extend between multi particle joins; that is, one can write down things like

$$((x_{1\star}, x_{2\star}, \ldots, x_{n_1\star})_{\otimes_{\alpha_1}} \wedge (y_{1\star}, y_{2\star}, \ldots, y_{n_2\star})_{\otimes_{\alpha_2}}) \vee (z_{1\star}, z_{2\star}, \ldots, z_{n_3\star})_{\otimes_{\alpha_3}}$$

which prepares the setting for an extension of quantum field theory. Here, some caution is necessary, until so far we have *assumed* that if we write down things like

$$(x_{1\star}, x_{2\star}, \ldots, x_{n\star})_{\otimes_\alpha} \wedge (y_{1\star}, y_{2\star}, \ldots, y_{n\star})_{\otimes_\alpha}$$

that both indices $j : 1 \ldots n$ in the "superposition" referred to the *same* particle or entity, but how to interpret this more general situation? As a fact, for an infinite number of particles, how do we know that the indices j in the above refer to the same particles? The answer given in quantum field theory to this problem is that this matter of identity is *irrelevant* regarding proper physical questions one can ask; in other words, we do not need to answer this identification problem in order to extract physical predictions like I measure a particle with properties x or I measure two particles with property y and z. Hence, in this interpretation, one *excludes* observables which do not measure properties of a definite number of particles since it would be unclear how to interpret this. So, basically, one measures *properties*, the act of measurement can only be performed thanks to existence of particles or identities, but one declares oneself ignorant about which particle it is one measures a particular property of. The fact that in the macroworld, we can measure properties of specific identities (me or you) happens then because we are not in superposition and the problem is not asked. The reason why one can do this is because one speaks about a one particle Hilbert space which is the same for all particles, the implication of this is well known and that is that every particle can have an infinite extend or has access to the entire universe meaning one can form conjunctions and disjunctions of properties in an *unlimited* way; it is this principle which we will criticize later on. To jump a bit ahead, in quantum field theory, one speaks about creation and annihilation operators corresponding to a particle with specific properties, but one never mentions about which particle it goes; this gives a two particle state $x \otimes_f y$, where f stands for "fermionic", the meaning of the join of a particle with property x and a particle with property y, both properties which contain the word fermionic, and this "product" does depend upon the order in which one writes the *properties* x and y. That is,

$$x \otimes_f y = -y \otimes_f x$$

and since we *do* speak in terms of which particle has which property, we will refine the notation by stating that

$$x_1 \otimes_f y_2 = -y_1 \otimes_f x_2$$

where x_k has the usual interpretation. We will later explain the meaning of this minus sign in the context of quantum mechanics; So, I am going to properly restore the identity question and leave it as a matter of *dynamics* to determine whether one should make some identifications in a probability interpretation or not. So, from now on, we shall always denote

$$(x_{i_1\star}, x_{i_2\star}, \ldots, x_{i_n\star})_{\otimes_\alpha}$$

where $i_j \neq i_k$ for $j \neq k$ and $i_j \in \mathbb{N}$. There is no a-priori philosophical, nor physical reason to be so easy going about the identity question in the micro domain. I admit it is a very strong principle to say that answers (probabilities) to dynamical questions do not depend upon it, but it is for sure no mandatory one since for example, there might be a principle concerning the number and size of joins a single particle might engage in, and we should be conscious about it when we decide to make either choice.

The reason why I have been so strict about the identity question is that one might in principle posit that for example 1 and 2 join and 2 and 3, but *not* 1 and 3, in quantum theory, this situation is impossible to describe as it does not matter (one would just say one has two particles) but it is clearly a very logical possibility and I shall develop it further now. As before, we cannot preclude that 1 is single also even if it joins with different parties so we make the liaison part of the dynamical content; moreover it should also be possible for 1 to remain separate from 2 and 3. How should we write such state down? We would write down for example

$$p_1 \cup (q_1 \otimes_\alpha r_2) \cup (t_2 \otimes_\beta u_3 \wedge v_2 \otimes_\gamma w_3)$$

where the correct interpretation is that one has a description separated from a description of the join of one and two and from the join of two and three. This does *not* reveal yet that one has *classical* properties and *quantum* properties from its joining with two for example; that is another question as we have addressed already. We have called a single particle state stochastic if and only if it is of form

$$x \vee y \vee z$$

that is $\wedge$ does not appear into it. Logically this reads, the single particle has property x *or* y *or* z. Now, one might be more liberal and assume the particle *must* always have a well defined event associated to it but it might posses the colours purple and green jointly. In that case x is a shorthand for (p, green) and we can write

$$((p, \text{green}) \wedge (p, \text{purple})) \vee (q, \text{red})$$

and so on, this state is neither classical nor quantum but stochastic-quantum. However, it posses classical features in the sense that it cannot have two properties of event; as we will discuss later on, a classical dynamics maps classical states to classical states and likewise does a quantum dynamics but the most general dynamics can map classical to quantum states, change liaisons by creating new joins and destroying existing ones and likewise so for separated entities. So, in our general example

$$p_1 \cup (q_1 \otimes_\alpha r_2) \cup (t_2 \otimes_\beta u_3 \wedge v_2 \otimes_\gamma w_3)$$

we say that 1 has a classical property p_1 and a joint *pure* property with 2 and 2 has an *impure* property with 3. One might at this point agree that this situation prototypes the most general one in the sense that the correct order of operations is given by $\cup, \vee, \wedge, \otimes_\alpha$ and that would make a lot of sense. It is completely reasonable for a single particle to have two disjoint descriptions as a single entity, for example a "classical" and a quantum mechanical one such as is the case for the Bohm-de Broglie approach; also, it is a-priori possible for a single particle to join twice with a second particle, as long as the "join" is different. The distinction between $p \vee q$ and $p \cup q$ is that in the first case only one conjunction is correct but nature has an *intrinsic* lack of knowledge about it (where usually this lack of knowledge is assigned to the limitation of a description by some observer), while in the second case both conjunctions are valid but *different* descriptions, meaning you cannot apply $\wedge$ nor $\vee$. Both arguments are familiar to people who know physics albeit they are more restricted there; for example, in quantum mechanics, we can have different descriptions of a single particle, but there those are assumed, by construction, to be equivalent (meaning the same up to a change of basis). Here, we extend this principle, by allowing them to be non-equivalent; in words we make a distinction between a conjunction of properties, which we might call a partial state and a conjunction of partial states. In the language of set theory, this is the distinction between $\{X, Y\}$ (the conjunction between partial states) and $\{X \cap Y\}$ (the conjunction of properties) where X, Y are subsets of some larger set; X or Y would then be given by $\{(X \cup Y) \setminus (X \cap Y)\}$ which is the disjoint union. One could try to imagine what something like

$$(1 \cup 2) \wedge (1 \otimes_\alpha 2)$$

would mean, literally one would say it has both descriptions as a system where 1 and 2 are disjoint and a description in which they are "joined". In my opinion, this would be equivalent to

$$1 \cup 2 \cup (1 \otimes_\alpha 2)$$

on the other hand

$$(1 \cup 2) \vee (1 \otimes_\alpha 2)$$

would mean that only one description is true but we don't know which one; it is here that I would launch a philosophical principle which is that of *definiteness of the description* which means that this "or" relation is forbidden. The reader must notice that so far, I have skipped sentences like

$$(p_1 \wedge q_1) \otimes_\alpha r_2$$

bringing the operations which are valid for a single particle under the multi particle join; actually, we have declared (by our agreement upon the order of operations) that such sentences should not be written. Of course, we should comment upon this and we will postpone this discussion to the future, but suffice it to say that in quantum mechanics it has something to do with

$$(p_1 \otimes_\alpha r_2) \wedge (q_1 \otimes_\alpha r_2).$$

As mentioned in the very beginning, one attaches "potentialities" to these sentences; in standard quantum mechanics this means that to every particle property we attach an *amplitude*, which is a quantifier from which the probability for these exclusive properties to arise can be measured. Note, that this definition is far more general than the quantum mechanical one as we do not even demand this amplitude to belong to a division ring. Therefore, our setting is much more flexible than the one of Jauch, Piron and Aerts; indeed, we did not make any restriction on the lattice of propositions one can make, the only nontrivial input is that of a union and a join and the definition of operations on properties of a single entity. Moreover, one can further widen our scope in a categorical sense and we shall just do that later on, this section was just meant as an appetizer and there is much more to say about it than we did so far.

2.1 On the definition of spacetime and causality.

Now that we have defined the words classical and quantum given disjoint generating properties which we associated to space, let us now come to a further specification of what we mean by "spacetime", before we can proceed to the notion of universe. Spacetime furnishes, in the most abstract sense, the ground for properties of elementary particles just like Minkowski geometry does for point like particles. To start with, everyone's description, which is different from one's experience, of nature is that of a sequence of processes

$$A^1_o \to_o A^2_o \to_o \ldots \to_o A^n_o$$

and the question now is how to "glue" these experiences and descriptions of open systems together in a theory of a closed system, the universe. This is what the philosophy of spacetime is about and what causes all the heated debates: experience has thought me that you need to weaken this gluing procedure as much as possible. Let me first give an idea of the scope of this question, its twists and turns as we know it in modern physics; as I have stated already, the theory of quantum mechanics is a theory of open systems meaning the observer does not belong to the system. Since in modern Quantum Field Theory particles are described *relative* to an observer, they have become an observer dependent concept, a philosophical stance which cannot be maintained for closed systems where the observer himself consists out of particles. So, the reader may guess that different observers might give different descriptions of what is going on: one observer might be an accelerating one while the other freely falling (I will explain these concepts in far greater detail later on). What the theorists now do is to *compare* these descriptions of the universe of both observers by assuming three *objective* properties (that is spacetime, the so called field algebra[4] and the field equations); the very structure of quantum mechanics allows for such comparison to be made once these objective properties are taken for granted. Now, what one arrives at is that while the free falling observer may describe the universe as empty, the accelerating one will describe it as being full of particles (we will spawn more detail out later). From the viewpoint of a closed system, this would mean that the process of accelerating will necessarily create particles if the universe were empty so that the accelerated observer effectively sees them and a remote, free falling, observer at some distance of the accelerated one might observe something like particles surrounding the accelerated observer. Now, it appears rather obvious that the specific spectrum of particles which get born together with the act of acceleration will depend on something more than just the magnitude of it; for example it might depend upon the mass and charge of the observer. This is the nonsensical aspect of the usual quantum mechanical calculation which has been called the Unruh effect, the spectrum of particles seen only depends upon the magnitude of acceleration of the observer; this is in some way logical as the very scope of describing open systems is limited to the very primitive character of the observer. Indeed, for a perfectly massless and chargeless observer, such as is the case in quantum field theory, physical intuition would tell you that no particles are born whatsoever because of acceleration. This has often led me to say that there does not exist an Unruh effect while at the same time admitting that particles may be found whose spectral properties would *probably* depend upon a lot more than just the magnitude of acceleration of the observer. Of course, this is all just a theoretical exercise and an Unruh effect has never been detected, nevertheless most researchers would claim it *implies* an observer dependent nature of the particle concept and as I have just illustrated that is just plain nonsense (and also erupts the question what this metaphysical observer really is). Before I will enter in the rather difficult general discussion of gluing our experiences together, let me elaborate first on how science has dealt with this question for the last couple of centuries starting out with Newton, Galilei and friends: they assumed (a) there is a fixed number of material objects constituting the universe, the processes $\to_o$ every "conscious" material object observes also *happens* at the very moment the observation takes place (here again one abuses language -and identifies a process with a now- because they assume the process of observation to be infinitely fast). Therefore, they completely tie the concept of happening to observation

[4]I will argue against an objective field algebra from the traditional viewpoint later on.

by demanding that there is a universal "now" associated to an infinitely fast process (observation) such that in reality there is an *objective* process happening when we observe something like it happening. This is Newton's principle of simultaneity of happening and observation, by our senses, of this event; hence the deduction by Newton and consorts of the existence of signals which "travel" infinitely fast. These are of course not all the assumptions Newton made, but it is one of the most important and basic ones. Second, Newton assumed that this "now" had the structure of a three dimensional Euclidean space $(\mathbb{R}^3, ds^2)$ where

$$ds^2 = dx^2 + dy^2 + dz^2$$

so he did not only assume that material bodies made up the universe, but the universe was more than that, a fictitious space (which could contain no matter at all) equipped with a fixed Euclidean line element which he interpreted as a fixed physical measure stick which was outside the realm of dynamics. Third, he assumed that every past and future now had precisely the same structure and that time is a continuous parameter running infinitely far to the past and future. So, Newton went far beyond the more elementary process view explained in the beginning of this chapter and described everything with respect to a meta time.

Einstein, Minkowski, Lorentz, Poincaré and friends offered a less rigid but similar interpretation, very much like Newton, he presumed eternal material bodies making up the universe and likewise did he give the universe a fictitious structure beyond that but some interpretations of relativity deny the existence of an absolute now of *being*, meaning that the theory contains an element of where I am when you read this work. By this, I do not intend to say that the theory must give a unique answer to this question, but at least that it gives some answer; Einstein's theory of relativity leaves this completely blank and proposes that the issue is not a physical but a metaphysical one. I am not a historian of science and I usually do not really care who thought what, but let me present the minimal assumptions which go into the theory of relativity and then present possible "supplementary" interpretations. I think that Einstein was impressed by the fact that Minkowski's geometry turned up in the theory of light and he wanted a theory of gravitation which was compatible with it. If one takes this point of view, then the causal *interpretation* of the conformal structure of the metric follows from additional assumptions regarding classical (field) equations of motion. It is however so that classically, you can turn this relationship around, that is start from a notion of causality and a volume element and deduce gravitational phenomena from the geodesic equation of motion. The latter says, as explained before, that free particles move on a straight line which correspond to the ellipses and circles of flat Euclidean geometry when one makes a *three* (and not two) dimensional projection. We will come back to the causal interpretation in a while. As said, general relativity (and perhaps Einstein too) sees the question more broad, than Newton did. One recognizes that the question of the happening of a process versus an image of it being "sensed" by an observer removed from it needs to be answered, but Einstein also thought that "signals" could travel at most at finite speed so that me measuring a signal coming from Venus means that Venus sent out this signal somewhere in *my past*; likewise he recognized that there were signals from places which had not received me yet so that they were "spacelike" to me. Moreover, in some bizarre twist, he designated the places in *spacetime* to which I can send (towards the future) a signal (in my personal) now. Einstein constructed his work furthermore on the insights of special relativity where it was the causal geometry of Minkowski which was being important, the latter being given by

$$(\mathbb{R}^4, c^2dt^2 - dx^2 - dy^2 - dz^2)$$

where c is the speed of light. To my mind, it is pretty clear that Einstein must have thought that the lightcone

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = 0$$

delineated the propagation of physical processes, light is to travel on the lightcone and all material bodies are to travel within, that is

$$c^2dt^2 - dx^2 - dy^2 - dz^2 \geq 0.$$

This is one facet of the causal interpretation: all signals propagate within the lightcone. More precisely, one has the following initial value point of view: suppose that all initial conditions for the universe are given on a three dimensional hypersurface Σ, that is the initial conditions for the metric (gravitational field) and matter within the universe, then the evolution of this data has the following dynamical property: the state of matter or the gravitational field in an event is fully determined by the initial data in the past of this event. This implies that if you would change the initial data outside the past of this event, there would be no influence whatsoever on the state of matter and the gravitational field in that event. This is a result valid for fields, for point particles one can show that if they initially travel within the lightcone, they will continue to do so; but on the other hand it is possible for them to travel outside the lightcone (and continue to do so), such particles are called tachyons. For a philosopher, the very existence of a gravitational field is already an assumption he cannot a-priori justify and it is natural to resort to a causal interpretation since this has something of a more substantial flavor. In other words, we need to find the *logical* origin of gravitation which is a question Einstein left far open. Now, there is a logical connection in theoretical physics between the concept of Minkowski geometry and some quantum mechanical assumptions on the one hand and the particle properties of mass and spin on the other. More specifically, the statement is the following: Minkowski spacetime has a symmetry group which we call the Poincaré group, quantum mechanics dictates that any symmetry should be elevated to an *operational* symmetry (so that invariance under it becomes an operational principle) and that one needs to study irreducible representations of the group. Those irreducible representations determine partially the distinct atomistic properties of a single particle; that is mass and spin are predicted but not say electric charge. Let me list the long chain of assumptions which goes in the deduction of these properties: (a) "eternalism" (b) continuum hypothesis (c) operational symmetry (I will spawn my comments on that later on) (d) irreducible linear representations determining distinct atomistic properties. There are other approaches towards this problem, one constructed several years ago by this author and another one relying on spinor bundles and the theory of fields. Both approaches elevate the above deduction to a more general setting and both have their problems. In classical general relativity, one could try to deduce the existence of a particle in terms of properties of the gravitational field, for example there exist solutions to the vacuum equation with so called singularities where the physical properties of the solution suggest that a particle should be present at the singularity or that the geometry represents a particle and its gravitational field even though we inserted no matter terms in the Einstein equations. This is for sure the case for the Schwarzschild and Kerr-Newman black hole solutions and such general programme might be called "matter from geometry" instead of "geometry equals matter" where the equality means that both influence one and another but are independently defined. It it unnecessary to say that here one still needs to define an identity from the gravitational field but the catch is that this identity is not going to influence the evolution of the gravitational field, the latter just evolves by itself (and the identity would need to be redefined during this evolution). It is just so that one can define an identity which one might reasonably suggest to be a source for the gravitational field but in reality it is not. This situation is less general of course than the one where one takes identities as fundamental as they are believed to have an influence on the notion of geometry. Let us now come back to the interpretation of causality in quantum mechanics, is it there also the case that particles propagate within the lightcone or that fields satisfy causal propagation properties? There, the answer is more subtle and is given by the statements that (a) the field operators for sure propagate in a causal way but (b) particles can "propagate" outside the lightcone and even into the *past* (another consequence of eternalism). More precisely, if I were to create a particle by acting on the vacuum here with the field operator, it *might* (with some probability) immediately be measured on Venus in some global reference frame. In that sense, signals do travel faster than light in quantum field theory, since I could replace the word particle by say Mozart's symphony. Nevertheless, the notion of causality presents a limitation on this kind of exotics by making the probability, for such detection to take place, exponentially small in terms of the spacelike separation; at least, this is so in Minkowski spacetime. Effectively, without reference to field operators at all, the notion of causality would mean *outcome independence*, a notion which is derived from a $3+1$ view on Einstein's four dimensional "block" spacetime and which will be explained in the next section. So, we are left with two distinct meanings on causality both of which may be right in their own domain of application; my suggestion is to dispell the notion of causality from foundational discussions and replace it by "evolution of an individual atomistic property" and maintain that it is the dynamics which decides what causality interpretation holds - that is outcome independence ought to be a constraint on the dynamics *resulting* from the notion of "evolution of an individual atomistic property".

Before we proceed with this more general discussion, let us further examine a bit the implications of Einstein's view on spacetime: for example, it implies that the future already exists. Moreover, it obscures the experience of the process of measurement between two "nows" by an individual observer given that there is no such thing as "happening" in the theory. Nothing comes to creation in the theory and everything is eternal unless you run into a singularity (which should not be seen as an internal contradiction in the theory as some prominent physicists stress, it really depends upon one's interpretation); as it turns out, according to our definition of a process as a change of state, modern physics arrived in the murky situation that with a traditional view on the notion of state within Einstein's theory, there is no process taking place whatsoever and therefore no observation. This is generally accepted to be a consequence of general covariance, a symmetry which is only possible because of the assumption of "eternalism" or in other words, due to the denial of some dynamical origin for this symmetry. It is a dead place for quantum mechanics and this is indeed the so called problem of time physics faces since 90 years. The intention of this section was to be general, so I am postponing more details on these problems of general relativity, instead I return to my original question and study if I can solve it into an even more general way than Einstein did and with a significantly different interpretation such as "evolution of individual properties" versus causality.

So, our discussion is still one of kinematics, meaning ways to phrase the basic ingredients to speak about a theory of nature, not about dynamics which dictates how these ingredients merge and evolve. To illustrate the great liberty at hand here, let me give a Newtonian way of describing Minkowski's geometry; suppose that *space* exists, which as we have said before is the collection of properties of elementary particles, and is given by $(\mathbb{R}^3, dx^2 + dy^2 + dz^2)$, a three dimensional metric space. Suppose that it just remains static during the process of evolution and introduce a time δt for a process of evolution to happen. Then, we would like to say which atom (point) evolved into which atoms (points) during this process of evolution; we declare by fiat that (x, y, z) evolves into all $(x + \delta x, y + \delta y, z + \delta z)$ for which

$$c^2(\delta t)^2 - (\delta x)^2 - (\delta y)^2 - (\delta z)^2 \geq 0.$$

Given that the geometry of space is fixed, one can now retrieve Minkowski's view by choosing an origin of time and taking $\delta t \to 0$. We are now confronted with the question, which point of view is the correct one? Minkowski's point of view which is "eternalist" and takes the Lorentz symmetry as fundamental, or this point of view with an absolute time and space but with emergent relations arising from reducible processes, with otherwise exactly the same physics? I think this question is again one of what one means by the observer: Minkowski would say the spacetime coordinate system is tied to a global observer while in my view one would need to add the observer as a physical entity to the system and then notice that its dynamical time (eigentime as measured by perfect clocks) does not need to coincide with the meta time t. Actually, I think that both interpretations are mistaken since in our view, why would the process of evolution of properties not depend upon previous established evolutions of properties and properties themselves ? On the other hand, one should just remark that the Lorentz symmetry is always a local symmetry and properties of particles should be determined by the local geometry and not the global one. For example, one might imagine that for some corner of the universe, the property spin is differently defined what would be the case if space were four dimensional instead of three dimensional over there or if the metric were four dimensional Euclidean instead of Lorentzian (such as is the case for the Hartle-Hawking wave function). Therefore, the virtues of Minkowski's viewpoint (the Lorentz symmetry) do not require his "eternalist" viewpoint. Hence, we arrive at the more general viewpoint that one needs to study processes from one universe, where one considers spacetime instead of space to a larger universe. Let us now try to make that precise, in our previous analysis, we stressed that properties of elementary particles were related to space and preferred this over saying that they were related to spacetime, the reason why I made that choice is because I had not defined space nor spacetime yet and it was more correct, given that we can only measure properties of space, to state space. So far, we have uttered the words properties which are necessary attributes for particles, evolution of properties, processes, space and spacetime: let me now try to find a proper language for dealing with those concepts and point out its limitations. First of all, why would we like to speak about evolution of properties? As we will see later on this concept is one of pregeometry, that is under *additional* assumptions Lorentzian geometry follows from it. Strictly speaking, one can write out a process just as evolution of conjunctions and disjunctions of properties of particles; but such dynamics would not be based upon any *principles* if we did not specify any relations between properties. Newton already recognized that and unified these relationships by means of the concept distance (metric) so that he could speak about properties which are far apart and close by. Now, I believe it is a general principle that nothing must be static, meaning that those relationships had to evolve too which is part of Einstein's view; but there it are the spacetime relationships which count and not the space relationships. These relationships are what I call evolution of properties, which are preserved and created by means of a process which is something even more radical in the sense that now, the properties get "reborn" also: it is the dynamics which must decide how far this birth process must go and their existence should not be a-priori determined such as is the case in Einstein's theory. This paves the ground for *dynamical* laws (that is the potentialities we attach to processes) which evolve too: away with "eternalism"!

Now, we just spoke about the fact that a universe with laws needs relationships between its properties and that those relationships might change and new properties get born by means of a process. As far as I know, there is no real justification whatsoever to limit those relations to the concept of evolution of properties which I will explain now. In a sense, the latter constitutes a generalization of what Einstein did, but excludes such possibilities as a Riemannian (part of the) universe such as exists in the Hartle-Hawking wavefunctions. As the reader may guess, this concept relates to the dynamical notions of finite signal propagation, as laid out by Minkowski, and the notion of outcome independence, but I will present it in a somewhat more general way than usual. We will make the agreement here that all our space times contain the relation "evolution of properties" although I warn the reader again that there is no good philosophical reason to do so - my attitude being grounded here in history. Now, in what language has been spoken *so far* about the evolution of properties? This has been done in the category of sets as we will work out now; when we say that a property x evolves *during a process*, it means we attach a set of properties to it $\{a, b, \dots c\}$ some of which might not belong to the previous universe. That is, the new state of the universe contains all actual properties (including the new ones) as well as the information about the evolution $x \to \{a, b, \dots c\}$. Unless we assume that evolution processes are irreducible, we must assume that we will allow for the birth of evolutions between *new* properties also. Now, and this is just a matter of convention; in order to "depict" the effect of a sequence of processes involving the birth of new evolutions of properties and properties themselves; it is convenient to assume that every property x is different from the properties $\{a, b, \dots c\}$ it evolves in. Another option to reconstruct the process would be to enumerate the evolutions of properties so that we know which evolution came before another in the sequence of processes. Traditionally, one opts for the first viewpoint and this "picture" is what is called *spacetime*; in case for a finite number of properties in spacetime, one can depict this by means of a so called Hasse diagram. More in detail, we draw an arrow from x to a if $a \in \{a, b, \dots c\}$ where the latter should be read as x evolves into $\{a, b, \dots c\}$ and *not* x has properties $a \wedge b \dots \wedge c$ or x has properties $a \vee b \dots \vee c$ or for that matter any proposition one can make from the properties $a, b, \dots c$. How does this relate to set theory? Well, if the evolution of properties itself is *not* irreducible, then the reader must understand that for some of the properties in which x evolves into must evolve amongst themselves, that it is possible that $b \to \{c, d, \dots e\}$. Consistency of what it means to evolve into, then should imply that

$$\{c, d, \dots e\} \subset \{a, b, \dots c\}$$

too. Now comes the fundamental assumption which is the following: our discussion suggests that we should expand the notion of evolution towards sets of properties. We assume that this evolution reduces to the individual one in the sense that $\{x, y, \dots z\}$ evolves into the union of evolutions of $x, y, \dots z$. Without this, it would be impossible to formulate initial value formulations for physics as x and y separately might not evolve into a but jointly they would. More abstract, consider the following category *prop* which consists of all subsets X, Y of properties as objects and with morphisms the inclusion $i : X \to Y$ if $X \subseteq Y$; then our evolution E is a functor $E : \textit{prop} \to \textit{prop}$ which maps every subset to the one it evolves into. E satisfies the four properties that

$$E(X \cap Y) \subseteq E(X) \cap E(Y),\ E(X \cup Y) = E(X) \cup E(Y),\ E(\{\emptyset\}) = \{\emptyset\},\ E^2(X) \subseteq E(X)$$

and moreover

$$E(\{x\}) \cap \{x\} = \{\emptyset\}$$

for any singleton $\{x\}$. All conditions, except the first one which follows from $E(X \cup Y) = E(X) \cup E(Y)$, are independent and one can wonder why we do not elevate E to a monodial category (containing *prop*) by allowing for a "join" $\otimes$ of properties just like we did for particles and properties theirof. The thing is that I would not know what it means. Spacetime, on the other hand is an identity and the description above, in terms of the category *prop* and functor E, which we will denote by the tuple $(prop, E)$ constitute the properties of spacetime; hence, we can allow for looking at spacetime as a quantum, classical, stochastic or classical-stochastic entity depending upon which operators $\wedge, \vee$ one allows for in the kinematical description. Also, here, one might go further and consider a single universe to consist out of multiple space times and therefore allowing for a join $\otimes_\alpha$ applied to space times; this would be considered as a second quantization of gravity. Even more exotic, one can quantize the universe, consider multiple universes and universes of universes, we will treat all this in greater detail later on. It is not as simple as it looks and one needs to be careful about it. So, what is now the definition of *actual* space given that we have a spacetime: it is the set of all properties $\{x\}$ such that $E(\{x\}) = \{\emptyset\}$. To relate to the literature here, the notion of evolution we have described here is equivalent to that of a partial order on the set of properties, and therefore our programme includes causal set theory. The virtue of my presentation is that one appreciates its limitations when speaking about nature while causal sets are usually perceived as a generalization of known physics. Moreover, in causal set theory, the partial order is suggested to determine a notion of causality, an interpretation which we rejected here; at best, it should have something to do with causality but the dynamics is the arbiter of that and not the kinematics. Indeed, causal set proponents regard this partial order to define the past and future of an event (which I call property), an interpretation which just does not make any sense and has lead them to consider the wrong notion of Bell causality.

2.2 The relations between the identity spacetime and the subordinate identities of matter.

For now, we have said that spacetime is constructed from properties of matter and evolutions thereof. These properties are assumed to be distinct but not necessarily atomistic in the sense that smaller properties of *particles* can be derived from relations between evolutions of these primary properties. For example, spacetime in general relativity is given by a four dimensional manifold equipped with a Lorentzian metric; the local Lorentz group is something which follows from the Lorentzian metric or one might also argue that it is encoded into it from the very start. As said before, the Lorentz symmetry, as part of the Poincaré symmetry, gives rise to the notion of spin while the translation part of the Poincaré algebra gives one a notion of mass (the arguments to get there do not need to be quantum mechanical in nature at all) and to my feeling there is a standard interplay between dynamics and kinematics here. That is, the concept of mass as well as its magnitude might follow from a mixture of kinematics and dynamical restrictions and not just from kinematics alone. We are still far removed from such an understanding in theoretical physics and certainly many theorists would dream to calculate the mass of an electron from first principles. Newton's concept of mass came from his intuition about the meaning of the concept "force" but nobody knows what that means either and certainly, as mentioned before, the units of meter and second cannot be fundamental either so that dimensional analysis really cannot be the main guide in one's thoughts albeit it is a very useful and a powerful way of thinking when dealing with theories for open systems which are written out with respect to an observer's reference frame. There appears to be no metaphysical argument beyond this which could settle further the kinematics of spacetime and therefore *deduce* the properties we are speaking of - here we must let ourselves be guided by our senses which may not be the best method after all. For example, our senses would say spacetime is four dimensional meaning that we have three dimensions of space and one dimension of time; it is very possible to construct theories with more dimensions of space and maybe of time which project down to our four dimensional experience which reminds me very much about the story of Plato's cave, that we observe a shadow world. Here, we *must* rely on our common sense and take Heisenberg's dogma that we only ought to speak about theories connecting direct relationships between our senses, and nothing more, into account. Modern physicists know that our senses aren't good enough and hope to find evidence for another structure beyond that, but then we enter into the realm of speculative theories while I would prefer a *dynamical* explanation for the emergence of our senses. That is, why to restrict kinematics beyond *reason* and refrain from finding a physical principle restricting the dynamics so that our four dimensional world rolls out on the scales we observe it? Some people try that, but not many, it is already a very hard problem to show how a smooth spacetime geometry emerges from something like a pregeometry, let alone that we can calculate an electron's mass from first principles. There will have to be done hard work indeed before we gain further insight into these matters - here I am not concerned as yet with how our properties of particles connect to our senses (which may be a complex business) and will continue to reason further on in terms of "fundamental properties" and emergent ones (deduced from relations between evolutions of fundamental properties). After all, I want to be general in this section and not be too much concerned about *our* universe; perhaps there does not exist a better explanation for the existence of our universe and we are part of a *landscape* of universes - this would be very deprimating indeed.

What I want to speak about in this subsection is the tower of relationships one can and must develop between the identities of matter and the identity of spacetime and spacetime of identities of spacetime and so on. One notices that our language falls a bit short here and one can better speak about subordinate identities; in sociology, you can compare this to the identity "state" and "citizen" although you will never hear a state say that it is a state. Spacetime *may* be different since it is a fundamental substance to reason about particles, a state though isn't of much importance for reasoning about it's inhabitants. Let me warn the reader again that what follows is how far I can see and depends upon my personal interpretation of metaphysical concepts which I try to explain patiently; this necessarily implies that I will also make by definition idiosyncratic interpretations on the current state of physics, but I will at least warn the reader when they are not mainstream. Actually, I have already done that regarding the Unruh effect, let me elaborate upon that: (a) the mainstream interpretation is that the accelerated observer has a different vacuum state and particle notion and that one needs to calculate a Bogoliubov transformation between those, this makes sense from the viewpoint of quantum mechanics as a theory of open systems relative to an observer but not with regard to closed systems such as the universe and therefore we have to dismiss that viewpoint (b) Maldacena's viewpoint, which he and I discussed a few years ago, is that the vacuum state is for sure objective and tied to the Minkowskian geometry, it is just so that the accelerated observer measures different observables which are *not* diagonal in the particle base, this already makes more sense but we have excluded such observables in our interpretation of quantum field theory as we cannot speak about a definite particle number anymore which is necessary if one acknowledges that the observer too is made out of particles, something which is badly needed for closed systems, hence we have to dismiss that interpretation also (c) my interpretation which is that the Unruh effect, as it stands, is not a viable physical effect but that something like it must be true in a theory of a closed system if one takes more physical characteristics of the observer into account, such as his mass. Particles are defined objectively but are created, in a process, *because* of the acceleration of the observer. There is no ambiguity in the particle notion of one observer with respect to another one. Let me also mention how I interpret the Hawking effect, similar to the Unruh effect, there is *no* Hawking radiation whatsoever for an observer which remains far away from the event horizon of a black hole. That is, a black hole does not objectively radiate as many sources wrongly state today! Hawking computed, just like Unruh, a Bogoliubov tranformation between the viewpoint of an observer in the asymptotic past and one which remains close to the event horizon of a black hole. It is just so that for an observer close to the event horizon, who wants to stay out of the black hole, a rather permanent acceleration is required. It is this acceleration which causes particles to be born and a radiation spectrum to be observed, but again, this depends upon many more characteristics of the observer than just his acceleration. Therefore, it is just plain nonsense, as I have repeated over the years, that modern physics would not allow anymore for a realist world view in which things objectively happen: it is rather the limited formulation of quantum theory which forbids this by the outset. In astrophysical observations, Hawking radiation could be seen by a distant observer if some matter is surrounding the black hole event horizon, but this has nothing to do with radiation being send out by a "naked" black hole.

Before we come back to our original project of speaking about the subordination of the particle identity to the spacetime identity, let me first speak about how modern science has partially dealt with this question, where partially refers to the fact that not every scientist walks this road. In the so-called perturbative approach towards quantum gravity or asymptotic safety for that matter, the insight relativists have gained in the sense that Einstein's theory should be regarded as one of dynamical spacetime is plain *rejected*. Indeed, those fellows restore the "eternalist" viewpoint of Minkowski and claim that gravity is just a force field like any other (which is against the philosophy behind the geodesic equation) and therefore is made up out of elementary interacting particles. Our kinematics, as it stands now, is fully equipped to tackle this petty world view. Obviously, those people regard the universe as an open system and are not only confronted with the fact that different observers will give inequivalent accounts but moreover must face the fact that the predictions of their theory crucially depends upon the background Minkowski spacetime they choose. That is, if they were to choose another "vacuum" classical cosmology, the predictions could not be mapped to one and another (are not equivalent), are not unique anymore (inequivalent choices of vacuum state) and moreover, they do not even know how to define (non-perturbatively) an interacting quantum field theory on a curved spacetime background. So even the very formulation of such theory is an open question! For those of us who recognize(d) that this programme is fundamentally flawed philosophically, it came as a relief that those Einstein bangers discovered an inconsistency in their own reasoning: that is the theory did turn out to be perturbatively non-renormalizable meaning one needs effectively an infinite number of coupling constants to make it consistent up to some energy level at which it goes completely havoc. Unfortunately, the tradition of quantum mechanical open systems remains to dominate the physics community until now as one has high hopes that these "technical" problems can be solved once one recognizes that particles cannot interact in points which they interpret as meaning that particles must be extended objects, like strings. I can safely make the bet upfront that this viewpoint will turn out to be fundamentally flawed too and that similar issues to non-renormalizability will show up at another level. By what I just said, I do not intend to say that it is impossible to make a consistent theory of "gravitons" if one were to develop a different quantum theory, something which we shall do in this book, rather that *standard* quantum theory falls short and that such programme has nothing to do with a *purely* quantal gravitational "force" given that the geometry of spacetime is determined by classical and not quantum degrees of freedom. We shall fully clarify this point of view in chapters eight and twelve; in the former, we develop a classical-quantum viewpoint on spacetime and define a consistent graviton theory whereas in the latter we work towards quantum spacetime.

To be entirely fair, I have had objections in the past against what I am going to say now designating such programme as too liberal and containing too many degrees of freedom and I remember having made such comments to Renate Loll about causal dynamical triangulations. My viewpoint has evolved a little over the years in the sense that the kinematical possibility of it should be allowed for but that we need an entirely new principle beyond known physics to make sure that the dynamics only profits a bit from those exuberant liberties. It is *that* what I am still lacking in Loll's programme as one needs to go beyond a quantum dynamics to solve that matter; I am pretty sure that it needs to be solved as our spacetime is observed as a classical manifold with a Lorentzian metric on it and undergoes an entirely classical dynamics on scales where matter has quantum properties. This is not so because the gravitational force is weak (that is only part of it) but it should explain why we can speak about a four dimensional continuum with a Lorentzian metric to start with, so the issue is a much more primitive one: I will explain later in greater detail what I mean.

As for the moment, our only goal is to investigate what one can and cannot speak about in physics and as the issue of dynamics is only slightly tangential to this quest, we will proceed now with "deducing" the appropriate language. I have decided to talk about the issue of this section step by step allowing each time for greater liberties and will indicate which programme in physics applies to which level of this process of generalization: as the reader will notice, causal set theory and causal dynamical triangulations are at the highest stage of kinematical liberty in modern physics but our framework goes beyond these programs too. So far, we have spoken about spacetime, actual space and properties of elementary particles being linked to *actual* space (I did not say that yet); since I have elaborated already on how one could speak about emergent properties and properties one might perhaps not derive from spacetime at all, let me introduce the following notation

$$p = (x, F_x(spacetime), \zeta)$$

where x denotes an element of actual space, F_x a functional relationship depending upon spacetime and ζ other parameters not related to spacetime whatsoever. As said before F_x should depend in a local way of spacetime around x but since we haven't even introduced any notion of topology yet, the reader does not need to know what it means precisely. So, p is an atomistic property of a particle (we will extend this framework to "extended objects" such as strings in a canonical way later on) and everything we said regarding the definition of classical and quantum theories applies to p (so we will keep x, y, z for properties related to spacetime and in particular to actual space). We will use the canonical projection

$$\pi : p \to x$$

of properties on the respective property of actual space. At least, what I am developing now is the standard accepted view in physics and the reader *should* wonder why we take only the particle properties related to actual space into account and not the particle properties which are not actual? For example, they might matter too in a future process of the universe; this would immediately lead to, amongst others, a higher quantum theory where the evolution of the wavefunction depends upon its value at previous times too. There is no philosophical principle to exclude this and from now on we shall attach spacetime properties to particles so that we will speak about evolving "histories" instead of "actualities" - the rule that only actual properties can be *measured* remains of course which is the first and primary reason why we only assigned those properties to elementary particles albeit there is no logical need for it. Fine, so this is our *final* settlement of that issue, it is the most general thing one can conceive and I have repeatedly stated that our framework would be extended later on. For all clarity, let me formalize this as follows: "**the universe consists out of spacetime and particles, where particles have properties which project down to properties (or events) of spacetime. Actual measurements can only pertain to properties which project down to actual space**". In a sense, we assume that our spacetime is "future finite" and closed meaning that for every property x, $E(x)$ has a finite measure and contains the limit events towards the "future" (this is bad, but ingrained language). As this implies that we need a topology and its Borel sigma algebra, as well as an equivalence class of spacetime measures to make that precise, we will refrain from doing so temporarily. Note that we do not need the existence of a preferred measure but merely of an equivalence class which is defined by the fact that the property of finiteness and being distinct from zero coincide. This leaves open the door for so called conformally invariant theories of gravitation which have recently been investigated again.

Let us now formalize this in a categorical language: we have that the properties of the identity spacetime are given by $(events, E)$, moreover one has that there exists a projection map $\pi : prop \to events$ which projects a property of elementary particles on its underlying event, and finally we have the operations $\cup, \vee, \wedge, \otimes_\alpha$ with which we can write down propositions about particle identities. Since spacetime is also an identity, we must wonder how to generalize these operations to the identity of spacetime keeping in mind the dependency of *prop* on spacetime and $\pi : prop \to events$. This is what we mean when we say that particles are subordinate to spacetime; first of all, logic would oblige one to speak about "evolution of properties of spacetime" but this was the result of a process taking place, remember that the reason why we had to introduce the concept of evolution of properties in the first place was that we did not want a dynamics without law. Actually, we have been a bit sloppy so far since spacetime should be endowed with other attributes than $(events, E)$ as we have said already; for example with a Borel sigma algebra and an equivalence class of measures on it. Hence, our new objects to which the operations of $\cup, \vee, \wedge, \otimes_\alpha$ should be applied are

$$\{\text{spacetime}_a, \text{particles}_i, (events, E, \mathcal{B}, [\mu])_a, prop, \pi_a : prop \to events,$$

$$\text{words in}\, p_i \,\text{where}\, p \in prop \ \text{ constructed using} \cup, \vee, \wedge, \otimes_{\alpha,s}\}$$

where it is understood that $\otimes_{\alpha,s}$ also depends upon $(events, E, \mathcal{B}, [\mu])$. $\mathcal{B}$ denotes the Borel sigma algebra and $[\mu]$ an equivalence class of measures. We will use the latin letters a, b, c to denote spacetime identities and i, j, k to denote particle identities; the composite object of one spacetime with identity a and particles with identities labeled by i is called a universe with identity (a, i) where we mean one a and multiple i. One might opt for including all particle identities in one universe even if some identity does not appear in a word, we will do this from now on and call a particle identity active in some universe if it appears in some word. So far, the question which is addressed in a small part of the physics community is that of the extension of operators to the identities of spacetime alone, and only ocasionally matter is included in some sense, meaning one looks for an extension of the operation $\wedge$ on objects of the type $(events, E, \mathcal{B}, [\mu])_a$. These programs so far stay far removed from the issues which I will address shortly; indeed, only global questions such as fluctuations on the total spacetime volume or the volume of actual space are addressed (as far as I know, one does not dispose of a well defined notion of curvature (operator) yet). The above notation for $universe_{(a,i)}$ implies that we have to talk about the *same* particles/space times in distinct universes and the *same* events/properties in different universes (possibly with the same "universe" identity). It is this extraordinary luxury I was talking about before which needs to be kept under control by a new dynamical principle (see section four) since a naive dynamics won't reproduce any universe like we know it.

Now the reason why we don't have the logical need to separately specify relationships between different properties $(events, E, \mathcal{B}, [\mu])$ of spacetime is that there are plenty of natural relations between them! For example, what are the common events and common evolutions between common events? Is the measure space of events equivalent and if not quantify in some sense how they differ; how good can one "match" one spacetime to another using measurable functions (this question is meaningful if one chooses a measure and not just an equivalence class)? As said before, in principle, one has an infinite chain where one can specify additional relationships between universes and extend the dynamics to those relationships too, but why do it? We *can* close the discussion in a simpler way by means of induced relationships due to the very definition of the properties of universes: so, it is reasonable to close the door at this level and we shall just do that for now. In principle, one can extend not only the operation $\wedge$ to universes, but also the operations $\vee$ and $\otimes_\alpha$ albeit it is unclear how the latter should depend upon the properties of the distinct universes; as we shall see,this is already no simple matter for properties of elementary particles! All one should keep in mind is that ultimately one only measures properties of particles by means of similar properties of other particles; in that sense, it is entirely plausible that a particle lives in multiple space times and distinct universes (meaning having a different identity) without us fully realizing it. We only can be guided by the classical picture of the three dimensional universe in our mind and how it relates to the actual "multiverse" we live in where by multiverse I mean a superposition of universes or even more general any word one can write down in different universe identities.

2.3 Generalization of our language: extended objects.

So far, the identity of a subatomic particle was given by a single number such as $i \in \mathbb{N}$ and we now turn to the situation of what happens if the particle identity itself contains structure such as is the case for the identity "string". While some people would say that a closed string is something which is differentiably equivalent to the circle and that one needs to examine the processes this circle is undergoing (without relying entirely upon the metaphysical concept of time); string theorists have chosen to stick with some concept of time and to formulate dynamics in an "eternalist" fashion using the string wordsheet, a hypothetical surface to be swept out be a moving string. The very idea that a physical particle carries some internal structure is an old one as one hopes to "explain" constraints on the dynamics from structural properties of the particle *and* spacetime. Of course, such explanation only gives valuable clues about nature depending upon how well one can motivate the internal structure as well as its more primitive character. As I have explained already, the quantization of spacetime is definetly a higher project than making a consistent theory of gravitons where the latter is grounded in some "Newtonian" view on gravity, that it is a force carried by means of elementary particles while Einstein's wonderful insight was that gravity is not a force at all but the very structure of spacetime. That is, it makes it possible to speak about laws for force fields in the first place; without gravitation, no law for force fields could ever be formulated. Of course, we do realize the elementary fact that in contemporary formulations of physics, which all rely upon the eternal concept of time and space, that the "gravitational field" has a mathematical structure rather similar to that of "force fields" and "particle fields" but one should not deny its fundamentally distinct status. From the point of quantum field theory, one would say that it is a vital ingredient in defining particles, so how could it be made up out of them? I have once played with the logical possibility that one can have dynamical spacetime *and* gravitons, but the latter do not gravitate meaning they do not contribute to the energy momentum tensor defining spacetime. This already goes beyond the framework of quantum field theory and we will not further pursue this option here.

Nevertheless, strings could turn out to be useful in finding out dynamical laws for elementary particles and it is from this point of view that we will address extended objects. Another *type* (we will see how in a categorical sense, there is a duality between fields and strings) of extended objects are fields; those have a long history such as the gravitational field of Newton, the electromagnetic field of Maxwell, the (classical) Klein Gordon and Dirac fields and so on. The last two can be seen in two ways; either as a *classical* field or the wavefunction of a single quantum mechanical relativistic particle depending on whether one resorts to a $\{\,\}$ or $\wedge$ interpretation as I will explain later on. Finally, I will mention the "weak equivalence" between a quantized field and the particle language developed previously. So far, we have not emphasized one piece of notation too much albeit we have spilled it out in words: when denoting x_1 we meant "particle one has property x" and $x_{1\star}$ was a shorthand for "particle one has property x_1". Actually, this very notation reveals that we assume a particle identity to be structureless and a more civilized notation would have been

$$\{p\} \stackrel{f_1}{\to} prop$$

where f_1 is the property map f of identity 1 which we also could have denoted by

$$\{p\} \times \{1\} \stackrel{f}{\to} prop$$

by putting the identity in the domain of the mapping. This point of view can now easily be extended to a more general situation

$$A \stackrel{f_1}{\to} prop$$

or its dual

$$prop \stackrel{g_1}{\to} X$$

where in the second relationship *prop* is often replaced by spacetime and X may contain some structure over spacetime but can also be independent from it. Here, A is understood to be a space with sufficient structure on it; at least one would expect it to be a measure space equipped with a (measurable) relation such as is "evolution of properties" for spacetime. One can regard the f_i, g_j as functors between categories but this is not the place to fully develop that view since we did not specify the nature of the relations on A and X yet. One can decide to keep the structures on A and X to be static or dynamical; for example, in string theory A is dynamical in some sense whereas *prop* is static but the mappings f_i, g_j are always part of the dynamical content.

The natural definition for a quantum field is then that of a field identity where one considers tuples $(\wedge_k g^k_1, \lambda)$ where λ attaches to each $g^k_1 : prop \to \mathbb{C}$ a potentiality, this is the so called Schrodinger picture. Note that there is a trivial equivalence between the words, and the potentiality theirof, of a single quantum particle (so we use $\wedge$ only) and the object of a complex valued field, assuming that the potentialities are complex valued. So we need to qualify the space of "differentiable" complex valued functions $F(g)$ of complex valued (square) integrable functions $g : prop \to \mathbb{C}$. Standard results from functional analysis reveal that a dense subset is given by

$$\sum_{n \in \mathbb{N}; p_k \in prop,\, k:1 \ldots n} \lambda(p_j : j = 1 \ldots n)(\chi_{p_1}, \chi_{p_2}, \ldots, \chi_{p_n})_{\otimes}$$

where

$$(\chi_{p_1}, \chi_{p_2}, \ldots, \chi_{p_n})_{\otimes}(g) = g(p_1)g(p_2) \ldots g(p_n)$$

giving the "equivalence" with our multi particle theory: indeed, the reader may see that the above notation is equivalent to

$$(\wedge_{n;p_j}(p_1, \ldots, p_n)_{\otimes}, \lambda).$$

Obviously, one should interpret p_i as property p_i and not as property p of particle i or $p_{i\star}$ for that matter; new identities cannot arise out of first quantization and the only identity here is the identity "field (one)". That is why I have used "equivalence" since our original framework of distinguished particles is much richer. So far, the theoretical physics community has not bothered to extend the operations $\vee$ and $\otimes_\alpha$ to field identities or even string identities albeit there one hears sometimes dreams of a "string field theory". When one interprets $g : prop \to \mathbb{C}$ as a classical field and not an "equivalent" description of a quantum particle, one allows for measurements of properties x and y whereas this is forbidden in the quantum mechanical interpretation.

2.4 Are macroscopic identities fundamental or emergent: weak reductionism.

So far, we have introduced from scratch a language which is intrinsically richer than the language used in physics up to this date; we will summarize and slightly extend our thoughts in the next section where we will address for the first time the process of measurement. So far, three main themes where relevant to our discussion: (a) the notion of a single identity (b) operations which one can perform on properties of single identities (c) the structure of spacetime and how it has evolved into history. In this section, we will once more examine the matter of identity and its possible relevance to physics: more in detail, we shall ask ourselves the question whether macroscopic identities must be regarded as fundamental or emergent. This relates to the issue which we have discussed already, that the description of a system of identities is richer than merely the union or join of them; what we wonder now is whether new identities can be attached to groups of identities and if these new identities change the dynamics in a way which is "unforeseen" by the dynamics for the constituting identities. Largely, this is of course a matter of the interplay between dynamics and kinematics where one has to resort to subjective notions such as unlikely or unplausible if one is going to judge whether something results from the interplay of molecules and atoms or whether something is inherent to the notion of what it means to be human. Let me give a programming example and one of a piece of art;

when programming a game like Farm Frenzy or Plants versus Zombies which my kids like to play, you give every type of plant or zombie a name and likewise arise the names cow, goose, sheep in Farm Frenzy. Using these variables, the programmer can define actions on them such as "zombie eats plant" or "cow produces milk"; usually these things are done in a very high level programming language which is far removed from the language of the machine which is one of bits and bytes. It is actually beneficial and more natural to write it down like that since it allows you to easily implement many more actions than those you could *reasonably* program in a direct way; the same kind of reasoning holds for a work of art which is created out of a "dead" piece of material and which transcends its materialistic configuration. It gets an identity such as does the Eiffel tower or the Mona Lisa which attract every season millions of tourists to Paris: something new has been born out of something rather plain by an act of creation, very much comparable to the birth of a biological creature out of an egg.

So this is the question of this subsection: does nature also "reasons" in terms of John or Jack, arm or leg, statue or painting, or does it each time has to *explicitly* refer to the composition of these entities in terms of elementary identities of (structureless?) particles? Does she, just like the software on a computer, speak in several languages depending upon what has to be said? I for sure believe she does and I have in the past launched the principle of weak reduction meaning that on higher scales *new* variables matter whose kinematics nor dynamics can be *reasonably* reduced to the dynamics and kinematics of the constituting lower scale (microscopic) elements. This is a weaker version of the ordinary principle of reduction which is upheld by most scientists, I believe, and which states that such reduction should exactly take place. So, in our kinematics, I could introduce

$$John = \{1, 3, 5, 7\}$$

if I were to consist out of four elementary particles only and likewise could I use a property map to find out properties of John. In cosmology, we do this all the time giving identities to stars, planets, asteroids and several pieces of interstellar junk out there; it is important for us to set up the theory (and indeed, the theory of gravitation has been discovered in that way). On the other hand, in microscopic physics one relies on the notion of identical particles, a highly debatable concept we will discuss later on.

2.5 About a definition of measurement in the "multiverse".

We will close off this section by discussing an ansatz, a thought, for the definition of a process of measurement of an elementary particle. This thought is as far as I know new and I haven't seen it discussed anywhere else in the literature. We will first spell it out for elementary particles in one universe and later on in the multiverse since the latter requires some more sophistication; finally, we present our operational language in some more abstraction, now that one has gotten acquainted meantime with its ideas and motivation theirof. Let me also stress that this section is somewhat speculative in the sense that a choice of definition is always just that "a choice", which may or may not be a very meaningful one. We will meet definitions of this kind such as is the case for the concept of indistinguishable or identical particles, the latter has a long history and has changed over time. However, I feel somewhat inspired by the founding fathers who had the idea that a measurement involved measurement apparatus and the system under study; the problematic aspect of the concrete meaning they gave to this sentence was that, in their description, they made a fundamental distinction in language between the particle and the measurement apparatus. That is, the particle was represented by a wave and the measurement apparatus by an operator acting on that wave whereas the measurement apparatus itself consists out of particles! We will propose a more symmetric definition which speaks about a change in relationships and of which the standard operator description constitutes a *part* of what is really going on. There are two key ideas to measurement, one concerns the change of "join" (entanglement in quantum theory) and the other one accompanies this principle and that is that a change in join should come with a further localization in space - a principle I will make precise shortly. It is always best to explain the idea by giving a couple of examples illustrating what you want to say; consider two particles, one with the property electron and the other one with the property photon represented by wave functions (where we suppress spin indices) $\Phi_1(x)$, $\Psi_2(y)$ or better by words and their potentialities but for sake of making the connection with the standard quantum mechanical definition, I will speak in terms of wavefunctions. In our language, there are a few possibilities for describing this system, namely as a disjoint union $\Phi_1(x) \cup \Psi_2(y)$, a classical join $\Phi_1(x) \otimes_c \Psi_2(y)$, as a quantum join $\Phi_1(x) \otimes_q \Psi_2(y)$ and also as a "superposition" $\Phi_1(x) \wedge \Psi_2(x)$, a notation which is reserved in standard quantum field theory for a single particle having the properties electron and photon. There are still other possibilities but those involve composite operators; associated to those different ways of writing things down are of course different rules for interfering probabilities based upon the potentialities. The standard situation in quantum field theory is of course given by $\Phi_1(x) \otimes_q \Psi_2(y)$ where the tensor product also depends upon the nature of the properties of the particles, in this case "Bose" and "Fermi". We will launch the following idea here, in case a process introduces a novel type of join for a single particle 1 with other particles j it were not joined with before *and* in case the properties of 1 in every word are uniformly localized in some sufficiently small spatial region, then we say that 1 had been *strongly* measured by the particles j. This is a very broad definition and I refuse to say how accurate this localization should actually be and if the reader wants to, we can speak about strong ϵ-measurements to cover for that deficit. In our example above, we could say that the electron is strongly measured by the photon if for example

$\Phi_1(x) \cup \Psi_2(y)$ evolves into

$$\Phi_1'(x) \otimes_q \Psi_2'(y) \wedge \Phi_1''(x) \otimes_q \Psi_2''(y)$$

where the joint support of Φ_1', Φ_1'' is contained within a spatial region of radius ϵ. So, what I want to convey here is that it is *not* sufficient for $\Phi_1(x) \cup \Psi_2(y)$ to evolve into $\Phi_1'(x) \cup \Psi_2'(y)$ for example, even if $\Psi_2'(y)$ is different from Ψ_2 and as such the photon's state has changed during the process and the electron's state has become localized. In such case we will speak about *weak* measurements or spontaneous localization; the very idea I want to launch now is that *all* our observations correspond to strong measurements. That is, a particle needs to get localized and entangled with some constituents (elementary particles) of the measurement apparatus before we can even speak about a measurement; this is the addition I wish to make to the standard measurement axiom in quantum mechanics. Remember here, that we introduced the novel idea before that a measurement apparatus has as well a classical as quantum description and what we posit here is that a change in its classical state necessarily is accompanied by a change in its very quantum structure, something which is impossible to describe in ordinary quantum mechanics. There is another issue, which I will highlight now and which has to do with the same fact I just mentioned, namely that the description of the world is not a pure quantum description. To illustrate what I want to say, consider a quantum-joined (Einstein Podolsky Rosen) pair of electrons, one moving left from the source towards Stern Gerlach apparatus A and the other towards B; suppose that the evolution of our universe is such that at A, the measurement occurs first. Then, after A made its measurement, what is the correct description of the relationship between the two particles? Is it a union, a classical join or still a trivial quantum join whose structure is equivalent at that moment to a classical join. Standard quantum mechanics would give answer three but here we see that this is not necessarily the case; it could be very well that the left mover joins with some particles in A and breaks its join with the right mover. This is not merely a matter of semantics but also reflects a dynamical issue; since I have no argument to prefer one over the other I will leave it at this. All I wanted to convey here is that measurement might involve a join between apparatus and particle, something which is impossible to describe in standard quantum theory. Note that I did not speak yet about the conditions a measurement takes place in, that is a matter of dynamics which we postpone to section four.

To define measurement of a particle by means of a bunch of particles (apparatus) when one allows for the $\wedge$ (and $\cup$) operation on universes, we need some more thought. First, let us note that the easiest thing to do would be to impose that the (geometry of the) boundary of our growing spacetime is classical when we make a particle measurement implying that the above definition can be straightforwardly generalized and obtain probabilities corrected by using amplitudes associated to distinct universes. This, however, does not need to be so and I could imagine dealing with "unsharp" boundaries of even different topology; we will work towards such definition in chapter twelve as it interconnects with how we are going to define the dynamics. So, let us finish this section by rehearsing and further clarifying the ontology developed in this section: we have defined spacetime by means of events or properties and the relation "evolution of properties", events were those elements which are common to all atomistic properties of elementary particles in the sense that every elementary property can be written as $(event, something\ else)$ where this "something else" could be derived from the spacetime structure, but this is not necessarily the case and distinct viewpoints exist. On properties of elementary particles, it possible to define the operations $\vee, \wedge$ to form words and determine their potentialities; note that the identity of a single particle exists out of spacetime and that the latter has to be seen as the stage in which those identities come to "live" and start to interact. Hence, it shouldn't come as a surprise that one single particle can have multiple properties at once; note that we defined four kinds of particles depending upon which of the operations are used. The next question we addressed, is how we should describe systems with multiple particles present: it is here that we introduced the join as having the same properties a single particle has and since a single particle came in four different types, likewise do joins. It is a this point that I will elaborate a bit further; for example, take the classical join of two quantum particles, then it is logical that only words of the type $word_1 \otimes_c word_2$ are allowed, such as

$$(x_1 \wedge y_1) \otimes_c z_2.$$

Here, one has to answer the question if there is any relationship between the latter word and

$$x_1 \otimes_c z_2 \wedge y_1 \otimes_c z_2$$

so that, in a sense, the wedge operation is still allowed but in a limited way. For example,

$$x_1 \otimes_c z_2 \wedge y_1 \otimes_c v_2$$

is forbidden since it is not of the right type. The same question concerning the quantum join leads in case of an affirmative answer to the linear structure of quantum mechanics. The main distinction between a union and a classical join is that the latter always involves distinct particles while the former can pertain to the same identities. The other rules regarding these operations, such as the order in which they come, were defined in a clear way and we finish the content of this chapter here.

The reader will undoubtedly wonder, given the richness of the exposition in this chapter what kind of incremental but important progress the theory to be developed in this book constitutes of. Regarding the important identity question, we remain at the level of quantum field theory by stating that it does not matter: indeed, the specific form of the theory does not contain enough data as to introduce such distinction. So, I have no pretense at all that this theory will be correct for larger, stable identities given our ideas regarding weak reductionism. This is no shame given that in life you have to start out by studying the "simple" things very carefully before you want to proceed to more complex situations. Indeed, I am rather convinced that the reader will meet a sufficient amount of non-trivial ideas and formulations which require a substantial amount of thought to remain with the deep sensation of doubt whether it might be possible to say something intelligent about the *real essence* of living creatures at all. Our theory however is a process theory and can be generalized to a wider class of space times than the ones we consider in this book. As promised in chapter one, I will be generous with ideas regarding (crazy) extensions of our framework and it is indeed my hope that the reader may ultimately enjoy this book as an organic, living and changing concept also.

Chapter 3

A quick course in mathematics.

If you found the previous chapter interesting and in need of further treatment, then you have to be well rounded in mathematics, the language in which we quantify the laws of the universe. Since this is mainly a book about physics, I have to make some choices and in spite of my willingness to dwell on the foundations of this beautiful language, this chapter is not going to represent an in depth view on mathematics. However, it will allow you to frame more important questions to be asked and it will give you a taste for what it is like to be a "real" mathematician. Therefore, my approach is intuitionist, I will feed you with a lot of examples and some abstract concepts which should allow you to know what the theory is about and why those things are so important: this will be sufficient to follow the technical details in this book. However, if you want to do research on your own and have a deeper understanding of these things, you should buy a couple of good books, some of which may be found in the reference list. I will assume the following knowledge: (a) natural, real and complex numbers (b) basics of Euclidean geometry in two and three dimensions (c) continuity of real valued functions as well as computing differentials and integrals of them. It would be beneficial for the reader to have studied the basics of vector space theory and matrices as well albeit this shall be generalized soon.

This chapter deals with the following topics in the order given: set theory, topology, simplicial complexes, the Euler number, homotopy theory, the Betti numbers, metric spaces, Lorentz spaces and posets, vector spaces, linear operators, Hilbert spaces, types of operators, Von-Neumann extensions, spectral theorem, quaternions and Clifford numbers, Nevanlinna spaces, higher dimensional derivatives and integrals, manifolds, vector fields and one forms, general tensors, Lie and exterior derivatives, de-Rahm cohomology, metric tensors, torsion and Riemann tensor, Bianchi identities, geodesics and extremization of length, exponential map, Synge's function, sectional curvature, maximally symmetric spaces, scaling of balls in hyperbolic and spherical spaces, volume comparison theorem, analytic functions, poles and residues, contour integration. There are of course plenty of other topics I could have taken into account, but these should suffice for the reader to obtain some literacy in mathematics and appreciate the content of this book. Finally, I will dwell on the issue of probability for which I need some measure theory but I will keep that treatment as short as possible.

3.1 Topology.

As mentioned before, my way of treating the subjects below will represent my view on the matters which is more intuitive than formalist albeit one should be able to fill in all fine details eventually of course. However, mathematics is not an occupation which is done by robots but nevertheless mathematicians have made it into a habit to write down things in a way such that a robot can verify its truth. Topology is one of the most basic aspects of all mathematics ranging from number theory to geometry which means it deserves to be treated first albeit high school students probably never heard of it. More than hundred years ago, people started working on the foundations of set theory and in doing so, they were inspired by things in ordinary life meaning they thought of a set as an unordered collection of more primitive items. The ideas about set theory are not well formed yet and if one wants to know all the details, then one might equally write a whole book about it; for example, the reader might enjoy the Zermelo-Frankel axiomatic system or the foundations laid out by Saunders MacLane. To give the reader a taste for the problem of saying that a set is an unordered list of items, one should note that most lists are infinite or that the item cannot be written down exactly. For example, the set of natural numbers cannot be written down given that it is infinite and it can only be "inductively" defined whereas it is in general impossible to write down an arbitrary real number. Moreover, the infinity associated to the real numbers is of a different kind than the one of the natural numbers and therefore, it is extremely hard, and mostly a matter of choice how to deal with infinity. The approach I will take is operational and I shall refrain from starting with the definition of an element, which makes the entire theory more flexible. Denoting by sets $A, B, C, \ldots$ people figured out that any decent set-theory $\mathcal{S}$ should allow for internal, associative and commutative operations $\cap$ and $\cup$, where the former refers to the "common items" and the latter to the "union of items", which are the identity on the diagonal meaning that $A \cap A = A \cup A = A$. Associativity of $\cap$ means that $A \cap (B \cap C) = (A \cap B) \cap C$ and commutativity refers to the fact that $A \cap B = B \cap A$. One, moreover, demands the existence of a unique "empty set" $\emptyset$ such that the following holds

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

where the last rule is referred to as the de-Morgan rule. From this, one can define operations like $\subseteq$ by means of $A \subseteq B$ if and only if $A \cap B = A$; one, moreover, calls A primitive if and only if for every nonempty $B \subseteq A$ holds that $B = A$ which implies that if $A \cap C \neq \emptyset$ then $A = A \cap C \subseteq C$ because $A \cap C \subseteq A$ given that $A \cap (A \cap C) = (A \cap A) \cap C = A \cap C$. So, primitive sets may be thought of as corresponding to "elements"; more precisely, for primitives A, we introduce the notation that $A = \{\widehat{A}\}$. $\widehat{A}$ is then called an element and we write $\widehat{A} \in B$ if and only if $A \cap B = A$; this implies that if $B \subseteq C$, then any element of B belongs to C as follows from the definitions. Indeed, $\widehat{A} \in B$ if and only if $A \cap B = A$ which is equivalent to $(A \cap C) \cap B = A$ implying that $A \cap C \neq \emptyset$ and therefore $A \cap C = A$ due to the primitivity of A. Also, we have that if $\widehat{A} \in B, C$ then $\widehat{A} \in B \cap C$ given that $A \cap (B \cap C) = (A \cap B) \cap C = A \cap C = A$; on the other hand if $\widehat{A} \in B \cup C$ then $\widehat{A} \in B$ or $\widehat{A} \in C$ since $A = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ implying that at least one of them must be nonempty and therefore equal to A. We also have that if $\widehat{A} \in B$ then it is in $B \cup C$ since $A \cap (B \cup C) = A \cup (A \cap C)$ and this last expression equals $A \cup A$ or $A \cup \emptyset$ due to the primitivity of A. In both cases we are done, since $A \cup \emptyset = A = A \cup A$. All of this does *not* imply that a set can be written as a collection of elements. For example, take $\mathcal{S}$ to be consisting out of $\emptyset, \{1\}, \{1,2\}$, then $\{1\}$ is primitive but $\{1,2\}$ is not a collection of elements. Therefore, we add the requirement that

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

which reads that a set equals the collection of its elements. Given all the previous remarks we have that the $\cap$ operation indeed refers to all common elements while the $\cup$ operation takes the union of them. Given that we define the natural numbers as a series $n = 1 + 1 + 1 + 1 + \ldots + 1$ by the following prescription

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

we propose that $\mathbb{N}$ is the set of all such numbers, defining the set theory of natural numbers by considering all subsets of $\mathbb{N}$. As said before, usually one starts to make a distinction between elements (or objects) and sets, one defines $\in$ and $\emptyset$ and from that one can *derive* the first three axioms of set theory. Our approach is far more general in the sense that we realize that the notion of element or point, defined by a primitive set, might be an unnecessary abstraction and perhaps not even exist. Therefore, our fourth axiom might be somewhat out of place and the reader may wish to consult category theory in that regard as higher language. Zermelo-Frankel set theory involves many more axioms than the ones mentioned above having to do with infinity, such as regards the axiom of choice. We shall not dwell on these issues here but mention that many statements in modern mathematics depend upon this axiom by means of Zorn's lemma. Since all of this is all rather formal and not intuitively obvious for anyone - indeed there are mathematicians who refute the axiom of choice - we shall not mention it here but the interested reader might consult other sources. You may also wish to play with the idea of a non-commutative or non-associative set theory in which the commutativity respectively associativity of the $\cap$ and

$\cup$ operations are abandoned. Traditional set theory also involves a fifth axiom which is that of taking the difference

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

and we shall also include this axiom on our list.

Given our rather elementary definition of set theory as well as the natural numbers, we can now proceed with our foundations of mathematics. Given two sets B, C we define the Cartesian product $B \times C$ as the *set* of all couples (x, y) such that $x \in B$ and $y \in C$, so the reader might wish to add this as a sixth axiom to $\mathcal{S}$ meaning the theory is closed with respect to the associative operation $\times$ where

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

and

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

and likewise so with the first argument exchanged by the second one. Again, I refer to category theory for a further generalization of this. A *relation* R between B and C is a subset of $B \times C$; in case $B = C$ further things can be said. As a matter of notation we denote xRy 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 a partial order, usually denoted by $\prec$ or $\leq$, and a reflexive, symmetric and transitive relation an equivalence usually denoted by $\equiv$. We have just defined the natural numbers in an abstract way using the symbol $+$; we now equip $\mathbb{N}$ with a partial order $\leq$ by saying that $n \leq n$ and $n \leq n + 1$ and taking the *transitive closure* of that. The reader may have fun now with extending $+$ to the usual sum, defining negative numbers from that resulting in the integers $\mathbb{Z}$ and defining the usual product. He or she should then show that the rational numbers may be defined from $\mathbb{Z} \times \mathbb{N}_0$ by means of the equivalence relation $(m, n) \equiv (m', n')$ if and only if there exist $k, l \in \mathbb{N}_0$ such that $lm = km', ln = kn'$ where $\mathbb{N}_0 = \mathbb{N} \setminus \{0\}$. The reader should see some possibilities for alternative constructions here, but this is how mathematics is framed in our times.

Let us now come to the issue of topology: we already spoke about sets and elements of them, now we wish to speak about open neighborhoods of elements. In framing the definition, we shall make an extension of set theory $\mathcal{S}$ so far by allowing for an infinite union, where ∞ does not necessarily refer to the "number of elements" in $\mathbb{N}$, of sets to be taken. Specifically, we define a topology $\tau(D)$ by means of a set D; here $\tau(D)$ consists of subsets of D such that (a) $D, \emptyset \in \tau(D)$ (b) $A, B \in \tau(D)$ implies that $A \cap B \in \tau(D)$ (C) $\cup_{i \in I} A_i \in \tau(D)$ for $A_i \in \tau(D)$. In other words, finite intersections of open sets are open as well as infinite unions of them. The reader may enjoy finding the intuition behind those axioms and we shall come back to them by means of specific examples; suffice it for now to understand that a topology is relative to a set D. Given $\tau(D)$, we call a subset $E \subseteq D$

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

Given a point $p \in D$, we call O an open neighborhood of p if and only if $p \in O$. Given p, a basis of open neighborhoods is given by a countable collection of open neighborhoods O_i of p, such that for any open V containing p there exists an i such that $O_i \subseteq V$. One could, moreover, demand that $O_{i+1} \subseteq O_i$ but we shall not do that in the sequel. Regarding closed sets X, Y, the reader may easily verify that the following holds: (a) $\emptyset, D$ are closed (b) $X \cup Y$ is closed and (c) $\cap_{i \in I} X_i$ is closed if all X_i are. Sets such as $\emptyset, D$ which are open and closed are called clopen. Given a set $B \subseteq D$, the intersection of a closed sets X containing B is closed and called the closure of B, denoted as $\overline{B}$. The reader smells already that a different kind of infinity is involved here as the one we met so far for the natural numbers: to deal with this, one has the so-called power set axiom in Zermelo-Frankel theory which we also adapt here. It is formulated as follows: given a set D, the power collection 2^D of all subsets of D, as well as its subcollections are sets and therefore belong to $\mathcal{S}$. The interested reader should read further upon this and might be interested in the construction of ordinal numbers by Cantor. So, we already need this kind of infinity to define the closure of a set in ordinary topology. This is how the fine details of set theory entangle with the very definitions in topology.

The reader may wish now to extend the definition of a Cartesian product to infinite so-called "index" sets; for this, we need 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 axiom is required to take unique limits as we shall see; if it were not to hold, then different sublimits may be taken as the reader might enjoy finding out. So, we denote by

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

where the I-tuples are partially ordered by means of $\prec$. Usually, we will use sequences labeled by natural numbers $\mathbb{N}$ which is totally ordered meaning that for any $n, m \in \mathbb{N}$ either $n \leq m$ or $m \leq n$. We shall denote these sequences by

$$(x_n)_{n \in \mathbb{N}}$$

and this finishes the rather abstract nonsense we had to go through. The reader may appreciate my intuitionist presentation from the perspective that it offers the feeling and insight that lots of other things can be said about the foundations of mathematics and that there are indeed those working on things like topos theory, where the axioms do not directly connect to our everyday intuition and comprehension. This may be necessary and the start of some fairly interesting adventures in the future given that the picture of the world quantum mechanics provides for is far removed from common understanding indeed. We already used the Cartesian product to define relations, now we shall use it to define functions $f : A \to B$. Again, we define a function f by means of a subset of $A \times B$ where now it is imposed that for every $a \in A$ there exists exactly one $b \in B$ such that $(a, b) \in f$. We will slightly abuse notation and write down that $f(a) = b$ which is the standard thing to do in mathematics. Again, we shall concentrate on some special cases: we call f injective if $a_1 \neq a_2$ implies that $f(a_1) \neq f(a_2)$; in case that for any $b \in B$, there exists an $a \in A$ such that $f(a) = b$ we call f surjective. When A and B are equipped with a topology, we say that f is continuous if and only if

$$f^{-1}(O) = \{x \in A | f(x) \in O\}$$

is open for any open $O \in \tau(B)$. We will connect continuity to the following obvious property; we call x a limit point of a sequence $(x_i)_{i \in I}$ if and only if for any open neighborhood O of x, there exists an $i_0 \in I$ such that for any $i_0 \prec i$ holds that $x_i \in \mathcal{O}$. Now, continuity means that if x is a limit point of (x_i) then $f(x)$ is a limit point of the $(f(x_i))_{i \in I}$ and the reader should be able to write down the formal proof. Now, we are of course not interested in all topologies; those which carry an interest have the so called Hausdorff property. $\tau(D)$ is a Hausdorff topology if and only if for any $x, y \in D$ holds that there exist disjoint open neighborhoods O, V of x and y respectively. Effectively, the condition means that distinct points are separated by means of disjoint neighborhoods.

We are now in a position to define the real numbers $\mathbb{R}$; introduce the absolute value $|q|$ of a rational number $q = \frac{n}{m} \in \mathbb{Q}$ by means of $|q| = q$ if $q \geq 0$ and $-q$ otherwise. Then the reader may verify that the following holds:

- $|p + q| \leq |p| + |q|$ called the triangle inequality,
- $|q| = 0$ if and only if $q = 0$ called the non-degeneracy condition.

A positive, non-degenerate bifunction $d(x, y)$, where the non-degeneracy means that $d(x, y) = 0$ if and only if $x = y$. satisfying the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$ is called a metric. Clearly, $d(p, q) = |p - q|$ defines a metric for all $p, q \in \mathbb{Q}$ and the reader should be able to verify that a metric defines a topology from the following axiom: the metric topology τ defined by d is generated by means of the open balls $B(x, \epsilon) = \{y | d(x, y) < \epsilon\}$ and $0 < \epsilon \in \mathbb{Q}$. Generated means that every open set in the topology can be written as a union of open balls. An infinite sequence of rational numbers $(q_n)_{n \in \mathbb{N}}$ is called Cauchy if and only if for any $0 < \epsilon \in \mathbb{Q}$ there exists an n_0 such that $n, m \geq n_0$ implies that $|q_n - q_m| < \epsilon$. On the Cauchy sequences, one can define an equivalence relation $\equiv$ by $(q_n)_{n \in \mathbb{N}} \equiv (p_n)_{n \in \mathbb{N}}$ if and only if

$$\lim_{n \to \infty} |p_n - q_n| = 0$$

where the limit lim is just another way of noting the ϵ, n definition as the reader should know. Finally, $\mathbb{R}$ is defined as the set of all equivalence classes of Cauchy sequences. The reader should extend the definitions of $+, ., ||$ from $\mathbb{Q}$ to $\mathbb{R}$.

These five pages are certainly the most abstract of this math tutorial and as I have pointed out several times, there is much more to be said about it but now we shall turn to more applied mathematics and proceed with simplicial complexes, the Euler number, homotopy and finally, the Betti-numbers. I shall be much more sloppy and intuitionist than I have been so far which is justified given that the main points can be understood at an intuitive level. We shall denote by

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

the set of n-tuples of real numbers and obviously, one can extend the notion of sum as

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

One also has the notion of a product of a real number with an n-tuple or vector by means of

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

The standard n-dimensional simplex S_n is defined as

$$S_n = \{(x_i)_{i=1}^n | x_i \geq 0 \,\text{and} \sum_{i=1}^n x_i \leq 1\}$$

where a zero complex is a point and $\sum$ stands for a general sum. The reader must see that S_1 equals an interval, S_2 a triangle and S_3 a tetrahedron and so on. The outer boundary of S_n is given by

$$F_{n+1} = \{(x_i)_{i=1}^n | x_i \geq 0 \,\text{and} \sum_{i=1}^n x_i = 1\}$$

and can be understood as "isomorphic" to a standard S_{n-1} simplex. Let us try to make this a bit more concrete, the boundary of S_2 consists out of the *oriented* lines $(x, 0), (1-x, x), (0, 1-x)$ where x runs from zero to one. Each of these lines are glued together on their boundaries $(0, 0), (1, 0), (0, 1)$ respectively which are ordinary vertices. More in particular, one can define the $n-1$ dimensional faces F_i, $i = 1 \ldots n$, by means of

$$F_i = \{(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) | x_i \geq 0 \,\text{and} \sum_i x_i \leq 1\}$$

which are isomorphic to standard $n-1$ dimensional simplices and constitute together with the outer boundary the total boundary of S_n. Now, all pairs in $\{O_n, F_i\}$ are glued together in one common simplex which we denote by $F_i \cap F_j$ and so on, until only a point is left. Note that the F_i are not in standard form, but are embedded in one dimension higher and glued together on one common $n-2$ dimensional face. At this point, it is good to formalize things a bit more: a function $f : X \to Y$ between two topological spaces is a homeomorphism if and only if f is injective, surjective, continuous and having a continuous inverse f^{-1}. It is easy to see that homeomorphisms define equivalences between topological spaces and therefore, we are often interested in classifying equivalence classes of topologies.

A n-dimensional simplex is a topological space homeomorphic to the standard simplex S_n and we have used this terminology a bit loosely before. One can glue together two simplices along a face; more precisely, let $\psi : S_n \to A$ and $\phi : S_n \to B$ be the respective homeomorphisms, then there exists a a topological space $A * B$ and injections $i_A : A \to A * B$ and $i_B : B \to A * B$ which define homeomorphisms onto their images such that $(i_A \circ \psi)(S_n) \cap (i_B \circ \phi)(S_n)$ corresponds to the boundary symplex $F_1 \subseteq S_n$. Moreover we have that

$$\phi^{-1} \circ i_B^{-1} \circ i_A \circ \psi : F_1 \to F_1$$

is a homeomorphism for which there exists a permutation $\pi \in S_n$ such that $F_1 \cap F_{i_1} \cap \ldots F_{i_k}$ is mapped onto $F_1 \cap F_{\pi(i_1)} \cap \ldots F_{\pi(i_k)}$ for $1 < i_1 < i_2 \ldots < i_k$ and $k < n$. The reader should extend this definition to the gluing of n simplices along a boundary simplex, which we shall refer to as a face in the following. A simplicial *complex* is then nothing but such gluing of simplices *or* a more general gluing onto lower dimensional simplices such as two triangles meeting in a point. We shall adapt this more general convention in what follows. All this may seem to a formal way to define something akin to lego or pyramids, but we come now to the crux of the story; that is, the most important object in the theory. This object has been the beginning, again, of much abstraction by Saunders MacLane and, indeed, it constutes a very powerful language. Denote by $\mathbb{Z}_3 = \{-1, 0, 1\}$ then it is well known that it constitutes a finite field. That is, it has a multiplication and sum and both operations define a commutative group structure on $\mathbb{Z}_3$. Concretly, this means that the sum is internal, associative, there is a unit element 0 as well as an inverse for any nonzero element and finally it is commutative. There is another, equivalent way of defining an n simplex which is by means on $n + 1$ points v_i in $\mathbb{R}^n$. The ordered $n + 1$ tuple $(v_i)_{i=1}^{n+1}$ then defines an oriented simplex homeomorphic to the standard simplex as the reader may want to verify and he or she might enjoy writing down some homeomorphism explicitly. The $n + 1$ boundary faces are then easily seen to correspond to an omission of some v_i. Given n, we define the boundary operator as

$$\partial(v_1, \ldots v_{n+1}) = \sum_{i=1}^{n+1} (-1)^i (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{n+1})$$

being a $\mathbb{Z}_3$ sum of $n-1$ simplices. Extending this boundary operator by linearity, the reader should verify that

$$\partial\partial(v_1, \ldots v_{n+1}) = 0$$

meaning that the boundary of a boundary is zero. The reader should extend this definition to a simplicial complex and notice that the same formula holds;

we shall treat this construction further in the next section on linear spaces. This is the *main* point about simplicial complexes and now we just look at topological spaces A homeomorphic to them; given two “simplicial divisions” of A the reader intuitively understands that one can find a common *refinement* of them where refinement is defined in the obvious way by means of further subdividing a simplex into smaller pieces for example by adding one internal vertex and three internal lines in a triangle. To prove these statements typically fills up some pages while one understands such thing immediately. So the reader understands immediately that it is of interest to look for characteristics of a simplicial complex which are invariant under the operation of refinement. One might enjoy understanding that one such quantity is given by the Euler number $\chi(C)$ of a simplicial complex C. It is defined as

$$\chi(C) = \sum_{i=0}^{n} (-1)^i V_i$$

where V_i is the number of i dimensional simplices in C. First check this formula in one or two dimensions and then generalize to higher dimensions. So, the Euler number is a characteristic of an equivalence class of topological spaces. We shall now provide the reader for an intuition of homology theory and then give the abstract definition in the next section. That is, the k'th homology module, $k = 0 \ldots n-1$, H_k over $\mathbb{Z}$ of a topological space A is generated by the closed, meaning having no boundary, *connected* k-dimensional subspaces which cannot be written as the boundary of a connected $k+1$ dimensional space. Here, two such subspaces are equivalent if both taken together constitute the boundary of a $k+1$ dimensional subspace. A subset B is connected if any two points in it can be joined by means of a continuous path lying in B. Also, the sum $[v_1] + \ldots + [v_m]$ represents *any* union of subspaces w_i equivalent to v_i, where the brackets stand for the equivalence class and v_i, w_i are called representants. H_n is simply defined as the module generated by the equivalence classes of closed, connected n-dimensional subspaces. The *dimension* of H_k is denoted by b_k, the k'th Betti-number and a powerful theorem which is rather simple to understand dictates that

$$\chi = \sum_{k=0}^{n} (-1)^k b_k.$$

Let us give now a couple of examples which are fairly standard and which should invite the reader to further contemplate this: first you should find a triangulation of the two dimensional sphere and then show that the Euler number is two. Also, the two sphere itself is closed, so $b_2 = 1$, clearly $b_1 = 0$ given that every closed line is the boundary of some surface and $b_0 = 1$ since a sphere has exactly one component. Hence $b_0 - b_1 + b_2 = 2$ as it should. Define now a two dimensional torus by taking a square and identifying the opposite sides; again, this manifold is closed, has one component and has basically two classes of closed lines which cannot be written as the boundary of a two dimensional subspace giving $b_1 = 2$. Hence, the Euler number should be zero as the reader can verify directly. Again, we treat the concept module more closely in the next section which concludes our first section.

3.2 Metric and linear spaces.

We have studied the definition of a metric before, a metric space is a set X equipped with a bifunction $d : X \times X \to \mathbb{R}^+$ satisfying

- $d(x, y) = 0$ if and only if $x = y$,
- $d(x, y) = d(y, x)$,
- $d(x, z) \leq d(x, y) + d(y, z)$.

d defines as before a metric topology $\tau(d)$. We shall now give away some definitions:

- (X, d) is complete if and only if any Cauchy sequence has a limit point in X,
- (X, d) is bounded if and only if there exists a $C > 0$ such that $d(x, y) < C$ for any $x, y \in X$,
- (X, d) has the midpoint property if an only if for any $x, y \in X$, there exists a z such that $d(x, z) = d(y, z) = \frac{d(x,y)}{2}$.

As usual, X is open and closed in the metric topology and the reader can give easily examples of metric spaces which are not complete implying that closedness and completeness of a space are not the same. A subset A, however, is closed if and only if it contains all its limit points in X which we shall prove now as an exercise; this prototype of proof should enable the reader to verify all other statements. Let A be closed and consider a sequence $(x_n)_{n \in \mathbb{N}}$ with an accumulation point $x \in X$, then $x \in A$ since otherwise there would exist an open ball around x disjoint from A which is in conflict that it is an accumulation point. Vice versa, suppose that any accumulation point of A in X belongs to A, then A^c is open since if not there would exist a point x in the complement and a series $(x_n)_{n \in \mathbb{N}}$ in A which has x as an accumulation point. Therefore $x \in A$ which is a contradiction. Likewise can one prove the following theorem: any compact subset K of X is closed and bounded. The opposite is not true however; for example, the closed ball of radius one around the origin in $\mathbb{R}^\infty$ is not compact.

We now come to implications of the midpoint property; given a continuous curve $\gamma(s), s \in [0, 1]$ in X, we define its length as

$$L(\gamma) = \sup_{s_0 = 0 < s_1 < s_2 \ldots < s_{n-1} < s_n = 1} \sum_{i=0}^{n-1} d(\gamma(s_i), \gamma(s_{i+1}))$$

where the supremum has to be taken given that further subdividing a partition only increases its length. A curve has finite length if and only if $L(\gamma) < \infty$; in a complete metric space having the midpoint property and $x, y \in X$, there exists a continuous curve from x to y with length equal to $d(x, y)$. It is clear that such curve minimizes the length of all curves between x and y and it is called a geodesic or distance minimizing curve. Conversely, a metric space having a curve between any two points such that $d(x, y) = L(\gamma)$ is called a path-metric space. So, given completeness, the path metric property is equivalent to the midpoint property. There is a beautiful area of abstract metric geometry where one can define a Hausdorff dimension by means of scaling properties of balls of radius r and associated to that, the notion of a measure and so on. Furthermore, one can define geodesic triangles and angles between geodesics and compare such triangles with their cousins in metric spaces of Alexandrov curvature bounded from below. In other words, a lot of the results obtained in the upcoming section on Riemannian and Lorentzian geometry can be generalized to the much more abstract setting of path metric spaces.

This author has generalized a lot of the results available on metric geometry to a Lorentzian setting, in either to the context of a spacetime geometry. In a Lorentzian spacetime, the pseudo-metric d defines a partial order $\prec$ on the events and we define the set $I^+(x)$ as the set of all events y such that $x \prec y$ or vice versa $I^-(x)$ as the set of all y such that $y \prec x$. Here, d satisfies

- $d(x, y) > 0$ implies that $d(y, x) = 0$,
- $d(x, x) = 0$,
- if $d(x, y)d(y, z) > 0$ then $d(x, z) \geq d(x, y) + d(y, z)$

and $x \prec y$ if and only if $d(x, y) > 0$; the last inequality is called the reverse triangle inequality. (X, d) has a natural topology, called the Alexandrov topology, generated by the sets $I^{\pm}(x)$ and the reader may enjoy giving examples in which the Alexandrov topology is non-Hausdorff. The issues regarding the midpoint property are much more subtle as the reader may want to figure out. In general, we define the length of a continuous causal curve $\gamma(s)$ by

$$L(\gamma) = \inf_{0=s_0<s_1<s_3<\ldots<s_{n-1}<s_n=1} \sum_{i=0}^{n-1} d(\gamma(s_i), \gamma(s_{i+1}))$$

where by definition of a causal curve any $d(\gamma(s_i), \gamma(s_{i+1})) > 0$. Here, the infimum has to be taken given that subdividing a curve usually lowers the sum. Finally, (X, d) is said to have the path metric property if and only if for any $x \prec y$ there exists a causal curve connecting x with y such that

$$L(\gamma) = d(x, y)$$

and clearly such curve is the *longest* causal curve connecting x with y. As before, such curve is called a geodesic and one can develop an entire pseudo-metric hyperbolic geometry.

In physics, one also has the causal set programme where spacetime is replaced by a discrete structure (a poset) as follows: consider a discrete set X equipped with a partial order, then $(X, \prec)$ is called locally finite if and only if for any $x \prec y$, the Alexandrov sets

$$A(x, y) = \{z | x \prec z \prec y\}$$

has finite cardinality. It is past finite if and only if the cardinal number of $I^-(x)$ is finite for any $x \in X$. General spaces of this kind, how simply defined they may be, are extremely hard to control and we now turn our head towards linear spaces which serve as the playground to define more special and controllable geometries that we did far. Indeed, it will concern specializations of the path metric geometries where limits towards the infinite small become meaningful.

In what follows, we shall need a bit of number theory: we have seen so far the natural, integer, rational and real numbers but of course, there are many more of them. We have also spoken about fields such as $\mathbb{Q}$ and $\mathbb{R}$ and now, we shall speak about rings, and division algebra's. A ring R is equipped with an addition $+$ and multiplication $.$ such that $R, +$ is a commutative group, the associative multiplication is distributive with respect to the sum meaning $(r+s).(t+v) = r.v+r.t+s.t+s.v$ and there exists a unit 1 such that $1.r = r = r.1$. A ring is a division algebra if and only if for any r there exists an inverse r^{-1} satisfying $r.r^{-1} = 1 = r^{-1}.r$. Now, we come to the definition of a left (right) module M over a ring R; a left (right) module is a commutative group $M, +$ endowed with a left (right) action of R. More specifically, there exists a scalar multiplication $rv \in M$ of an element $r \in \mathbb{R}$ with an element $v \in M$. The latter satisfies

$$\begin{aligned} 1v &= v \\ 0v &= 0 \\ r(v+w) &= rv + rw \\ (r+s)v &= rv + sv \\ (r.s)v &= r(sv) \end{aligned}$$

A vector space is a left module over a field; we have already seen that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$ where we call n the dimension. An interesting question is to find division algebra's which are at the same time bi-modules over $\mathbb{R}$. It turns out there exist exactly three of them: the real numbers themselves, the complex numbers $\mathbb{C}$ and the real quaternions $\mathbb{R}\mathbb{Q}$. All of them are suitable to define a theory of quantum mechanics but history has decided that $\mathbb{C}$ is the best candidate. The complex numbers arose from the desire to have a complete factorization of any finite polynomial over the real numbers; by this we mean that any function

$$f : \mathbb{R} \to \mathbb{R} : x \to \sum_{k=0}^{n} a_k x^k$$

with $a_k \in \mathbb{R}$, $a_n \neq 0$, could be written as

$$a_n \prod_{k=1}^{n} (x - b_k)$$

where the b_k belong to another field containing $\mathbb{R}$. As it turns out, the b_k are not real numbers in general as the polynomial

$$x^2 + 1$$

has no real roots. To accommodate for this, mathematicians invented the number i such that $i^2 = -1$; this turned out to be sufficient to factorize any complex polynomial by means of complex roots, a theorem which is called the fundamental theorem of algebra. Indeed, it is that important that it shall be used throughout this text at several stages; the reader interested in a proof should consult any decent textbook on elementary number theory and complex analysis. Later on, we shall study $\mathbb{RQ}$ as well as the so-called Clifford numbers which are of primary importance in physics.

In what follows, we shall study complex vector spaces, the restriction to real vector spaces being obvious: again, the reason for this choice is the fundamental theorem. We shall again deal with infinity here and present immediately the general case, the separable and finite cases being special examples. In $\mathbb{R}^n$ for example, one has that $v_i = (0, \ldots, 0, 1, 0 \ldots, 0)$ where the one appears on the i'th index satisfies the following properties:

- $\sum_{i=1}^{n} c^i v_i = 0$ if and only if all $c^i = 0$,
- every $v \in \mathbb{R}^n$ can be written as a sum like that.

More, in general, for any module over a ring, we call $(v_i)_{i \in I}$, where I is an index set, a basis if and only if the above two properties hold where we consider infinite sums. Now, any basis has the same cardinality, and the interested reader may want to read more about the definition of this word as well as the truth of this statement in the literature albeit we shall prove it here for finite index sets I. It can be shown, by appealing to the axiom of choice, that any module over a ring has a basis and we consider such abstract nonsense to be a fact.

We now jump to those spaces which are important in the standard formulation of quantum mechanics, the so-called Hilbert spaces $\mathcal{H}$. Those concern a generalization of standard Euclidean geometry and are defined by a so-called scalar product or sesquilinear form $\langle v|w \rangle$ where $v, w \in \mathcal{H}$ which is a module over the complex numbers. For real Hilbert spaces, the scalar product between v and w equals the product of the oriented length of the projection of v on w times the length of w and is therefore a real number. This quantity satisfies:

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

The axioms for a sequilinear form are exactly the same except that $\langle v|w\rangle = \overline{\langle w|v\rangle}$ where the bar denotes the complex conjugate defined by $\overline{a+bi} = a - bi$. For example, $\mathbb{C}$ is a Hilbert space with scalar product $\overline{v}w$; to complete the definition, the reader must understand that a scalar product is a higher structure than a metric d. More specifically, we first define a norm $||v||$ by means of

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

To understand the properties it satisfies, we need the following result

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

which reads that the absolute value of the product of the oriented length of the projection of v on w with the length of w is smaller or equal than the length of v times the length of w which is intuitively obvious. The formal proof goes as follows

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

where $\mathrm{Re}(a + ib) = a$ in either the real part of a complex number. The reader should verify that the real part of a complex number z can be written as $\frac{1}{2}(z+\overline{z})$ while the imaginary part reads $-i\frac{1}{2}(z - \overline{z})$. Here, the modulus of a complex number is defined by

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

and satisfies

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

where the last term equals up to a factor two

$$aa' + bb'$$

the absolute value of which is smaller or equal to $|a|\,|a'| + |b|\,|b'|$ the square of which

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

Therefore,

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

and thus

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

which is the so-called triangle inequality. Therefore, we obtain again a metric on the complex numbers by

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

and the reader notices that the norm is just a higher dimensional analogue of the modulus. Coming back to our original proof, the reader may see that we can choose λ such that

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

and therefore, we obtain that

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

which is a polynomial equation of second degree in the positive variable $|\lambda|$. The reader should know that for at most one zero to exist we need that the discriminant is smaller or equal than zero

$$0 \geq 4 \, |\langle v|w\rangle|^2 - 4||v||^2||w||^2$$

which gives precisely our result and equality holds only if $w = -\lambda v$. Hence,

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

which defines again a metric topology by means of

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

A Hilbert space is therefore a module over $\mathbb{C}$ equipped with a sesquilinear form such that the associated metric space is complete. The completeness assumption is extremely important for the theory of linear operators but before we come to that, let us understand a bit better the geometry defined by the sesquilinear form with regard to a so-called orthonormal basis. Two nonzero vectors v, w are called orthogonal to one and another if and only if $\langle v|w\rangle = 0$ and we say that a vector v is normalized if and only if $||v|| = 1$. Again, by means of the axiom of choice, we can show that there exists a basis $(e_i)_{i \in I}$ such that $\langle e_i|e_j\rangle = \delta_{ij}$ where δ_{ij} is the Kronecker symbol which equals 0 if $i \neq j$ and one otherwise. In a finite dimensional Hilbert space, where $v = \sum_{i=1}^{n} v^i e_i$ this implies that

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

a generalization of the standard inproduct in three dimensional Euclidean geometry. From now on, we shall adopt the Einstein convention and simply ignore the summation sign when dealing with expressions such as $\overline{v^i} w^j \delta_{ij}$. δ_{ij} is our first example of a tensor which we shall define in full generality later on.

In quantum mechanics, the following two operations are important, one of which equals the minimal join $\otimes$ called the direct product in the literature. Given two Hilbert spaces $\mathcal{H}_i$, the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ is again a Hilbert space defined in the following way: one starts from vectors $v_1 \otimes v_2$ where $v_i \in \mathcal{H}_i$ and the latter is just a different notation for (v_1, v_2). Next, one takes finite formal sums $\sum_{i=1}^{n} z_i v^i \otimes w^i$ of these elements and considers the following equivalence relation

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

We define $\mathcal{H}$ as the linear space of these equivalence classes and now turn it into a Hilbert space. The scalar product is fixed by the definition

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

and we simply take the completion of $\mathcal{H}$ in the associated metric topology. Likewise, one can define the direct sum $\mathcal{H}_1 \oplus \mathcal{H}_2$ but this time the equivalence relation is defined by

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

resulting in the scalar product

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

The reader should check that a basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ is given by $v_i \otimes w_j$ where v_i constitutes a basis for $\mathcal{H}_1$ and w_j for $\mathcal{H}_2$. Likewise, a basis for $\mathcal{H}_1 \oplus \mathcal{H}_2$ is given by $v_i \oplus 0, 0 \oplus w_j$.

Now, we come to the theory of linear operators which constitute the natural functions $A : \mathcal{H}_1 \to \mathcal{H}_2$ between Hilbert spaces. They are natural in the sense that

$$A(zv + w) = zA(v) + A(w)$$

which implies $A(0) = 0$. It is evident that this property implies that A is completely determined by its action on a basis $(e_i)_{i \in I}$ which we write down as

$$Ae_i = A_i^j f_j$$

where $(f_j)_{j \in J}$ constitutes a basis of $\mathcal{H}_2$ and the reader should keep in mind the Einstein convention. Therefore,

$$A(v^i e_i) = (A_i^j v^i) f_j$$

which, for finite dimensional Hilbert spaces, boils down to an ordinary matrix multiplication where the j index is the so called row index and the i pertains to the columns. This constitutes the second example of a tensor and you may notice that the indices are placed differently this time - we have one upper and lower index. Hence, the point of view of operators is better than the one of matrices in the sense that the former is basis independent while the latter is not; sometimes, it is convenient to work in the latter picture while making computations but usually, the reader will understand that the former is much more economic to work in. Regarding the matrix picture, we have the notion of a basis transformation which is *not* a linear operator since it acts trivially on the vectors given that it just changes the basis representation. Such changes of basis are however denoted in the same way:

$$e_i' = \sum_{k \in I} O_i^k e_k$$

and

$$f'_j = \sum_{k \in J} V_j^k f_k$$

where the O and V are invertible in the sense that there exist O^{-1}, V^{-1} such that

$$(O^{-1})^i_k O^k_l = \delta_{il} = O^i_k (O^{-1})^k_l$$

and likewise for V. We have then that

$$A'^j_i f'_j = Ae'_i = O^k_i Ae_k = A^l_k O^k_i f_l = (V^{-1})^j_l A^l_k O^k_i f'_l$$

implying that

$$A'^j_i = (V^{-1})^j_l A^l_k O^k_i .$$

In the next section, we shall further work on the theory of linear operators especially in the context of the spectral theorem but let us come back for a moment to the previous section. Here, given a topological space, we introduced the formal vector space over $\mathbb{Z}_3$ generated by all closed, connected, k-dimensional subspaces, an object which is infinite dimensional. Then, we put the equivalence relation on them

$$S^k_1 \equiv S^k_2$$

if and only if there exists a connected $k+1$ surface T^{k+1} such that $\partial T^{k+1} = S^k_1 - S^k_2$. This reduces the infinite dimensional vector space to a finite dimensional one over $\mathbb{Z}$. To construct H_k, we divide out by the module of exact k surfaces, that is those which can be written as the boundary of $k+1$ surface (those are automatically closed since the operator ∂ satisfies $\partial^2 = 0$). This brings us to the notion of a quotient space: let M, N by K modules where $N \subseteq M$, then the quotient module

$$\frac{M}{N}$$

has as elements the equivalence classes in M determined by the equivalence relation

$$v \equiv w \text{ if and only if } v - w \in N.$$

There is a long category theoretical treatment of homology and cohomology to which we shall come back later in the context of differentiable manifolds where the cohomology is determined by the exterior derivative d satisfying again $d^2 = 0$ and d and ∂ are isomorphic in a sense.

3.3 Operator theory.

We shall directly treat the infinite dimensional case, leaving the finite dimensional one as an easy subcase. However, since I shall not give any proof of any theorem which follows, the reader might be capable of formulating his or her own proof in the finite dimensional case. The latter is much easier to do given that infinite dimensions come with many subtleties not present in the finite dimensional case and consult a book on functional analysis for the more formal proofs. In standard quantum mechanics, the Hilbert space is time independent, something which we shall sharply criticize in the next chapter given that it implies the theory is not generally covariant. All mathematical results of interest pertain to operators $A : \mathcal{H} \to \mathcal{H}$ and the reader shall understand after a while why those results are not open for generalization.

Before we lift off, let us study some topologies on the Hilbert space $\mathcal{H}$ as well as on the space of operators. On $\mathcal{H}$, we have already studied the norm topology determined as before, we now come to the weak topology: this one is generated by so-called linear functionals which are linear maps ω from $\mathcal{H}$ to $\mathbb{C}$. The space of linear functionals is again a vector space and usually we restrict to those functionals which are continuous in the norm topology. The latter form again a vector space called the topological dual $\mathcal{H}^\star$: now, one has the result that a functional is continuous in the norm topology if and only if

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

for some $C > 0$. Let us give a formal proof of this result: obviously, the inequality implies continuity by linearity of ω. Conversely, take a linear functional and assume that there exists a sequence of linear independent, normalized vectors v_n such that $\omega(v_n) \to \infty$ in the limit for n to ∞. Then, by taking a subsequence, one can 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 $w = \sum_{n=0}^{\infty} \frac{1}{n^2} v_n$ whose norm is finite and $\omega(w_k) \to \infty$ which is in contradiction to the continuity.

Given this result, one can show that any bounded functional in $\mathcal{H}^\star$ is of the form

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

with $||v|| < \infty$ and the reader should try to prove this result for him or herself. Geometrically, this is obvious given that ω is fully determined by its null hyperplane $W = \{w|\omega(w) = 0\}$ as well as the action on its normal vector $\frac{v}{||v||}$. This motivates the following definition: we define the open sets

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

which constitute a basis for the so-called *weak* or $\star$-topology.

The weak topology is weaker than the norm topology meaning that every open in the weak topology is also open in the norm topology. The proof is evident by means of the inequality

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

and therefore if

$$||w - w'|| < \frac{\epsilon}{\max_{i=1}^{n} ||v_i||}$$

then $w' \in \mathcal{O}_{\epsilon;v_1,\ldots,v_n}(w)$. Another criterion for compactness of a metric space like $\mathcal{H}$ is that K is compact if and only if every sequence $(x_i)_{i\in I}$ in K has a subsequence $(x_{i_j})_{j\in J}$ converging to a limit point in K. We will now prove the equivalence with the old definition: if K is compact and $(x_i)_{i\in I}$ is a sequence, then consider a cover C^n by means of $\frac{1}{n}$-open balls. For each n, by the fact that K can be covered by means of a finite number of open balls of radius $\frac{1}{n}$, we have a sequence with a cardinal number of I elements contained in a decreasing sequence of open sets $\mathcal{O}_{n+1} \subseteq \mathcal{O}_n$ of diameter less than $\frac{2}{n}$. Hence, we find a subsequence and accumulation point in K. To prove the converse, we shall assume that K has a countable dense subset of points; therefore, any open cover of K by means of open neighborhoods can be replaced by a countable one $\mathcal{O}_n$. Suppose there does not exist a finite subcover, then we can find a subsequence of points $x_m \notin \cup_{n=1}^{m}\mathcal{O}_n$ which must have an accumulation point $x \in \mathcal{O}_k$ for some k. But then, some $x_m \in \mathcal{O}_k$ for m arbitrarily large in contradiction to the assumption.

In the sequel, we shall only work with separable Hilbert spaces, meaning Hilbert spaces having a countable basis, corresponding to the lowest kind of infinity. The reader can easily show that for such spaces, the closed unit ball is not compact in the norm topology but it is compact in the weak topology, where the above theorem also holds for the weak topology in this case given that it has a countable basis. We now come to some definition of some norms one can define on the linear space of all operators and we start by the most important one:

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

which is the so called supremum norm. In case it is finite, we call the operator bounded and a whole theory of bounded operators exists poured in the abstract framework of $C^\star$-algebra's. We shall not deal with this subcase here as most physical operators are unbounded. Two other topologies are of interest, the strong and weak $\star$ topology. The former is defined by means of the open neighborhoods

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

while the latter has as open neighborhoods

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

The reader may show that both topologies are Hausdorff and that the weak-$\star$ topology is weaker than the strong one. Also, we leave it as an exercise for the reader to show that all three topologies coincide for operators on finite dimensional Hilbert spaces.

We are now in a position to develop the theory of interest: I shall refrain from giving the physically important examples, since we need the general theory of the next section for that. So, in general, operators have a domain $\mathcal{D} \subset \mathcal{H}$

which we shall assume to be a dense subspace in the sequel; the adjoint $A^\dagger$ of A is defined by means of the following procedure. Consider the subspace $\mathcal{D}^\star$ of vectors v such that

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

for all $w \in \mathcal{D}$. Then, the functional $w \to \langle v|Aw\rangle$ can be continuously extended to $\mathcal{H}$. Therefore, there exists a vector z such that

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

and we define $A^\dagger v = z$ which can be easily seen to be a linear operator. Therefore, the domain of $A^\dagger$ equals $\mathcal{D}^\star$. The following cases are of special interest:

- $A = A^\dagger$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator is called self-adjoint,

- $AA^\dagger = A^\dagger A$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator is called normal,

- $UU^\dagger = U^\dagger U = 1$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator is called unitary,

- $P^2 = P = P^\dagger$ and $\mathcal{D} = \mathcal{D}^\star = \mathcal{H}$ in which case the operator is called a Hermitian projector.

As we shall see later on, self-adjoint or Hermitian operators are related to the unitary ones which play an important role in the standard formulation of quantum mechanics which is not suited at all for cosmology as we shall study in the next chapter. We shall work towards two different theorems: one which says that a suitable class of operators may be extended to Hermitian operators, where an extension of an operator is a novel operator defined on a larger domain coinciding with the original operator on its domain. The second result states that any normal operator may be decomposed into orthogonal projection operators in the weak $\star$ topology.

Let us first treat the extension of a so-called partial isometry V with a domain $\mathcal{D}$ which is not required to be dense; the latter satisfies the property that

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

for all $v, w \in \mathcal{D}$. Hence, by continuity, we can extend V to the closure $\overline{\mathcal{D}}$ of $\mathcal{D}$ which results in a linear homeomorphism between $\overline{\mathcal{D}}$ and $\overline{\mathrm{Im}(V)}$ where $\mathrm{Im}(V) = \{Vw|w \in \mathcal{D}\}$ is the image of V. Now, only in case the orthogonal complement $\mathcal{D}^\perp = \{w|\langle w|v\rangle = 0 \, \forall v \in \mathcal{D}\}$ has the same dimension as the orthogonal complement of the image $(\mathrm{Im}(V))^\perp$ can we obtain an extension by means of a unitary operator $W : \overline{\mathcal{D}}^\perp \to (\mathrm{Im}(V))^\perp$ resulting in a unitary operator $U = V \oplus W : \mathcal{H} \to \mathcal{H}$ which is a unitary extension of U. Now, the reader notices that for any subspace W, $W^\perp$ is closed in the weak and therefore also in the norm topology; moreover, $W^{\perp\perp} := (W^{\perp})^{\perp}$ is the weak closure of W.

Now, Von Neumann was aware of the Cayley transform between Hermitian and unitary operators in finite dimensional Hilbert spaces: more precisely, a self-adjoint operator A gets mapped to

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

where $(A \pm i1)$ is invertible in finite dimensions given that the equation $Av = \mp iv$ has no solution since otherwise

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

implying $v = 0$. Moreover, $(A + i1)$ commutes with $(A - i1)$ implying that U is unitary. He wondered what conditions should apply on A for U to be a partial isometry, so that one could extend the latter to a unitary operator defining a Hermitian one by means of the inverse Cayley transformation:

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

The operator $A \pm i1$ should be injective by the same argument as before which suggests the condition that $\mathcal{D} \subseteq \mathcal{D}^{\star}$ and $A = A^{\dagger}$ on $\mathcal{D}$ which is the defining condition for A to be a symmetric operator. This does, in contrast to the finite dimensional case not imply that $A \pm i1$ is surjective. Therefore,

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

and we now have to do three things : (a) verify that it is a partial isometry indeed (b) close $\mathrm{Im}(A \pm i1)$ and finally (c) verify whether $\mathrm{Im}(A + i1)^{\perp}$ and $\mathrm{Im}(A - i1)^{\perp}$ have the same dimension. To prove (a) we notice 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 by symmetry of A

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

for all $v, w \in \mathcal{D}$. Usually, in the literature, one closes the operator A prior to making the Cayley transform albeit this is not really necessary; U trivially extends to an operator

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

and now we need (c) for U to be extensible to $\mathcal{H}$. This last condition can be phrased somewhat more conveniently by means of

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

Indeed

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

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

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

This is true if and only if $(A^\dagger \mp i1)w = 0$ since $\mathcal{D}$ is dense in $\mathcal{H}$; by definition $\mathrm{Ker}(B) = \{w|Bw = 0\}$.

This finishes our first major result: symmetric, densely defined operators have self-adjoint extensions if and only if the dimensions of $\mathrm{Ker}(A^\dagger \mp i1)$ are equal to one and another. Now, we come to the second major result for normal operators A showing the following: there exists a projection valued measure $dP(z)$ on the complex plane such that in a weak $\star$ sense

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

Here, we meet for the first time an integral, something I presume you have studied in high school; to make precise that what I am going to say, I shall have to introduce some measure theory and theory of distributions as well. But before we come to that, we see that if we want to achieve such result, we need to study the invertibility properties of the operator $(A - z1)$. In particular, we need to look at those cases where $(A - z1)$ is not invertible; logically, there are three possibilities:

- $(A - z1)$ is not injective nor surjective; then z is said to belong to the discrete spectrum,

- $(A - z1)$ is not injective, but surjective; in that case z belongs to the residual spectrum,

- $(A - z1)$ is injective, but not surjective; in that case z belongs to the continuous spectrum.

For normal operators, we have the result that the residual spectrum is empty. Note that if A is normal, then $A_z = A - z1$ is likewise normal; also, we have that A is injective if and only if $A^\dagger$ is which can be proven by means of $Av = 0$ if and only if $A^\dagger v = 0$. However, the fact that A is surjective does not necessarily imply that $A^\dagger$ is. So suppose that z belongs to the residual spectrum, then

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

for all w implies that $v = 0$. But then $\mathrm{Ker}(A_z^\dagger) = \mathrm{Ker}(A_z) = 0$ which is a contradiction. Therefore, the residual spectrum is empty. In case z belongs to the discrete spectrum, there exists a unique Hermitian projection operator P_z on $\mathrm{Ker}(A_z)$. P_z commutes with A, $AP_z = P_z A = zP_z$ since $\langle v|AP_z w\rangle = z\langle v|P_z w\rangle = \langle \overline{z} P_z v|w\rangle = \langle A^\dagger P_z v|w\rangle = \langle v|P_z A w\rangle$, and therefore also with $A^\dagger$

given that P_z is Hermitian; moreover, if $z \neq z'$ belongs to the discrete spectrum, then $P_z P_{z'} = 0$ as follows from

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

This looks already very much like the result we want to obtain; it is clear that on separable Hilbert spaces, the discrete spectrum is discrete meaning that it consists of at most a countable number of elements. We give the example of a bounded linear operator, for which one can prove that the spectrum is compact, given by $Ae_n = \frac{1}{n} e_n$ where $n > 0$ 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 for example $\sum_{n=1}^{\infty} \frac{1}{n} e_n$ does not belong to the image of A. So, the continuous spectrum can have "measure zero" and does therefore not contribute to the spectral decomposition in this case.

The continuous spectrum is obviously empty for operators on finite dimensional Hilbert spaces and the reader should be able, at this point, to prove the spectral theorem by means of the fundamental theorem of algebra. Indeed, here the complex numbers become important since they garantuee a non-empty spectrum - a statement which we shall prove shortly - which should allow the reader to complete the proof by induction on the dimension of $\mathcal{H}$. Here, some notation becomes important: given a unit vector v, denote by

$$P = vv^{\dagger}$$

the operator defined by $Pw = v\langle v|w\rangle$. Then P is a rank-one Hermitian projection operator and in case $AP = zP$ we obtain that $Av = zv$ meaning v is a so-called eigenvector and z belongs to the discrete spectrum $\sigma_d(A)$. Therefore, in finite dimensions, our statement becomes

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

where $\sum_{z \in \sigma_d(A)} P_z = 1$. Before coming to the continuous spectrum in general, let us first show that the spectrum is non-empty in the finite dimensional case; the infinite-dimensional proof relies on methods in complex analysis, which we shall only study in a while and are therefore out of reach for the moment. I warned you I was not going to prove every statement in full detail but merely outline the general ideas, which is after all the most important thing to do. In finite dimensions, one introduces the so called determinant $\det(A)$ of a matrix, not operator, A^i_j as follows:

$$\det(A) = \epsilon_{i_1 \ldots i_n} A^{i_1}_1 \ldots A^{i_n}_n$$

where again, Einstein summation has been taken into account. Here, we meet our second tensor $\epsilon_{i_1 \ldots i_n}$ which transforms as a density. To understand this tensor, the reader must know something about the permutation group S_n; the latter consists of all bijections $\rho, \tau : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. The reader should check that S_n equipped with the standard composition $\rho\tau$ forms a non-commutative group indeed. A transposition is a bijection which swaps two indices i, j and we denote by (ij); clearly, every nontrivial permutation can be written as a product of transpositions. We now show that every product of transpositions equivalent to the identity must contain an even number of them. For this, we use the following rules: (a) $(i, j)(k, l) = (k, l)(i, j)$ if all four indices are distinct (b) $(jk)(ij) = (ij)(ki) = (ki)(kj)$ if i, j, k are all distinct an finally (c) $(ij)^2 = 1$. By means of these operations, we can rewrite our product of transpositions as

$$(i_{2n} i_{2n-1}) \ldots (i_{2k+2} i_{2k+1})(1 i_{2k-1})(1 i_{2k-3}) \ldots (1, i_1)$$

where all i_k are different from 1. Since in the product $(1 i_{2k-1})(1 i_{2k-3}) \ldots (1, i_3)$ i_1 can only get to 1 we must have another $(1, i_1)$ in this product which can again be shuffled trough so that it arrives in place of $(1, i_3)$. In this way k is even and the original product can be reduced to $(i_{2n} i_{2n-1}) \ldots (i_{2k+2} i_{2k+1})$ which does not contain 1; therefore, in a similar way, one arrives to the fact that n is even which concludes the proof. Given two distinct products p_i for a permutation ρ, then $p_1 p_2^{-1}$ is a product for 1 and therefore even, which implies that both p_i contain an even or odd number of transpositions. This motivates the following definition:

$$\mathrm{sign}(\rho) = 1$$

if ρ is written as an even product of transpositions and minus one otherwise. Now, $\epsilon_{i_1 \ldots i_n}$ is defined by $\epsilon_{i_1 \ldots i_n} = \mathrm{sign}(\rho) \epsilon_{i_{\rho(1)} \ldots i_{\rho(n)}}$ and $\epsilon_{12 \ldots n} = 1$. $\epsilon_{i_1 \ldots i_n}$ is called the totally antisymmetric symbol and vanishes if any two indices are equal to one and another. Under a basis transformation O^i_j, the symbol transforms as

$$\epsilon'_{j_1 \ldots j_n} = \epsilon_{i_1 \ldots i_n} (O^{-1})^{i_1}_{j_1} \ldots (O^{-1})^{i_n}_{j_n} = \det(O^{-1}) \epsilon_{j_1 \ldots j_n}.$$

Therefore, the ϵ symbol is basis independent if and only if one only considers transformations O with $\det(O^{-1}) = 1$.

The reader should prove that the determinant enjoys the following properties:

- $\det(AB) = \det(A)\det(B)$,
- $\det(1) = 1$,
- $\det(A^\dagger) = \overline{\det(A)}$.

From this, it follows that $\det(A^{-1}) = (\det(A))^{-1}$ and therefore, A is invertible if and only if $\det(A) = 0$. Another formula for the determinant is given by

$$\det(A) = \sum_{\rho \in S_n} \mathrm{sign}(\rho) A_1^{\rho(1)} \ldots A_n^{\rho(n)}$$

which may be helpful in the above. This brings us back to the existence of elements in the spectrum, called eigenvalues. $z \in \sigma_d(A)$ if and only if A_z is not injective nor surjective if and only if

$$\det(A_z) = 0.$$

The latter equation is of the form $P(z) = 0$ where $P(z)$ is a complex polynomial of degree n in the variable z. The fundamental theorem tells us that it can be written as

$$\prod_{i=1}^{n}(b_i - z) = 0$$

and therefore we have n roots corresponding to $z = b_i$ some of which may be equal to one and another. This finishes the proof that the spectrum is nonempty.

We now investigate the continuous spectrum in case A is bounded. For z in the continuous spectrum, we have the following: there exists a sequence of unit vectors v_n such that

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

in the limit for n to ∞. So, elements in the continuous spectrum give rise to approximate eigenvectors. We have that $\mathrm{Im}(A_z)^{\perp}$ is zero since A_z is injective and therefore $\mathrm{Im}(A_z)$ is dense in $\mathcal{H}$. Suppose, on the contrary that

$$\inf_{||v||=1} ||A_z v|| \geq C$$

then $\mathrm{Im}(A_z)$ is closed given that if $||A_z(v_n) - A_z(v_m)||$ is a Cauchy sequence then $||v_n - v_m||$ is with limit vector v and $A_z v$ is the limit point of $A_z v_n$ by continuity which shows A_z is surjective in contradiction to the fact that it belongs to the continuous spectrum. We have likewise that if $z \neq z'$ then

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

where $(v_n)_{n \in \mathbb{N}}$ corresponds to A_z and $(w_n)_{n \in \mathbb{N}}$ to $A_{z'}$ which resembles the property of Hermitian projection operators in the discrete spectrum.

We now come to the construction of the spectral measure: given a measurable subset $\mathcal{O} \subseteq \mathbb{C}$, we define $P_{\mathcal{O}}$ to be the smallest Hermitian projection operator with the property that if $z \in \sigma(A) \cap \mathcal{O}$ and $(v_n)_{n \in \mathbb{N}}$ a sequence of (approximate) eigenvectors for z, then $||P_{\mathcal{O}}(v_n) - v_n|| \to 0$ in the limit for $n \to \infty$. Here, a measurable subset A is constructed in the following way:

- any open set is measurable,
- the complement of a measurable set is measurable,
- the infinite union of measurable sets is measurable.

We shall come back to these foundational statements of measure theory later on but let us continue with the main argument. From all the above, it is clear that

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

as the reader may wish to verify. Given a countable partition $(B_n)_{n \in \mathbb{N}}$ of $\mathbb{C}$ by means of measurable sets[1] we can take partial sums

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

where $z_n \in B_n$. As usual, the integral is defined by refining the partition and the remainder of the proof consists in showing that these sums converge in the weak-$\star$ topology to the aforementioned integral as well as A. The first statement is a delicate technical exercise, while the latter result involves again the axiom of choice. We omit the proofs of these statements as they contain very little novel ideas.

The reader should try to show that the spectrum of a self-adjoint operator is a subset of $\mathbb{R}$ and for a unitary operator a subset of the unit circle $|z| = 1$. Regarding the standard formulation of quantum mechanics, one has the important Stone-Von Neumann theorem whose proof relies on the spectral theorem but which we shall not outline here. In standard quantum theory, one has a fixed Hilbert space and the time evolution is given by given by unitary operators $U(t)$ with the property that $U(t + t') = U(t)U(t')$ and $U(0) = 1$. In case the mapping $t \to U(t)$ is continuous with respect to the strong operator topology and standard metric topology on $\mathbb{R}$, there exists a unique Hermitian operator such that

$$U(t) = e^{iHt}$$

where e is a generalization to operators of the usual exponential map you know from high school. More generally, given any continuous map $f : \mathbb{C} \to \mathbb{C}$ and normal operator A, we have that

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

where we use the spectral decomposition

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

There is a rich literature on these issues and the reader in invited to learn more about operator theory albeit it does not constitute a suitable language for quantum mechanics in the presence of a gravitational field. Before we turn our head towards general analysis, let me mention two interesting generalizations of

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

the setting so far without giving any proof.

The first generalization consists in replacing the complex numbers by means of the real quaternions $\mathbb{RQ}$ which is a division algebra as well as module over the real numbers. As a vector space, it is determined by means of the elements:

$$q = a + bi + cj + dk$$

where $a, b, c, d \in \mathbb{R}$ and $k = ij = -ji$, $i^2 = j^2 = k^2 = -1$. Define the quaternion conjugate

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

then $|q|^2 = q\overline{q} = \overline{q}q = a^2 + b^2 + c^2 + d^2$ which vanishes if and only if $q = 0$. If $q \neq 0$ then $q^{-1} = \frac{\overline{q}}{|q|^2}$ which shows that $\mathbb{RQ}$ is a division algebra. One can define quaternion bi-modules and quaternion valued scalar products which constitute a natural generalization of Hilbert space. It has been shown that all the above results can be generalized towards quaternionic Hilbert spaces and some physicists have studied a good deal of quaternionic quantum mechanics.

A second generalization consists in dropping the requirement that

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

and allowing for this scalar product to become negative. Later on, we shall study such finite dimensional Nevanlinna spaces in the context of Lorentzian geometry. The infinite dimensional context is rather much more detailed and requires subtle completeness definitions. We now study analysis in general finite dimensional spaces.

3.4 Higher dimensional analysis.

In this section, we generalize pretty much everything you have learned in high school for functions $f : \mathbb{R} \to \mathbb{R}$ to functions $g : \mathbb{R}^n \to \mathbb{R}^m$. In particular, we shall study the differential Dg, vectorfields, dual fields, general tensor fields, the exterior derivative and general integration theory. All these things can be generalized to infinite dimensional spaces where one defines the so called Fréchet derivative but we shall avoid all technicalities associated to the infinite dimensional context and its inequivalent topologies. The norms which we shall use here are the usual Pythagorian ones associated to a real scalar product. The coordinates in $\mathbb{R}^m$ defined with respect to an orthonormal basis shall be denoted by x'^μ, x'^ν whereas those defined with respect to an orthornormal basis in $\mathbb{R}^n$ are denoted by x^α, x^β. Given $f : \mathbb{R}^n \to \mathbb{R}^m$, the notation $f^\mu(x^\alpha)$ therefore has an obvious meaning and we start by defining the *partial* derivatives $\partial_\alpha = \frac{\partial}{\partial x^\alpha}$. The latter is defined by means of the condition that

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

The reader may want to prove that if all partial derivatives exist in a point x, then f is continuous in that point. More generally, the function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called differentiable if and only if there exists a linear mapping $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

Clearly $Df(x) = \partial_\alpha f(x) dx^\alpha$ where $dx^\alpha(h) = h^\alpha$ and therefore if the derivative exists in a point x, all partial derivatives exist in x. The converse is not necessarilly true: take for example any continuous function which has the following restrictions $f(x, 0) = x^2, f(0, y) = y^2$ and $f(x, x) = |x|$ for x, y close to zero. Then the partial derivatives exist, but not the total derivative; one has the following theorem: if all partial derivatives exist in a neighborhood of x and are continuous in that neighborhood, then the total derivative exists in x. The continuity requirement precisely avoids the kind of pathologies present in the above example.

One can obviously take multiple partial derivatives as well as absolute derivatives of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and one has the following important property: if all second partial derivatives $\partial_\alpha \partial_\beta f$ exist and are continuous, then

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

the proof of which follows easily from the definition of the partial derivative. Now, we come to differentiable generalizations of our 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 all $k \leq n$ absolute derivatives $D^k g$ and $D^k g^{-1}$ exist. Mostly, we shall deal with C^2 or C^∞ diffeomorphisms but exceptional circumstances may occur. It is natural to form the mapping $f \circ g : \mathcal{O} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ which we write as

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

and we now want to take derivatives with respect to x^α and denote $x'^\beta(x^\alpha) = g^\beta(x^\alpha)$. The rule, which is most easily proven, and which you have also studied in high school in a restricted form, is given by

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

where again Einstein summation occurs in the δ indices. Often, this rule is written as

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

which leads to the formula

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

and of course the same with x^α, x'^β interchanged. δ^β_γ is the third tensor we meet and it is defined by $\delta^\beta_\gamma = 1$ is $\alpha = \gamma$ and 0 otherwise. On the level of the vector space, we can identify e_α with ∂_α so that the basis acquires an operational significance. Therefore

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

and therefore dx^α obtains the status of an element in $(\mathbb{R}^n)^\star$ the topological dual of $\mathbb{R}^n$. Under a diffeomorphism

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

as it should since from the above identities follows that

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

meaning that the δ^α_β tensor is an object invariant under local diffeomorphisms of $\mathbb{R}^n$. We now define vectorfields as differential operators

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

and under a diffeomorphism of $\mathbb{R}^n$ this transforms as

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

which implies we have that

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

Likewise, we have dual fields

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

transforming as

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

The vector fields $\mathbf{V}, \mathbf{W}$ form a Lie-algebra meaning that

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

is again a vector field. One can now take tensor products of vectors and one forms where we put all vectors to the left and one forms to the right. This leads to objects of the kind

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

and it transforms as

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

We call the above object an (r, s) tensor with r contravariant and s covariant indices.

We now define the wedge product of one forms dx^α, it satisfies the properties of associativity, anti-symmetry $dx^\alpha \wedge dx^\beta = -dx^\beta \wedge dx^\alpha$ and finally $dx^{\alpha_1} \wedge dx^{\alpha_2} \wedge \ldots \wedge dx^{\alpha_k}$ defines a $(0,k)$ covariant tensor. From all the above, it is clear that

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

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

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

which is coordinate independent given that the expression looks the same in any coordinate system. For a k form

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

we have that

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

which is coordinate independent given that second partial derivatives are symmetric in the indices while one takes the antisymmetric part due to the $\wedge$-product. Therefore, the entire expression behaves as a totally anti-symmetric $(0, k+1)$ tensor. As I have promised in the section on topology, the boundary operator of a topological space and the exterior derivative are cousins of one and another given that

$$d^2 = 0.$$

So, d allows for the same construction as we did for ∂ which results in de-Rahm cohomology theory, to be treated in the next section.

In the remainder of this section, we shall introduce Lebesgue integration theory, which we shall restrict later on to the differentiable context. We shall treat it in the most general way for topological spaces X, equipped with a Hausdorff topology $\tau(X)$. As mentioned in the section regarding the spectral decomposition theorem, the Borel-Sigma algebra $\mathcal{B}(X)$ defined by $\tau(X)$ is generated by the opens $A \in \tau(X)$ by taking complements and infinite unions (and therefore also intersections). Why to define such a thing in the first place? The reason was invented by Banach who showed that one can decompose the standard unit sphere in pieces and reassemble them such that one gets two unit spheres. Obviously, if such a thing is possible, then it becomes meaningless to define the area of the sphere and therefore one has to restrict measure theory to the Borel sigma-algebra where such thing cannot occur. Given $\mathcal{B}(X)$, a measure μ is defined by

- $\mu(A) \geq 0$ 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 call the measure non-degenerate if and only if $\mu(B) > 0$ for any $B \in \tau(X)$. The construction of the Lebesgue integral is rather elaborate and based upon strong convergence criteria which I have criticized in the past; nevertheless, we shall treat it here given that it has become standard material. A function $f : X \to \mathbb{R}$ is called measurable if and only if $f^{-1}(C)$, where $C \in \mathcal{B}(\mathbb{R})$, belongs to $\mathcal{B}(X)$; so, the inverse of a measurable set is measurable. The reader might want to show the obvious statement that any continuous function is measurable. Lebesgue now split the function into a positive and negative part $f = f^+ - f^-$ by $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ and defined the Lebesgue integral for positive functions alone:

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

In particular, he assumed this number to exist and defined the the integral of a real function f as

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

The extension to complex valued functions is now obvious and there are roughly speaking two important theorems regarding Lebesgue integration. The first one is Fubini's theorem which we shall use constantly when calculating integrals; its formulation goes as follows. Let X, Y be two Hausdorff topological spaces equipped with a Borel-sigma algebra $\mathcal{B}(X), \mathcal{B}(Y)$, then one can construct $\mathcal{B}(X \times Y)$ from the topology on $X \times Y$ generated by open squares of the type $A \times B$ where $A \in \tau(X)$ and $B \in \tau(Y)$. Clearly $\mathcal{B}(X) \times \mathcal{B}(Y) \subset \mathcal{B}(X \times Y)$ and one can obviously define product measures $\mu \times \nu$. Let $f : X \times Y \to \mathbb{C}$ be a measurable function, then we have the following statement. If $\sup_{y\in Y} \left|\int_X f(x, y)d\mu(x)\right| < \infty$ and likewise $\sup_{x\in X} \left|\int_Y f(x, y)d\nu(y)\right| < \infty$ then the integral

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

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 uniformity of the bound, by means of the supremum, is a necessary requirement. Consider for example the space $\mathbb{R} \times \mathbb{R}^+_0$ and the function $f(x, y) = e^{-x^2 y}$ then

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

and

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

both of which are not uniformly bounded. We have that

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

whereas in the first expression the behaviour of $\sqrt{\frac{\pi}{y}}$ at $y = +\infty$ causes the divergence, while in the latter the pole at $x = 0$ of $\frac{1}{x^2}$ is the cause for the pathology.

The second theorem is called Lebesgues dominated convergence theorem and it goes as follows: suppose one has a sequence f_n of measurable functions converging point wise to a measurable function f such that $|f_n| \leq g$ for all n and g is integrable. Then,

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

and $\left| \int f_n d\mu \right| \leq \int g d\mu$. The proofs of those theorems are rather boring but the interested reader may enjoy a course on abstract measure theory. In the next section, we shall look at differential forms as defining differentiable measures, study a generalization of partial integration called Stokes theorem, but at this point it is sufficient that the reader understands how integration works.

3.5 Riemannian and Lorentzian geometry.

Now we come to the most important section of this chapter and the reader should learn to think like an analyst geometer. The geometer reasons in a very simple language and the interested reader shall be astonished by the very richness of the conclusions one can draw; Gromov has produced astounding results for Riemannian geometry while this author has opened the study of the even more difficult case of Lorentzian geometry. The analyst will study differential geometry and look at it as an exercise in solving ordinary and partial differential equations, an approach which we shall also embrace. I shall try to persuade the reader of the point of view of the geometer-analyst which is rich and full of insight and technical power. This section shall be organized as follows: first we define differentiable manifolds with or without boundary and we generalize the tensor and exterior calculus developed in the previous section to that context. Next, we study Stokes theorem and construct the de-Rahm isomorphism between the homotopy and cohomology theory constructed in the previous sections. Then, we move on to define geometry, that is, we shall study path-metric and pseudo path-metric spaces which are locally determined by a $(0, 2)$ covariant tensor defining a real scalar product. This leads to an analytic characterization of geodesics, the definitions of the exponential map as well as Synge's function. Finally, we study some important theorems in Riemannian geometry regarding the scaling behaviour of balls of radius r given lower bounds on the sectional curvature.

A C^n manifold $\mathcal{M}$ is a topological space which locally looks like $\mathbb{R}^n$; more specifically, there exists a cover $\mathcal{O}_\alpha$ by means of open charts $\phi_\alpha : \mathcal{O}_\alpha \to \mathcal{V}_\alpha \subseteq \mathbb{R}^n$ where the ϕ_α are homeomorphisms onto their image and $\mathcal{V}_\alpha$ is either open or the intersection of an open neighborhood $\mathcal{W}_\alpha$ of the origin with the half space $\{x|x_n \geq 0\}$. In the latter case, $W_\alpha \cap \{x|x_n = 0\}$ captures a part of the boundary. In case $\mathcal{O}_\alpha \cap \mathcal{O}_\beta \neq \emptyset$, one has that $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(\mathcal{O}_\alpha \cap \mathcal{O}_\beta) \to \phi_\beta(\mathcal{O}_\alpha \cap \mathcal{O}_\beta)$ is a C^n diffeomorphism. We shall assume the manifold to be paracompact, meaning that it can be covered by a countable number of compact sets, and not necessarily connected albeit the latter assumption would have certain advantages and constitutes a standard requirement in the classic work by Hawking and Ellis. $\mathbf{V}$ is a vector field on $\mathcal{M}$ if and only if for any function $f : \mathcal{M} \to \mathbb{R}$ and any chart $(\phi_\alpha, \mathcal{O}_\alpha)$ on $\mathcal{M}$ holds that there exists a vector field $\mathbf{V}_\alpha$ 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 $\mathbf{V}$ in the chart $(\phi_\alpha, \mathcal{O}_\alpha)$. From this, one defines dual fields by means of their action on vector fields and from thereon general tensor fields. Likewise, we generalize the definition of a k-form as well as the exterior derivative; we now come to the definition of the push forward, pull back and Lie derivative. A diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ is a homeomorphism such that for any chart $(\phi_\alpha, \mathcal{O}_\alpha)$ and chart $(\phi_\beta, \mathcal{O}_\beta)$ such that $\psi(\mathcal{O}_\alpha) \cap \mathcal{O}_\beta \neq \emptyset$ holds that

$$\phi_\beta \circ \psi \circ \phi_\alpha^{-1} : \mathcal{V}_\alpha \to \phi_\beta \circ \psi(\mathcal{O}_\alpha) \subseteq \mathcal{V}_\beta$$

is a C^n differentiable mapping and likewise for the inverse ψ^{-1}. This is just an agreement we make, that the degree of differentiability of diffeomorphisms is the same as the one of the "chart transformations". Given a diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ and a 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 defined as the push forward by means of ψ^{-1} and is denoted by $\psi_\star f$. The reader should generalize the above to 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 way as before, we can, by 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 obvious replacement for one forms and tensor fields. We now intend to measure the "difference" of $\phi^\star \mathbf{V}$ with $\mathbf{V}$; since the latter has to behave in a coordinate covariant way, only infinitesimal differences are allowed for. Here, it is useful to introduce a one parameter family of diffeomorphisms ψ_t such that $\psi_{t+s} = \psi_t \circ \psi_s$ for t, s sufficiently small and $\psi_0 = \mathrm{id}$ the identity transformation. Hence, we define the differential

$$\frac{d\psi_s^\star f}{ds}|_{s=0}(x) = \mathbf{V}(f)$$

where $V^{\alpha}(x) = \frac{dx^{\alpha}(\psi_{-s}(x))}{ds}_{|s=0}$. This vector field determines completely the diffeomorphsims ψ_s by means of

$$\frac{dy^{\alpha}(\psi_s(x))}{ds} = -V^{\alpha}(\psi_s(x)).$$

So, we have found a correspondence between a one parameter family of diffeomorphisms and vector fields; this allows us to define the Lie-derivative of a general tensor

$$\mathcal{L}_{\mathbf{V}}(T)(x) = \lim_{s \to 0} \frac{\psi_s^{\star} T(x) - T(x)}{s}$$

and we now calculate the Lie-derivative of a function and a vector field. By definition, one has that $\mathcal{L}_{\mathbf{V}}(f)(x) = \frac{d}{ds}|_{s=0} \psi_s^{\star} f(x) = \mathbf{V}(f)(x)$ and likewise

$$(\mathcal{L}_{\mathbf{V}}\mathbf{W})(f)(x) = \frac{d}{ds}|_{s=0}(\psi_s^{\star}\mathbf{W})(f)(x) = \frac{d}{ds}|_{s=0}\mathbf{W}(\psi_{-s}^{\star} f)(\psi_{-s}(x)) =$$

$$-\mathbf{W}\mathbf{V}(f)(x) + \mathbf{V}\mathbf{W}(f) = [\mathbf{V}, \mathbf{W}]\,(f)(x).$$

In order to further understand some crucial properties of integrals, one immediately notices that

$$d(\psi^{\star}\Omega) = \psi^{\star}(d\Omega)$$

for any k-form Ω, and therefore

$$\mathcal{L}_{\mathbf{V}} d = d\mathcal{L}_{\mathbf{V}}.$$

This can also be shown by means of

$$\mathcal{L}_{\mathbf{V}} = d i_{\mathbf{V}} + i_{\mathbf{V}} d$$

on the space of k-forms, where $i_{\mathbf{V}}$ is the contraction with the vector $\mathbf{V}$ defined as

$$i_{\mathbf{V}} \Omega_{\mu_1 \ldots \mu_k} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_k} = V^{\mu_1} \Omega_{\mu_1 \ldots \mu_k} dx^{\mu_2} \wedge \ldots \wedge dx^{\mu_k}.$$

Measure theoretically, one has that by definition $(\psi^{\star}\mu)(\psi(A)) := \mu(A)$ for any measure μ and measurable set A; therefore

$$\int_{\mathcal{B}} f d\mu = \int_{\psi(\mathcal{B})} (\psi^{\star} f)(\psi^{\star} d\mu).$$

Now, we come to the full power of our approach which is that an n form $\Omega = \Omega_{\mu_1 \ldots \mu_n} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_n}$ defines a measure μ by regarding every open set A as a limit from the outside and inside by means of a cover of "small" cubes determined by *constant* vectors $v_1, \ldots, v_n$ and recognizing that $(dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_n})(v_1, \ldots, v_n)$ equals $\det(v_j^{\mu})$. The latter constitutes indeed the Euclidean volume of the "small" hypercube as measured in that coordinate system so that the total volume equals $\Omega_{\mu_1 \ldots \mu_n} \det(v_j^{\mu})$ which is clearly coordinate independent.

Therefore, both our definitions of $\psi^\star\mu$ and $\psi^\star\Omega$ coincide and we write in general integrals by means of

$$\int_{\mathcal{M}} f\Omega.$$

Now, we come to the multidimensional generalization of partial integration, a theorem which is most easily proven by choosing special coordinate systems adapted to the boundary and recognizing that all our definitions above have been coordinate independent. The result reads

$$\int_{\mathcal{M}} d\Omega = \int_{\partial\mathcal{M}} \Omega$$

for any $n-1$ form Ω. Diametrically opposite to homology theory, we are interested in the $\mathbb{R}$ module C_k of k-closed forms Ω_k satisfying $d\Omega = 0$, and which are equivalent up to an exact k form $d\Omega_{k-1}$; the difference with standard homology theory being that d increases dimension whereas ∂ decreases it. This suggests a duality which is that $H_k^\star = C_k$; indeed, consider the action of an element $\Omega_k \in C_k$ on a closed surface $S_k \in H_k$ defined by

$$\widehat{\Omega_k}(S_k) = \int_{S_k} \Omega_k$$

then we show the action is well defined. In case S_k is equivalent to S'_k being the boundary of T_{k+1} we have that

$$\widehat{\Omega_k}(S_k) - \widehat{\Omega_k}(S'_k) = \int_{T_{k+1}} d\Omega_k = 0$$

where we have used Stokes theorem as well as the closedness of Ω_k. Also,

$$\widehat{\Omega_k + d\Omega_{k-1}}(S_k) = \widehat{\Omega_k}(S_k) + \int_{\partial S_k} \Omega_{k-1} = \widehat{\Omega_k}(S_k)$$

where we used Stokes theorem and the closedness of S_k. We first show that the mapping is injective; clearly, a nontrivial closed $\Omega_k \in C_k$ demands the existence of a nontrivial S_k such that $\widehat{\Omega_k}(S_k) \neq 0$. If not, then the integral over *any* 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 $\partial R_k = \partial R'_k$. Moreover, the dependency on the boundary is additive and local as easily follows from the additive and local character of the integral. Therefore, there exists a $k-1$ form Ω_{k-1} such that $d\Omega_{k-1} = \Omega_k$. One shows, moreover, surjectivity meaning that for any nontrivial $S_k \in H_k$ there exists a closed, but not exact, Ω_k such that $\widehat{\Omega_k}(S_k) = 1$; the obvious thing to do is to take any differential form Ω_k defined on S_k having this property and show that the equation $d\Omega_k = 0$ has at least one solution (in general, there is an infinite number of them). This constitutes an easy exercise in imposing periodicity conditions on k-dimensional surfaces so that one obtains a globally well defined solution. So, the Betti numbers can likewise be calculated from cohomology instead of the simplicial homotopy.

Until so far, we have said nothing about geometry and concentrated merely on calculus on manifolds; now, it is time to unwaken the geometer and use these tools to manipulate geometrical concepts such as geodesics: they constitute the true way to think about geometry and associated physical processes. In a previous section, we introduced a real scalar product as being more delicate than a norm which in its turn gave rise to a metric d. We shall now further generalize the geometry by letting the scalar product depend upon the point $x \in \mathcal{M}$ where the latter is a differentiable manifold. So, we are interested in $(0,2)$ covariant tensors $h_{\alpha\beta}$ and $g_{\alpha\beta}$ which produce local "orthonormal" bases v_a and 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

$$\eta_{11} = 1, \eta_{ij} = -\delta_{ij}; i, j = 2 \ldots n$$

and all other components vanish. Often, for $n = 4$, e_a is called a tetrad or vierbein; metrics like δ_{ab} are called Riemannian and constitute a generalization of finite dimensional real Hilbert spaces whereas η_{ab} are Lorentzian given that they define non-compact null sets

$$\eta_{ab} v^a v^b = 0.$$

Our first task consists in showing that these two geometries constitute specifications of the usual and Lorentz path-metric theories. In those cases, we had that the length of the shortest curve or longest causal curve was equal to the metric or pseudo-metric distance respectively. Before we can obtain this result, we have to say something about the local causal structure determined by η_{ab}. A vector v^a is called (a) causal if and only $\eta_{ab} v^a v^b \geq 0$ (b) null or lightlike if and only if $\eta_{ab} v^a v^b = 0$ and (c) spacelike if $\eta_{ab} v^a v^b < 0$. Moreover, a causal vector is future pointing with respect to e_a if and only if $v^0 > 0$. In what follows, we shall assume that there exists a globally well defined tetrad which excludes certain topologies of the Mobius strip type. Technically, this assumption is known under the name that the geometry $(\mathcal{M}, g)$ is time-orientable and 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 at any point is so. The length of a causal curve is given 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 a Riemannian geometry. We define $J^+(x)$ as the set of points y which can be joined to x by means of a future pointing causal curve starting at x. Similarly, $J^-(x)$ is defined as the set of all y which can be joined to x by means of a future pointing causal curve starting at y. We now look for extremal causal curves between two points x and y which are kept fixed. Therefore, we have to take a kind of "differential" with respect to the curve and put the latter equal to zero; the result reads

$$0 = \delta L(\gamma) = \int_a^b \frac{1}{2\sqrt{g(\dot{\gamma}(s), \dot{\gamma}(s))}} \left(2g_{\mu\nu}\delta\dot{\gamma}^\mu(s)\dot{\gamma}^\nu(s) + \partial_\alpha g_{\mu\nu}\dot{\gamma}^\mu(s)\dot{\gamma}^\nu(s)\delta\gamma^\alpha(s)\right)$$

and the reader should further rewrite this equation, by using that $\delta\dot{\gamma}^\mu(s) = (\delta\dot{\gamma^\mu}(s))$, until it reduces to the form

$$\int_a^b F(\gamma(s), \dot{\gamma}(s), \ddot{\gamma}(s))_\mu \delta\gamma^\mu(s) ds.$$

Noticing then that this expression must vanish for all $\delta\gamma^\mu(s)$ leads to the 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 an affine parametrization given that the above remains invariant under reparametrizations of the type $s \to bs + a$. Given that the integral is invariant under reparametrizations and coordinate transformations, this equation should transform as a vector under coordinate transformations and the reader may verify the result of changing the parametrization. We now compactify somewhat the notation of this equation by means of

$$\dot{\gamma}^\nu \nabla_\nu \dot{\gamma}^\mu = 0$$

where

$$\nabla_\nu V^\mu = \partial_\nu V^\mu + \Gamma^\mu_{\nu\kappa} V^\kappa$$

for any vector V^μ. The reader must check here that $\dot{\gamma}^\mu(s)\partial_\mu = \frac{d}{ds}$ and that $\nabla_\mu V^\nu$ transforms as a $(1,1)$ tensor; also, $V^\nu \nabla_\nu W^\mu - W^\nu \nabla_\nu V^\mu - [V, W]^\mu = 0$ for all vectors $\mathbf{V}, \mathbf{W}$. The former property means that the derivative has a geometric significance while the latter says 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 called the geodesic equation and the associated curves are geodesics. We now extend the definition of the covariant derivative as follows

$$\nabla_\mu f = \partial_\mu f$$

or in a coordinate free notation,

$$\nabla f = df, \nabla \mathbf{V} = \nabla_\mu V^\nu dx^\mu \otimes \partial_\nu.$$

Denoting by

$$\nabla_{\mathbf{V}} = V^{\mu}\nabla_{\mu}$$

we now extend the definition by

$$\begin{aligned} \nabla_{\mathbf{W}}(\omega(\mathbf{V})) &= (\nabla_{\mathbf{W}}\omega)(\mathbf{V}) + \omega(\nabla_{\mathbf{W}}\mathbf{V}) \\ \nabla_{\mathbf{W}}(S \otimes T) &= (\nabla_{\mathbf{W}}S) \otimes T + S \otimes (\nabla_{\mathbf{W}}T) \end{aligned}$$

where ω is a one form and S, T constitute general tensors. From the specific form of the Christoffel connection, the reader notices that $\nabla g = 0$ which means that the metric is covariantly constant. It is quite obvious that geodesics maximize the Lorentzian distance in case they are causal and minimize it in case of a Riemannian metric.

We shall study now a couple of tensors one can construct from the covariant derivative which are of primary importance in geometrical analysis and the theory of relativity. The first one constitutes a $(1,2)$ tensor field

$$\mathbf{T}(\mathbf{V},\mathbf{W}) = \nabla_{\mathbf{V}}\mathbf{W} - \nabla_{\mathbf{W}}\mathbf{V} - [\mathbf{V},\mathbf{W}]$$

and the reader verifies indeed that $\mathbf{T}(\mathbf{V},\mathbf{W}) = -\mathbf{T}(\mathbf{W},\mathbf{V})$ and

$$\mathbf{T}(f\mathbf{V} + \mathbf{Z}, \mathbf{W}) = f\mathbf{T}(\mathbf{V},\mathbf{W}) + \mathbf{T}(\mathbf{Z},\mathbf{W}).$$

In case of our metric connection, the torsion tensor vanishes and the reader is invited to study connections with torsion. The second tensor constitutes a $(1,3)$ tensor field denoted by

$$\mathbf{R}(\mathbf{V},\mathbf{W})\mathbf{Z} = \nabla_{\mathbf{V}}\nabla_{\mathbf{W}}\mathbf{Z} - \nabla_{\mathbf{W}}\nabla_{\mathbf{V}}\mathbf{Z} - \nabla_{[\mathbf{V},\mathbf{W}]}\mathbf{Z}$$

which is again antisymmetric in $\mathbf{V}, \mathbf{W}$. In components, this reads $R_{\mu\nu\alpha}^{\ \ \ \ \beta}$ and we can raise and lower indices with $g^{\alpha\beta}$ and $g_{\alpha\beta}$ respectively. From the general Jacobi identity $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$ for operators, we obtain that

$$R_{[\mu\nu\alpha]}^{\ \ \ \ \ \ \beta} = 0, \nabla_{[\alpha}R_{\beta\gamma]\kappa}^{\ \ \ \ \ \ \delta} = 0$$

where the square brackets stand for total anti-symmetrization. These identities are called the first and second Bianchi identity respectively; in case of our torsionless metric connection, the reader should verify that

$$R_{\mu\nu\alpha}^{\ \ \ \ \kappa}g_{\kappa\beta} = R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu}$$

by working directly in the component form. This tensor field is known as the Riemann curvature tensor and is of ultimate importance in gravitational physics. Its first contraction

$$R_{\mu\nu} = R_{\mu\alpha\nu}^{\ \ \ \ \alpha}$$

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 which is given by

$$dV(x) = \sqrt{\det(g_{\mu\nu})} dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n$$

and the reader should verify that this expression is coordinate independent indeed. This gives rise to the so-called Einstein-Hilbert action

$$\int dV(x)R(x)$$

whose variation towards the metric reproduces the Einstein equations.

We now come to a more in depth study of geodesics: the reader may verify that for a geodesic, the length of a unit vector remains the same meaning

$$\frac{d}{ds} g(\dot{\gamma}(s), \dot{\gamma}(s)) = 0.$$

Hence, we choose a parametrization such that the latter equals one for timelike geodesics, zero for null geodesics and minus one for spacelike geodesics. Denote by $T^\star\mathcal{M}_x$ the linear space of all vectors at x, then we can define a mapping

$$\exp_x : T\mathcal{M}_x \to \mathcal{M} : v \to \exp_x(v)$$

where $\exp_x(v)$ equals the endpoint of the geodesic of affine parameter length one such that the tangent vector at x is given by v. Obviously,

$$\mathrm{D}\exp_x(0)(w) = w$$

where $w \in TT^\star\mathcal{M}_{(x,0)} \sim T^\star\mathcal{M}_x$ which is nothing but the expression of the fact that the geometry is locally Minkowskian. Denoting by $T^\star\mathcal{M} = \cup_{x\in\mathcal{M}} T^\star\mathcal{M}_x$, we can define a topology which equals the product topology $\mathcal{O}\times\mathcal{V}$ where $\mathcal{O} \subseteq \mathcal{M}$ and $\mathcal{V} \subseteq \mathbb{R}^n$. The exponential map is then 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 local diffeomorphism. Given two points x and y there can exist multiple geodesics joining these two points due to a nontrivial topology or the existence of focal points; in that regard do we derive the geodesic deviation equation as follows. Given a one parameter family of geodesic congruences $\gamma(s,t)$ where $s : a \ldots b$ and $t \in (-\epsilon, \epsilon)$, we define $\mathbf{V} = \left(\frac{\partial}{\partial s}\right)^\star$ and $\mathbf{Z} = \left(\frac{\partial}{\partial t}\right)^\star$ then

$$[\mathbf{V}, \mathbf{Z}] = 0$$

and therefore

$$\nabla_V \nabla_{\mathbf{V}} \mathbf{Z} = \nabla_{\mathbf{V}} \nabla_{\mathbf{Z}} \mathbf{V} = \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}$$

where we have used in the first equality the torsionless character of the connection, whereas in the second equality the geodesic equation $\nabla_{\mathbf{V}}\mathbf{V} = 0$ was of importance. This is a kind of Newtonian harmonic oscillator with friction where the friction and spring constants arise from the connection symbols, and first derivatives thereof. Negative friction causes expansion whereas positive friction causes contraction. Along a geodesic, or actually any curve, one can define the notion of parallel transport, respectively Fermi-Walker transport, of a tensor by means of

$$\nabla_{\mathbf{V}}\mathbf{T} = 0$$

where $\mathbf{V}$ is the tangent vector to the geodesic. It is easy to check that parallel transport preserves scalars and in particular

$$g(\mathbf{Z}, \mathbf{W})$$

for any $\mathbf{Z}, \mathbf{W}$. Therefore, parallel transport defines a Lorentz transformation $\Lambda(x,w)_b^{a'}$ where $w \in T^\star\mathcal{M}_x$ and b a Lorentz index with respect to $e_b(x)$ and a' a Lorentz index with respect to $e^{a'}(\exp_x(w))$. A Lorentz transformation is a symmetry of the Minkowski metric, meaning that

$$\Lambda(x,w)_b^{a'}\Lambda(x,w)_d^{c'}\eta_{a'c'} = \eta_{bd}.$$

The reader should verify that those transformations form a continuous group of dimension $\frac{n(n-1)}{2}$ and that it has 4 disconnected components in case $n = 4$ determined by the transformations $1, T, S, ST$ where T stands for time reversal and S for space reversal.

We now come to the construction of Synge's function, a long forgotten mathematical gadget generalizing the quadratic form on Minkowski

$$\sigma(x,y) = \frac{1}{2}(y-x)^\mu(y-x)^\nu\eta_{\mu\nu}$$

where x, y represent points in $\mathbb{R}^4$. Therefore, it is appropriate to define $\sigma(x,y)$ from the action principle

$$I(x,y) = \frac{1}{2}(t_1 - t_0)\int_{t_0}^{t_1} g_{\mu\nu}\frac{dx^\mu(s)}{ds}\frac{dx^\nu(s)}{ds}ds$$

where $x^\mu(s)$ is a curve connecting x with y. This expression is invariant under affine reparametrizations $s \to as + b$ of the curve and therefore, one can assume any variation $\delta x^\mu(s)$ to be as such that the end values t_0, t_1 remain fixed. Hence,

$$\delta I(x,y) = (t_1 - t_0)\int_{t_0}^{t_1} g_{\mu\nu}\left(\frac{D}{ds}\frac{dx^\mu(s)}{ds}\right)\delta x^\nu(s)ds$$

where

$$\frac{D}{ds} = \nabla_{\frac{d}{ds}}$$

and $\delta x^\mu(t_0) = \delta x^\mu(t_1) = 0$. Demanding this to vanish is equivalent to

$$\frac{D}{ds}\frac{dx^\mu(s)}{ds} = 0$$

which is the geodesic equation and therefore one defines

$$\sigma(x, y, w) = \frac{1}{2}g(w, w) = \frac{1}{2}\epsilon L^2(x, y, w)$$

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 $\exp_x(w) = y$ and finally, $\epsilon = 1$ for timelike geodesics and -1 for spacelike ones. Assuming that w varies continuously when x and y do, w merely serves here as an indicator of the fact that different geodesics between x and y may exist, the reader may verify that

$$\sigma(x, y, w)_{,\mu} := \partial^x_\mu \sigma(x, y, w) = -w_\mu = g_{\mu\nu}w^\nu, \sigma(x, y, w)_{,\mu'} := \partial^y_{\mu'}\sigma(x, y, w) = -g_{\mu'\kappa'}\Lambda^{\kappa'}_\nu(x, w)w^\nu.$$

Hence it follows that

$$2g^{\mu\nu}\sigma(x, y, w)_{,\mu}\sigma(x, y, w)_{,\nu} = \sigma(x, y, w)$$

and likewise for the primed derivatives. Synge's function has some other properties and I shall make use of those in the crucial chapters of this book.

Now, we shall gradually turn our head towards Riemannian geometry: whereas the presentation of the above concerned Lorentzian metrics, everything safely projects down to the Riemannian case. The results below however only apply to Riemannian metrics; given two vectors v, w, the surface of the parallelepid spanned by v, w is given by $g(v, v)g(w, w) - (g(v, w))^2$ and the sectional curvature $s(v, w)$ is defined as

$$s(v, w) = \frac{g(R(v, w)v, w)}{g(v, v)g(w, w) - (g(v, w))^2}.$$

A metric is said to be of constant sectional curvature if and only if

$$R_{\mu\nu\alpha\beta} = \frac{R}{n(n-1)}(g_{\mu\alpha}g_{\nu\beta} - g_{\mu\beta}g_{\nu\alpha}).$$

Hence, the Ricci tensor takes the form

$$R_{\mu\nu} = \frac{R}{n}g_{\mu\nu}$$

meaning that every space of constant sectional curvature is an Einstein space. The question we ask ourselves now is whether certain properties of spaces with bounded sectional curvature from below cannot bounded from above by model spaces of constant sectional curvature. The model spaces I am talking about are the so-called maximally symmetric spaces where a symmetry is represented by a diffeomotphism $\psi : \mathcal{M} \to \mathcal{M}$ such that $\psi^\star g = g$. In case of a one parameter family ψ_t of diffeomorphisms, this leads to

$$\mathcal{L}_{\mathbf{V}}g = 0$$

where $\mathbf{V}$ has been defined as before. The reader may verify that this equation may be rewritten as

$$\nabla_{(\alpha} V_{\beta)} = 0$$

where $V_\beta = g_{\beta\alpha} V^\alpha$ and the brackets stand for symmetrization; that is

$$Z_{(\alpha_1 \ldots \alpha_n)} = \frac{1}{n!} \sum_{\sigma \in S_n} Z_{\alpha_{\sigma(1)} \ldots \alpha_{\sigma(n)}}.$$

This equation is known as Killing's equation and it is easily seen that there are $\frac{n(n+1)}{2}$ of them leaving for at most $\frac{n(n-1)}{2} + n$ free parameters where the n originates from the number of free parameters available in a point; therefore, a space with $\frac{n(n+1)}{2}$ linearly independent Killing fields is called maximally symmetric. From the Killing equation and the definition of the Riemann tensor, it follows that

$$\nabla_\alpha \nabla_\beta V^\alpha = R_{\alpha\beta\gamma}{}^{\alpha} V^\gamma.$$

More generally, from

$$\nabla_\alpha \nabla_\beta V_\gamma - \nabla_\beta \nabla_\alpha V_\gamma = -R_{\alpha\beta\gamma}^{\ \ \ \ \kappa} V_\kappa$$

it follows that

$$\nabla_\alpha \nabla_\beta V_\gamma + \nabla_\beta \nabla_\gamma V_\alpha = -R_{\alpha\beta\gamma}^{\ \ \ \ \kappa} V_\kappa$$

and using the symmetries of the Riemann tensor, one derives that

$$\nabla_\alpha \nabla_\beta V_\gamma = R_{\beta\gamma\alpha}^{\ \ \ \ \kappa} V_\kappa.$$

We now study the prototype *Riemannian* maximally symmetric spaces having zero, positive and negative constant sectional curvature respectively. They are called flat, spherical and hyperbolic geometries respectively. Let us start with flat space which is defined by $(\mathbb{R}^n, \delta_{\alpha\beta})$ where the Kronecker $\delta_{\alpha\beta}$ refers to the canonical coordinate system. The reader may enjoy finding out that the $\frac{n(n+1)}{2}$-dimensional symmetry group is given by

$$SO(n) \times \mathbb{R}^n$$

where $SO(n)$ is the n dimensional rotation group determined by operators O^μ_ν such that

$$O^\alpha_\mu O^\beta_\nu \delta_{\alpha\beta} = \delta_{\mu\nu}$$

and $\mathbb{R}^n$ is the n-dimensional translation group. The action on $\mathbb{R}^n$ is given by

$$((O, a)x)^\alpha = O^\alpha_\beta x^\beta + a^\alpha$$

and from this the reader can find out the group multiplication rules. $SO(n)$ is generated by the anti-Hermitian operators A, meaning $A^\dagger = -A$, with respect to the given scalar product and is therefore $\frac{n(n-1)}{2}$ dimensional. Between any two points x and y, there exists precisely one geodesic which is the usual straight-line segment connecting them. The canonical volume measure is given by $dx^1 \wedge \ldots \wedge dx^n$ which can be rewritten as

$$r^{n-1} dr \wedge \Omega_{S^{n-1}}$$

where $r = \sqrt{\sum_{i=1}^n (x^i)^2}$ and $\Omega_{S^{n-1}}$ is the canonical volume measure on the unit $n-1$ dimensional sphere which is closed but not exact. This fact can be easily deduced from the property that ∂_r is perpendicular to $S^{n-1}(r)$ and the scaling formula $d(rx, ry) = rd(x, y)$ for any $r > 0$. Therefore, balls of radius r around a point x have volume equal to

$$\frac{r^n}{n} \mathrm{Vol}(S^{n-1}).$$

Another important property which can be proven is that the sum of angles in a geodesic triangle always equals π; these two characteristics, the scaling of balls in terms of their radius as well as the property of triangles are the most important indicators for the "kind" of geometry one is dealing with. As an exercise, the reader should find appropriate coordinates on S^1 and S^2 and write out the respective induced measures. The reader immediately finds out that the Riemann tensor vanishes so that we have a space of constant sectional curvature zero.

We now come to the second model space which is that of an n dimensional sphere of radius r in $(\mathbb{R}^{n+1}, \delta_{\alpha\beta})$; the symmetry group here is $SO(n+1)$ which has the correct dimension. The distance between two points x, y is given by the shortest angle measured on the circle lying in the plane determined by the origin and x, y which is given by

$$\cos(\theta) = \frac{x^\alpha y_\alpha}{r^2}.$$

where x, y are $n+1$-dimensional vectors. It is a lot easier to determine the metric on the sphere by means of finding a n-bein for the tangent space of the sphere, the latter which we divide in two parts, one corresponding to $x^{n+1} \geq 0$ and another to $x^{n+1} \leq 0$. In both parts do we have that the normal vector is given by $\frac{x}{r}$ and therefore, on the part $x^{n+1} > 0$ an orthonormal basis is given by

$$E_i = (-x^{n+1} w_i, w_i.\vec{x})$$

where $x = (\vec{x}, x^{n+1})$, $w_i \in \mathbb{R}^n$ and

$$w_i.w_j = -\frac{(w_i.\vec{x})(w_j.\vec{x})}{(x^{n+1})^2}.$$

To understand the geometry however, not a single computation needs to be made; from the symmetry properties of the space, one understands that it is completely determined by the length parameter r; in particular, this implies that the sphere has constant sectional curvature with a constant determined by $\frac{a}{r^2}$ on dimensional grounds. A small computation in two dimensions reveals that $a = 1$ and the reader is invited to verify that

$$\int_{S^2} R\sqrt{h} = 2\pi\chi(S^2) = 4\pi$$

a result which can be shown to be true for any Riemannian metric on the two sphere. The reader is invited to prove this result as a nontrivial exercise, called the Gauss-Bonnet theorem. It is fairly easy to see that geodesics in this maximally symmetric space have a negative expansion meaning that they reconverge to a single point and this reconvergence happens proportionally to the Riemann tensor; therefore, the volume of any ball with radius r in $S^n(r_1)$ where $r_1 < r_2$ is less than the volume of a ball with the same radius in $S^n(r_2)$ and in Euclidean space. Also, this implies that the sum of angles between geodesic sides of a triangle sum up to a value larger than π.

Finally, we come to the third model which is that of hyperbolic space $\mathbb{H}^n(r)$, $r > 0$, which can be retrieved from $n+1$-dimensional Minkowski as the space

$$\mathbb{H}^n(r) = \{x | x^\alpha x^\beta \eta_{\alpha\beta} = r\}.$$

This space is maximally symmetric with as symmetry group $SO(1,n)$ which are nothing but the $n+1$-dimensional Lorentz transformations; the space has again constant sectional curvature which can be computed to be $-\frac{1}{r^2}$ in $n = 2$ and therefore geodesics have positive expansion proportional to the Riemann tensor meaning that if $r_1 < r_2$ then the volume of a ball of radius r in $\mathbb{H}^n(r_1)$ is larger than the volume of that same ball in $\mathbb{H}^n(r_2)$ and certainly larger than the Euclidean volume. The same argument shows that the sum over angles in a geodesic triangle is less than π. It can be shown that in such model spaces, the volume of a ball blows up as

$$r^n e^{(n-1)\frac{r}{r_1}}$$

and the reader is invited to find the correct formula. Likewise, in $S^n(r_1)$ the volume goes as

$$r^n \left(\cos\left(\frac{r}{r_1}\right)\right)^{n-1}$$

as the reader must verify.

As a last topic on geometry, we present a theorem which is intuitively obvious but requires some fine details to complete and we shall not treat those here. The statement is: if $(\mathcal{M}, h)$ is a n-dimensional Riemannian manifold with sectional curvature larger or equal to $R \in \mathbb{R}$ then the volume of a ball of radius r is less or equal to the volume of that same ball in $\mathbb{H}^n\left(\frac{1}{\sqrt{R}}\right)$ or $S^n\left(\frac{1}{\sqrt{R}}\right)$ depending on whether $R < 0$ or $R > 0$ respectively. Of course, there is a lot more to say about geometry but the insights above should suffice for the reader to understand the main content of this book.

3.6 Complex analysis.

The magic of the number i comes really to full fruitition when doing analysis in complex variables $z = x + iy$. One can see z as a composite of two real variables x, y which suggests one to work with $z, \overline{z}$. Likewise can one define the differentials

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

having the properties

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

and vice-versa. Hence, a function in the complex variable z, $f(z)$ can be seen as a function in two real variables satisfying

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

A complex valued function in one complex variable can be understood as a real one form on $\mathbb{R}^2$ by means of the formula

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

where it is understood that the mapping happens in a fixed coordinate system. The transformation properties of $f(z)$ under $z \to z(z')$ do not correspond to the correct transformation rules of the associated one form and therefore the mapping from complex valued functions in one complex variable to one forms on $\mathbb{R}^2$ is not canonical. This can be understood by means of the example $f(z) = z$ and $z = z'^2$; hence $f(z'^2) = (x'^2 - y'^2) + 2ix'y'$ where $\frac{\partial x}{\partial x'} = 2x'$, $\frac{\partial y}{\partial x'} = -2y'$, $\frac{\partial y}{\partial x'} = 2y'$ and $\frac{\partial y}{\partial y'} = 2x'$. Therefore,

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

where $f'(z') = f(z'^2)$. Nevertheless, denoting by $\mathbf{F}(z) = \mathrm{Re} f(z) dx - \mathrm{Im} f(z) dy$, 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. Denoting by $dz = dx + idy$, we have that

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

The imaginary part is closed by means of $d(F_x dy - F_y dx) = (\partial^{\alpha} F_{\alpha})\, dx \wedge dy = 0$. By the de-Rahm theorem, we have that on a simply connected domain $\mathcal{O}$, defined by $H_0(\mathcal{O}) = \mathbb{Z}$ and $H_1(\mathcal{O}) = \{0\}$, every closed one form Ω is exact, meaning $\Omega = d\mathbf{A}(z)$ and therefore, by Stokes theorem

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

where in the first formula, γ is a closed curve in $\mathcal{O}$ and in the second S is any surface in $\mathbb{R}^2$, not necessarily contained within $\mathcal{O}$. Therefore, the integral of f over any closed curve just depends upon the homotopy class of that curve within the domain where f is analytic, meaning

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

Moreover, if $f(z)$ is analytic, then $\frac{\partial}{\partial z} f(z)$ exists and we show now that it is again analytic. Clearly, we only have to prove that the second derivatives exist and are continuous; in that case

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

To show that, we first prove that

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

which is shown by

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

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

$$\sim \frac{\partial}{\partial a} f(a) \frac{|z - a|^2}{|z - a|^2}$$

where in the last step we have used $\frac{\partial}{\partial \bar{z}} f(z)$ which is what we needed to show. From this and our previously obtained results, it is easy to show that

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

where $S^1(a, \epsilon)$ is the circle of radius ϵ around a. Therefore, $f(a)$ can be differentiated an infinite amount of times with regard to a and all its derivatives are analytic. It is easy to see that this result can be generalized to

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

where γ is a curve in a sufficiently small neighborhood of a winding n times around a.

One of the most important properties of analytic functions is that it can be written in a so called power series expansion; that is

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

where the right hand side is finite for $|z-a| < \epsilon$. The proof is pretty easy and based upon our previously obtained formula

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

Differentiating this n-times with respect to a yields

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

from which it follows that

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

Let us now first mention Taylor's result; that is, for any analytic function, the following holds

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

The proof follows immediately from

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

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

Here, we have used that

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

for $|z| < 1$ as the reader may verify. Convergence then follows from the fact that

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

so that the series converges for

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

A complex valued function $f(z)$ is called meromorphic if and only if it is analytic except at a countable number of isolated points a_i such that $f(z)(z-a_i)^{n_i}$ is analytic with $\lim_{z \to a_i} f(z)(z-a_i)^{n_i} \neq 0$. From the forgoing, it follows then 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 is a closed curve winding around a_j n times. In case, $n_j = 1$ we call a_j a pole and $\mathrm{res}(a_j) := \lim_{z \to a_j} f(z)(z-a_j)^{n_j}$ the residue; the above formula then reads

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

where m_j is the winding number of γ around a_j. This formula is of extreme importance to calculate integrals of meromorpohic functions and shall be used a couple of times in the text. Obviously, more results regarding complex analysis exist but those are the most important ones finishing this section.

3.7 (Relative) probability theory.

The real physical world might consist out of an infinite number of degrees of freedom which makes it impossible to construct a Lebesgue measure. Nevertheless, nature has no problems in behaving in a *local* or *relative* probabilistic fashion as we see every day in the laboratory. In order to understand the issue with the standard theory, consider a pot filled with an infinite number of point like balls and ask someone to take a ball out of it; then the probability for a ball to be taken out is exactly zero. Nevertheless *some* ball is taken and we can make meaningful statements about the relative frequency between two balls to be chosen; the latter should be one after an infinite number of trials. Indeed, in the real physical world, it might be meaningless to ask the question concerning the probability that something happens but it could be more opportune to answer in a way revealing how many times more or less something occurs than something else. After all, this is the real basis of probability theory given that the latter says nothing about the number of events occurring. In quantum mechanics, this implies that the wave function should not be normalized, it might even have an infinite norm and still a consistent interpretation would exist. So, we are questioning the fact here if it is meaningful to define a probability function on a measure space to start with, better would be to specify a relational quantity behaving in an appropriate way.

As a matter of philosophy, we all like to believe that what is happening to us is in a sense unavoidable, close to the border of being deterministic; I believe this attitude to be wrong. In my experience, sample space is that large that everything which is happening is almost a pure coincidence by itself; the conspiracy being hidden in the relative amplitudes. Often, we tend to forget this as for example a medical doctor proclaims that you have only 3 months to live, he actually means given that the sun will rise 90 times you won't live for the 91 rise. The first condition is so obviously satisfied that we can take it for an absolute statement unless the sun explodes of course during these 90 days. Let $(\mathcal{M}, \Sigma)$ denote a sigma-algebra on some set $\mathcal{M}$, then consider

$$\lambda : \mathcal{D} \subset (\mathcal{M} \times \mathcal{M}, \Sigma \times \Sigma) \to \mathbb{R}_+ \cup \{\infty\}$$

where by definition $0.\infty = a$ where a is any number in $\mathbb{R}_+ \cup \{\infty\}$ and $b.\infty = \infty$ for every $b > 0$ so that the multiplication is still associative and commutative and we define the "inverse" of 0 to be ∞. Then, the inverse still satisfies the property that $(x^{-1})^{-1} = x$ and $(xy)^{-1} = y^{-1}x^{-1}$; however, we have not exactly a group structure but everything we say applies for any field. A symmetric[2] subset $\mathcal{D} \subset (\mathcal{M} \times \mathcal{M}, \Sigma \times \Sigma)$ is a symmetric subset of the sigma-algebra $\Sigma \times \Sigma$ on the same underlying space $\mathcal{M} \times \mathcal{M}$. λ defines a relative probability function if and only if

$$\begin{aligned} \lambda(A, B) &= \lambda(B, A)^{-1} \\ \lambda(\sqcup_n A_n, B) &= \sum_n \lambda(A_n, B) \\ \lambda(A, B) &\equiv \lambda(A, C)\lambda(C, B) \end{aligned}$$

where in the last sentence the equivalence means that some value of the right hand side must equal the left hand side. The union $\sqcup$ is the disjoint union meaning that the intersections $A_n \cap A_m$ are empty. The implication of this point of view is nontrivial; as is well known, it is impossible to define a Lebesgue integral on $(\mathbb{R}^\infty, \mathcal{B})$ where $\mathcal{B}$ is the Borel sigma-algebra, but it is very well possible to define a relative Lebesgue measure and therefore integral by considering those Borel sets A which have a finite relative volume with respect to B. Here, the relative measure can be defined in a weak and strong sense; the former is given with respect to an increasing sequence of subspaces $\Gamma_n = \mathbb{R}^n$ with $\Gamma_n \subset \Gamma_{n+1}$ and the inductive limit of Γ_n is $\mathbb{R}^\infty$ by means of

$$\lambda(A, B) = \lim_{n \to \infty} \frac{\mu_n(\Gamma_n \cap A)}{\mu_n(\Gamma_n \cap B)}$$

where the limit is supposed to exist. The strong definition requires the above limit to exist and to be independent of any sequence chosen. Such strong relative measures are translation invariant and in the weak case, the relative measure is invariant with respect to any finite dimensional Euclidean group associated to the Γ_n. To define the relative integral is easy; denote by $\mathcal{B}(A)$ the set of all C such that $(C, A) \in \mathcal{D}$, then $\mathcal{B}(A)$ is not necessarily a Borel-Sigma algebra but it has all its salient features since it is a subset of Σ and the relative measure just reduces to an ordinary one. This suggests one to simply take over the definition of the standard Lebesgue integral with respect to $\mathcal{B}(A)$ to the infinite dimensional case. This calls for a point of attention though which is that decimating

[2]Symmetric means symmetric with respect to the (x, y) interchange in $\mathcal{M} \times \mathcal{M}$.

an infinite dimensional cube in every direction produces an $\aleph_1$ number of cubes. Clearly, this is not what we are doing and also the domain of integration is relative to A so that only a sub cover with an $\aleph_0$ number of elements with finite relative volume are used. Hence, it appears that every theorem for Lebesgue theory generalizes in a way to relative measures; this work can henceforth also serve as a basis to rigorously define relative path integrals and it might be that one can detach oneself from limiting procedures.

The content of this chapter should enable the reader to follow the arguments in the main body of this book and understand those issues necessary to fill in the remaining gaps in his or her knowledge of the mathematical language. One topic I have not touched upon and which I shall use throughout this text is the theory of distributions and in particular the δ-Dirac function. I have omitted this topic because I feel the reader may easily learn about this from Wikipedia or other resources available on the internet or any good book for that matter.

Chapter 4

The need for general covariance.

Historically, quantum mechanics was born around the same time as was general relativity and most physicists at that time had only very elementary notions about geometry. Therefore, it does not come as a surprise that both theories have been formulated in an incompatible language meaning it is quantum theory which falls short here. By this, I certainly do not intend to say that the ideas of Heisenberg and Bohr regarding the nature of reality were incorrect as the reader should have noticed by reading the introduction as well as chapter two. On the contrary, I believe and will partially show in the subsequent chapters that their views need to be extended to situations of a more general nature; this was also the very message I intended to convey in chapter two regarding my ideas concerning measurement and the according change in definition of the kinematical structure of the theory. What I disagree with, however, is that the definition of a particle cannot be made in an objective way and I deem this attitude to stem from a wrong formulation of the theory. More precisely, it is the lack of general covariance of the theory or, in other words, the non-geometrical formulation theirof which is responsible for this bullshit. On the other hand, I perfectly agree that observations are observer dependent and that there exists no general notion of objective truth in the sense that everyone reads the objective wavefunction in a slightly different way. As I have speculated in chapter two, this process will require drastic changes to all known formulations of physics and the current formulation of quantum mechanics is inadequate to address these issues in a proper way. So, I am a real Copenhager and I hope to have explained in a proper way in the introduction why this philosophy is very clever and not defaitist as one might think at first; it is however not as complete as I would wish and I hope to have contributed in chapter two some valuable ideas towards its extension.

Therefore, by an objective quantum theory, I mean objective particle notions and nothing else. To some people I know, it is obvious that quantum mechanics, as formulated between $1920 - 1950$ is a theory intended to work only in flat spacetime. Others have more difficulty with this and it is for them that I have written this chapter. For me, this issue was clear since I started to work on my PhD eighteen years ago since that was the very first time I encountered general relativity and I have always maintained the point of view that we were not ready yet to formulate the proper dynamical laws without changing the formulation of quantum mechanics. We shall address the issue of covariance from as many angles as possible so that even the most skeptic reader will have to rest in defeat. So, I am of the opinion that Einstein's problems with quantum mechanics were two-fold: on one side, it was a problem of language as the very formulation theirof was limited to the flat geometry of Newton or Minkowski and that must have been the reason why Einstein was seduced by first quantized covariant equations such as the Klein-Gordon equation. On the other hand, there was also something wrong with Einstein's view on spacetime, which is grounded in a philosophy of eternalism and not one of processes; we have sharply condemned this view in chapter two. When going over to the process view, the collapse of the wavefunction is a rather natural axiom and certainly not in conflict with relativity whatsoever.

4.1 On some foundational issues of the old quantum theory.

In this section, we start with a rather personal account on the foundations of quantum mechanics which is written in a way inspired by the comments in the previous chapter; I do not claim that all details have been sufficiently covered, but there are for sure more "foundational comments" in here than in all textbooks I have encountered so far. Dirac was the first person to reconcile Heisenberg's and Schrodinger's quantum theory from the point of view of the Poisson bracket $\{f, g\}$ evaluated on functions f, g of the phase space coordinates q_α; a mathematical gadget used in classical physics to write down the so called Hamiltonian equations of motion. The Poisson bracket has the following properties:

$$\begin{aligned}
\{f, g\} &= -\{g, f\} \\
\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} &= 0 \\
\{q_\alpha, q_\beta\} &= \Omega_{\alpha\beta} \\
\{fg, h\} &= \{f, h\}g + f\{g, h\} \\
\{af + bg, h\} &= a\{f, h\} + b\{g, h\}
\end{aligned}$$

where $\Omega_{\alpha\beta}$ is the non-degenerate symplectic form and $a, b \in \mathbb{R}$. Now, Dirac was thinking about the procedure of quantization by replacing this "algebra" by means of a "quantum algebra" which is defined from the latter by replacing the third relation by

$$\{q_\alpha, q_\beta\} = i\hbar\Omega_{\alpha\beta}1$$

where the unit has to be interpreted as another generator of the algebra which commutes with everything

$$\{f, 1\} = 0$$

which follows from the fourth condition. We will come back to this definition in a while; at the time quantum mechanics was born, people were convinced that the essential part of the dynamics, one in which a single quantum particle was essentially free, must be linear in terms of the potentialities *and* that probabilities must be expressed in terms of the modulus squared of the potentialities. That is, the essential quantity, which is the wave function, which we associated to words and their potentialities before, was observed to undergo a quasi-linear dynamics, a most important feature indeed. Since it appeared natural to postulate a first order, time irreversible dynamics for the wave function, one needed an equation of the form

$$i\dot{\Psi} = H \triangleright \Psi$$

where Ψ is vector in a complex vector space V and H a linear operator (the i is chosen out of convention here since we did not impose any properties on H yet) and $\triangleright$ denotes the action of H on Ψ. The action of a linear operator on a state satisfies

$$H \triangleright (a\Psi + b\Phi) = a(H \triangleright \Psi) + b(H \triangleright \Phi)$$

and the trick now is that the action defines an operator multiplication "." by

$$(X.Y) \triangleright \Psi = X \triangleright (Y \triangleright \Psi)$$

which is associative by definition since there is only one way to read successive actions. With respect to this product and for time independent H, one can *formally* integrate this equation and obtain that

$$\Psi(t) = e^{-iHt} \triangleright \Psi(0)$$

where we have extended our definition of an action to

$$(aX + bY) \triangleright \Psi = a(X \triangleright \Psi) + b(Y \triangleright \Psi).$$

The evolution operator $U(t) = e^{-iHt}$ for time dependent $H(t)$ reads

$$U(t,s) = \lim_{n\to\infty} (1 - i\delta_n H(s + (n-1)\delta_n)) (1 - i\delta_n H(s + (n-2)\delta_n)) \ldots (1 - i\delta_n H(s + \delta_n)) (1 - i\delta_n H(s))$$

where $\delta_n = \frac{t-s}{n}$ for $t > s$. Note that this is all formal in the sense that H can have a nontrivial domain $\mathcal{D}_H \subset V$ and it is by far not necessary that $H \triangleright \mathcal{D}_H \subset \mathcal{D}_H$ so that the composition is only well defined on $\mathcal{D}_{H^2} \subset \mathcal{D}_H$. For example, if H is a second order partial differential operator, then for the expansion to be well defined, it is necessary (but not sufficient) that Ψ can be differentiated an infinite number of times, while it is very well known that a unique strong solution exists for initial data which are twice differentiable. Indeed, it should be well known that the definition of e^{-iHt} is *not* given by

$$\sum_{n=0}^{\infty} \frac{(-itH)^n}{n!}$$

or by

$$\lim_{n\to\infty}\left(1-\frac{iHt}{n}\right)^n$$

since these expressions produce infinities at every order, but one has that

$$U(t)=\int e^{i\lambda t}dP_\lambda$$

where

$$\int \lambda dP_\lambda = H$$

is the spectral decomposition of H and the equality is to be interpreted in a weak sense. A good definition, in case $H(t)$ is time dependent, is given by

$$U(t,s)=\lim_{n\to\infty}U_{s+(n-1)\delta_n}(\delta_n)U_{s+(n-2)\delta_n}(\delta_n)\ldots U_{s+\delta_n}(\delta_n)U_s(\delta_n)$$

where the limit is understood in the weak sense and $U_r(\delta)=\int e^{i\delta\lambda}dP^r_\lambda$ where $H(r)=\int \lambda dP^r_\lambda$. Note that, at this point, $H(t)$ can be any operator whatsoever and does not need to be connected to the "quantization" of a classical Hamiltonian; a second ingredient is needed for the interpretation of Ψ. In fact, we did already encounter such ingredient, which was the existence of a spectral decomposition to *explicitly* integrate the flow, but why should we let ourselves be guided by such criterion to construct a physical theory. Most classical theories have no explicit formulae for the time flow, so why care? Moreover, why should it be that time dependent Hamiltonians are all self-adjoint on the *same* Hilbert space? In classical physics, one is not worried about fall-off criteria of the geometry towards spatial infinity, but in quantum physics one definitely is. To be more specific, the very definition of H is tied to the Hilbert space one chooses. Traditionally, to make sense out of the integration of the time flow in the way we did before, one posits the existence of a *time independent* scalar product $\langle | \rangle$ which defines a Hilbert space such that "time evolution" preserves this scalar product; that is

$$\langle U(t,s)\Psi|U(t,s)\Phi\rangle=\langle\Psi|\Phi\rangle$$

which is equivalent to

$$U(t,s)^\dagger U(t,s)=1$$

where 1 is the identity operator on Hilbert space. Strictly speaking, this is the condition which needs to be satisfied for a partial isometry, but the founding fathers went beyond that and also required

$$U(t,s)U(t,s)^\dagger=1$$

which is necessary to make $H(t)$ self adjoint and $U(t,s)$ well defined in the first place starting from our equation of motion[1]. Indeed, having a spectral decomposition with real eigenvalues is equivalent to the operator being self adjoint

[1]One could could start from the weaker condition that $U(t,s)$ defines a map on rays satisfying $|\langle U(t,s)\Psi|U(t,s)\Phi\rangle|^2=|\langle\Psi|\Phi\rangle|^2$ to deduce that $U(t,s)$ can be chosen to be linear and satisfy $U(t,s)^\dagger U(t,s)=1$.

and one could contemplate so called normal operators with a complex spectrum but non-unitary evolution. These fine points concerning the very definition of a time evolution operator are constantly ignored when making perturbative calculations in Quantum Field Theory; indeed, there one assumes the formal definition of the exponential operator which is ill defined and it should not come as a surprise that infinities arise in the calculation.

It is often said that Schrodinger needed both the insight of the linearity of the time evolution in terms of the potentialities and the fact that $|\Psi(x)|^2 dx$ is proportional to the probability for an "event" $\left[x - \frac{1}{2}dx, x + \frac{1}{2}dx\right]$ to happen to arrive at quantum theory. Let us ask ourselves what rules one might posit based upon the demand of linearity only. For sure, we cannot derive the complex numbers out of this as we know quantum theory can be consistently defined for real numbers and quaternions as well. A rule related to a linear classical stochastic theory supplied by the demand of conservation of probability would result for example in the following mathematical framework: we demand that $\Psi(t, x) \geq 0$ and a linear functional ω_t to exist such that $\omega_t(\Psi(t, x)) = 1$. Differentiating this with respect to time results then in

$$\dot{\omega}_t(\Psi) + \omega_t(H'\Psi) = 0$$

where $H' = iH$ since we do not want to impose the complex numbers yet. In case ω is t independent, this results by continuity in the fact that H' must map V into the kernel of ω and therefore 0 is an eigenvalue (of the discrete or residual type). In a matrix language, such feature is for example realized if and only if $\sum_i H'^i_j = 0$ for all j and with $\omega(\Psi) = \sum_i \Psi^i$. Suppose we would make a change of basis $\Psi \to O\Psi$, then this operator needs to satisfy that $\sum_i O^i_j = 1$, a condition which is consistent with the matrix product OV since

$$\sum_{i,j} O^i_j V^j_k = 1$$

for all k. Moreover, the identity transformation 1 and the inverse of O also constitute valid transformations so that any such theory has an $n(n-1)$ dimensional transformation group if V is n dimensional. However, there is an additional condition here as the evolution H' and transformations O still need to preserve the condition that $\Psi^i \geq 0$; in general this will only hold if and only if all $O^i_j, H'^i_j \geq 0$, which makes the group rather small but it still contains the permutation group which should suffice to account for the Galilean transformations in Newtonian physics. One might be tempted to generalize this and drop the condition that $\Psi(t, x) \geq 0$ which means they cannot represent probabilities anymore. Fine, so why not take, for example, $\frac{|\Psi^i|}{\sum_j |\Psi^j|}$ as a measure for the probability since that would settle the matter. Indeed, our transformation group would constitute the entire special linear group $SL_n(\mathbb{R})$ and the Hamiltonian H' is completely arbitrary in this framework. Therefore, having a large enough symmetry group cannot be a criterion for quantum mechanics to emerge.

Before we raise any further objections, let us come back to the Heisenberg picture and how Dirac reconciled both formulations by quantization of a classical theory. The Heisenberg picture is usually presented for a time independent Hamiltonian generating a one parameter group of time translations $U(t)$; in the Schrodinger picture, a self adjoint operator O^S_t representing a physical observable (at time t) is kept fixed and the quantities corresponding to real measurements are of the form

$$\langle \Psi_t | O^S_t \Phi_t \rangle$$

which is the same as

$$\langle \Psi_s | U(t,s)^\dagger O^S_t U(t,s) \Phi_s \rangle = \langle \Psi_s | e^{iH(t-s)} O^S_t e^{-iH(t-s)} \Phi_s \rangle.$$

Hence, it is said that measurement of O^S_t on Ψ_t results in $P_\lambda e^{-iH(t-s)}\Psi_s$ at time t in the Schrodinger picture, while it results in $e^{iH(t-s)} P_\lambda e^{-iH(t-s)} \Psi_s$ at time s in the Heisenberg picture. This is a consistent view, since a second later measurement at r in the Schrodinger picture results in $Q_\mu e^{-iH(r-t)} P_\lambda e^{-iH(t-s)} \Psi_s$ while the Heisenberg view produces $e^{-iH(r-s)} Q_\mu e^{-iH(r-s)} e^{iH(t-s)} P_\lambda e^{-iH(t-s)} \Psi_s = e^{-iH(r-s)} Q_\mu e^{-iH(r-t)} P_\lambda e^{-iH(t-s)} \Psi_s$ all of which produce the same probabilities. Taking the differential of $O^H(s,t) \equiv e^{iH(t-s)} O^S_t e^{-iH(t-s)}$ with respect to t implies that

$$\dot{O}(s,t) = i \left[H, O^H(s,t) \right]$$

where the bracket is the commutator, that is $[A,B] = AB - BA$. This is all well known and accepted as standard material; but things become somewhat more complicated if we take the *Schrodinger* Hamiltonian H^S to be time dependent. Indeed, so far, we obtained the result that the Hamiltonian in the Schrodinger picture equals the Hamiltonian in the Heisenberg picture albeit the latter should depend upon two times and not just a single one as is the case for the Schrodinger Hamiltonian. Since by definition,

$$O^H(s,t) = U(t,s)^\dagger O^S_t U(t,s)$$

and its differential to time t equals, since $\dot{U}(t,s) = -iH^S(t)U(t,s)$,

$$\dot{O}(s,t) = i \left(U(t,s)^\dagger H^S(t) O^S_t U(t,s) - U(t,s)^\dagger O^S_t H^S(t) U(t,s) \right) \neq i \left[H^S(t), O^H(s,t) \right]$$

so that the famous Heisenberg equation appears not to be amenable to time dependent Hamiltonians. One may guess now that it is more natural to derive the Heisenberg operators with respect to the reference time s arriving at

$$\frac{d}{ds} O^H(s,t) = -i \left[H^S(s), O^H(s,t) \right]$$

where one should keep in mind that $t \geq s$ and the final condition $O^H(t,t) = O^S_t$ that is, if the actual and reference time coincide, the Schrodinger operator equals the Heisenberg operator. Note also the relative minus sign to our previous expression which of course came from switching actual with reference time. This is however not the correct way to go and one should define the Heisenberg Hamiltonian as

$$H^H(s,t) = U(t,s)^\dagger H^S(t) U(t,s)$$

and the latter satisfies

$$\dot{H}^H(s,t) = U(t,s)^\dagger \partial_t H^S(t) U(t,s)$$

and any general observable with explicit time dependence obeys likewise

$$\dot{O}^H(s,t) = i\left[H^H(s,t), O^H(s,t)\right] + U(t,s)^\dagger \partial_t O^S(t) U(t,s).$$

The crucial distinction between the Heisenberg and the Schrodinger Hamiltonian is that, albeit they constitute precisely the same expressions in terms of the canonical variables, the latter come in terms of the Heisenberg and Schrodinger operators respectively. For time independent Hamiltonians, this distinction does not matter and gives the same result, while for time dependent Hamiltonians it does. Note that therefore, the Schrodinger picture is the easiest to start with as it gives direct formulae for all observables, while the Heisenberg picture can be somewhat more complicated for time dependent systems due to its dependence on time dependent canonical variables.

So, Dirac recognized the formal equivalence of the structure of the Heisenberg equation and the Poisson bracket structure of classical mechanics even though at that point, nobody should ever have mentioned the word quantization. All we did so far is to deduce these structures from the Schrodinger equation which had a direct ground in experiment; nothing so far was said about some magical trick between the classical line of thought and quantum framework. Another issue shows up if we really take the Dirac programme seriously; that is, there is no a priori reason why the replacement of the Poisson bracket should have anything to do with the commutator defined by the product inherited from the action of operators on vectors $\triangleright$. Obviously, it should be like that if we want the Heisenberg and Schrodinger picture to be equivalent, but there is no a priori reason for it from the point of view of Dirac. Let me illustrate this by means of an example; as before, consider a Hilbert space $(\mathcal{H}, \langle | \rangle)$ with an action $\triangleright$ of operators on vectors and its associated product ".". Consider A to be a positive definite operator and define the product $\star$ by $X \star Y = X.A.Y$ where from now on, we will drop all the dots. This product is associative and has a unit element A^{-1} which we will interpret as the unit appearing in the Dirac Quantization programme. Consider now a time independent Hamiltonian H which will serve to build a Heisenberg dynamics

$$\frac{d}{dt} O(t) = i(H \star O(t) - O(t) \star H) = i\left[H, O(t)\right]_\star$$

where we have dropped the reference time s. If our Hamiltonian arises from the quantization of a classical Hamiltonian, we shall impose

$$\left[q_\alpha, q_\beta\right]_\star = i\hbar \Omega_{\alpha\beta} A^{-1}.$$

This gives

$$q_\alpha = A^{-\frac{1}{2}} \hat{q}_\alpha A^{-\frac{1}{2}}$$

where

$$[\hat{q}_\alpha, \hat{q}_\beta] = i\hbar \Omega_{\alpha\beta} 1.$$

Hence, our Hamiltonian $H(q_\alpha, \star) = A^{-\frac{1}{2}} H(\hat{q}_\alpha, .) A^{-\frac{1}{2}}$ and therefore, the evolution operator $U(t)$, such that

$$q_\alpha(t) = U(t)^\dagger \star q_\alpha(0) \star U(t)$$

is given by $A^{-\frac{1}{2}} \hat{U}(t) A^{-\frac{1}{2}}$ where

$$\hat{U}(t) = e^{-iH(\hat{q}_\alpha)t}$$

is the evolution operator with respect to the standard product. In other words,

$$q_\alpha(t) = A^{-\frac{1}{2}} \hat{q}_\alpha(t) A^{-\frac{1}{2}}$$

and for the probability interpretation one just needs the spectral decomposition of this operator which is unrelated to the spectral decomposition of $\hat{q}_\alpha(t)$. For the Schrodinger picture, it would be mandatory to take the spectral decomposition of $q_\alpha(0)$ and apply it to $A^{\frac{1}{2}} \hat{U}(t) A^{-\frac{1}{2}} \Psi$ which gives a totally different result! This would immediately be repaired if one adjusted the action $\triangleright$ to the new product $\star$ but the point of this argument was to show that they did not need to be equal to one and another. Therefore, Dirac, crucially, had to depend upon this piece of information to maintain equivalence between both pictures. Note that the new Schrodinger evolution operator $T_t = A^{\frac{1}{2}} \hat{U}(t) A^{-\frac{1}{2}}$ is unitary with respect to scalar product

$$\langle \Psi | A^{-1} \Phi \rangle$$

but the operators q_α nor $\hat{q}_\alpha$ are Hermitian with respect to this product. Note here that, in the derivation of our argument, we have disentangled the meaning of what it is to be an identity; for us, we just defined it as an operator which commutes with everything while in the standard interpretation it also means acting as the identity on vector states. We will continue to do this, even to a further extend, in our comments upon quantum field theory. As a final question, one may wonder what A should depend upon; a natural suggestion would be that it behaves invariantly under coordinate transformations so that it must be some invariant of the spatial metric. This is how geometry can creep into the foundations of quantum theory and destroy the equivalence between the Heisenberg and Schrodinger picture.

Let us list the points we mentioned before:

- Fine, so you have a quasi linear dynamics regarding the potentialities for isolated particles, but we don't observe anything like this in the macroworld. Effectively, for big objects, we can forget about potentialities all together. Is it possible at all for such limit to emerge from a fundamentally quantum system?

- Why a time independent Hilbert space? We agree one needs a measure on the space of all potentialities but the spacetime structure might dictate this measure to evolve and not remain static. Surely, it is tempting to let oneself be persuaded by a powerful tool such as a spectral decomposition but why should a time evolution operator posses such a thing? Why should we be able to write everything in terms of stationary states if the universe is evolving irreversibly?

- Why should we take Dirac's programme seriously? Quantum theory, as formulated above, is entirely motivated by the linearity of the time evolution and has no a priori grounding in a Poisson structure. Actually, if we were not able to integrate the time flow in the way we did (based on the demand that the Hamiltonian corresponds to a Hermitian operator), we might not have spoken about a Heisenberg picture in the first place since time evolution might not map self adjoint operators to self adjoint operators. In any case, our line of argumentation shows somehow that the Schrodinger picture allows one to ask more foundational questions about quantum mechanics than the Heisenberg picture does.

- This last conclusion is only enforced by looking again at the evidence; the main observation was that the dynamics for the potentialities of a single isolated particle was more or less linear, not that it was deterministic! In that sense, Schrodinger might have already overstretched himself by writing down the ordinary differential operator $\frac{d}{dt}$ and could have instead resorted to a stochastic operator (which, admittedly, did not exist yet at his time) in which the wave potentialities themselves become stochastic variables. In our language this would mean that he might import the $\vee$ operation in quantum mechanics! This would not sound very strange at all given that the measurement process is nonlinear and stochastic; but this would most likely completely destroy the Heisenberg picture and Dirac's programme.

Formal analogy may often be a good guideline but ultimately physical arguments carry more power; as I will argue later on, when we want to dismiss the notion of time, the Schrodinger equation needs some revision too in the lines argued above. Let me proceed by giving an example which has time and time again be discussed in history regarding the fundamentally linear character of quantum mechanics; it wasn't taken as an axiom before such as bachelor's are spooned today, but one wondered whether it only pertained to single particle systems or also to systems containing more particles. Let me provide an example where one did not take it for granted that linearity would just extend as usual (by effectively describing the classical join of two quantum particles or equivalently, the union of them). Let particle i be represented by a wavefunction Ψ_i which will conveniently depend upon the spatial coordinate $\vec{x}_i$ and no other properties of particles are assumed; then, one can write down

$$i\dot{\Psi}_i(\vec{x}_i) = H_i\Psi_i(\vec{x}_i) + \sum_j \left(\int d\vec{x}_j |\Psi_j(\vec{x}_j)|^2 A_{ji}(|\vec{x}_j - \vec{x}_i|) \right) \Psi_i(\vec{x}_i)$$

where H_i is the "single particle Hamiltonian" for particle i and A_{ji} is some real valued function depending upon the distance between particles i and j. Such models have been examined in the literature in the past as one did not immediately want to resort to a quantum join of quantum particles; that concept was radically new at the time and it still is in some sense. So far, only the suggestion of a time dependent scalar product is to my knowledge really new (and we will examine some of its consequences shortly); self adjoint observables have been given up before by some people (not by many) and modified Schrodinger equations with stochastic noise have been examined in the context of the measurement postulate.

4.2 Non-covariance of the Schrodinger formalism and Heisenberg equations.

By covariance, I mean that the choice of coordinates should not matter in the definition of a physical law; this principle can pertain to the spatial coordinates only or to the spacetime coordinates all together. The question we shall address here is if this simple principle does not already call for a revision of the Dirac programme; regarding the Schrodinger picture, there is nothing wrong a-priori and one can just say by fiat that the Hamiltonian H enjoys this property. But perhaps, working through the Dirac programme might lead to additional insights which could reflect on the Schrodinger picture too *depending upon the questions you are going to ask*. Indeed, it is this very last sentence which is important and we shall see that the Schrodinger picture allows for more economic quantum theories if one does not ask about the momentum operator for example. Concerning momentum operators, it is well known one can define an inequivalent number of them; for example $i\partial_x$ and $i\partial_x + x^2$ the latter having different domain properties than $i\partial_x$ and therefore defining an inequivalent representation of the Heisenberg algebra. Indeed, it is well known that the Stone - Von Neumann theorem only applies to the exponentiated version of the Heisenberg position and momentum operators, the so called Weyl elements. From a modern point of view, one could say that the second operator should be excluded since it does not transform covariantly under $x \to y(x)$, but in the standard view on the Schrodinger picture, this transformation does not make any sense as the scalar product in three dimensions transforms as

$$\int d^3x \overline{\Psi(x)} \Phi(x) \Rightarrow \int d^3x' |\det\left(\frac{\partial x}{\partial x'}\right)| \overline{\Psi(x(x'))} \Phi(x(x'))$$

if $\Psi(x)$ transforms as $\Psi'(x') = \Psi(x(x'))$. One could cure this by letting $\Psi(x)$ be a density of factor $\frac{1}{2}$, meaning that

$$\Psi'(x') = |\det\left(\frac{\partial x'}{\partial x}\right)|^{-\frac{1}{2}} \Psi(x(x')).$$

In that case, the integral

$$\int d^3x \overline{\Psi(x)}\Phi(x)$$

would remain invariant but the expectation value of the momentum operator would transform as

$$\int d^3x' \overline{\Psi'(x')} i\partial'_j \Phi'(x') = \int d^3x |\det\left(\frac{\partial x'}{\partial x}\right)|^{\frac{1}{2}} \overline{\Psi(x(x'))} \frac{\partial x^k}{\partial x'^j} \partial_k \left(|\det\left(\frac{\partial x'}{\partial x}\right)|^{-\frac{1}{2}} \Phi(x(x'))\right)$$

which is fine in the sense that it is still a Hermitian operator. The problem however is that $i\partial_x \Phi(x)$ does not transform nicely under coordinate transformations anymore and it certainly does not transform as a $\frac{1}{2}$ density so that one has operators mapping densities of factor $\frac{1}{2}$ to something with no suitable covariance properties at all; this is the reason why the above formula does not make sense. Let us first deal with this fact and then point out a shortcoming in the Dirac programme; there are some real lessons to be learned here. Suppose $\Psi(t,x)$ is a spatial density of factor $\frac{1}{2}$, then the natural definition for a differential operator is given by

$$\partial_j^N \Psi(t,x) = h^{\frac{1}{4}} \partial_j \left(\frac{\Psi(t,x)}{h^{\frac{1}{4}}(t,x)}\right) = \partial_j \Psi(t,x) - \frac{1}{2} \frac{\partial_j h^{\frac{1}{2}}(t,x)}{h^{\frac{1}{2}}(t,x)} \Psi(t,x)$$

where $h(t,x)$ is the determinant of the spatial metric. Obviously, this definition can be extended to any density of factor r by simply replacing the factor $\frac{1}{2}$ in the last expression by r. Hence, one can generalize the usual covariant derivative ∇_j to ∇_j^N and interpret $-i\hbar\nabla_j^N$ as the natural momentum derivative; the latter however is *not* a symmetric operator since

$$\int dx \, \overline{i\partial_j^N \Psi(t,x)} \, \Phi(t,x) - \int dx \, \overline{\Psi(t,x)} \, i\partial_j^N \Phi(t,x) = \int dx \, \overline{\Psi(t,x)} \Phi(t,x) \, i \frac{\partial_j h^{\frac{1}{2}}(t,x)}{h^{\frac{1}{2}}(t,x)}.$$

Since a momentum operator applied to any tensorvalued density must transform as the same object with one covariant index more, this choice of momentum operator is unique up to first and second derivatives of the metric tensor. That is, any other vector field added, constructed from the geometry alone must contain at least third order derivatives of the metric such as $\partial_j R(t,x)$ where $R(t,x)$ is the Ricci scalar. This means that we are obliged to recognize that not all physical observables are Hermitian and in particular the momentum observables and Hamiltonian are not; a similar result follows from keeping Ψ as an amplitude (scalar under coordinate transformations) and choosing the usual $i\partial_j$ as momentum operator, but this time with time dependent scalar product

$$\int dx \, \overline{\Psi(t,x)} \Phi(t,x) h^{\frac{1}{2}}(t,x).$$

It is in this sense that gravity is very different from gauge theories since its coupling constant is imaginary whereas for gauge theories it is real; this mere fact has serious consequences for the foundations of quantum mechanics as we shall investigate further. Indeed, there are other consequences to be learned: *if* one takes it seriously that the momentum operator should be defined in the theory and that the Hamiltonian is constructed from the momentum operator, then one must conclude that the correct momentum operator is given by $i\hbar\nabla^N_j$ or $i\hbar\nabla_j$ depending upon whether we consider the wavefunction to be a density of factor $\frac{1}{2}$ or just a scalar. Let us first explain the latter case, since it is closer to what we know about quantum mechanics: suppose, for the contrarian viewpoint, that the classical Hamiltonian is given by

$$H = h^{jk} p_k p_j$$

then[2] substituting $p_j = -i\hbar\partial_j$ results in a quantum Hamiltonian of the kind

$$H = -\alpha h^{jk}\partial_j\partial_k - \beta\left(\partial_j h^{jk}\right)\partial_k - \gamma\partial^2_{jk}h^{jk}$$

which means it is *impossible* for the covariant expression

$$h^{jk}\nabla_j\partial_k$$

to be found since that one equals

$$-\hbar^2 h^{jk}\partial_j\partial_k + \frac{\hbar^2}{2}h^{jk}h^{lr}\left(\partial_j h_{rk} + \partial_k h_{jr} - \partial_r h_{jk}\right)\partial_l$$

which can be rewritten as

$$-\hbar^2 h^{jk}\partial_j\partial_k - \hbar^2\left(\partial_j h^{jk}\right)\partial_k - \frac{\hbar^2}{2}h^{lr}h^{jk}\partial_r h_{jk}\partial_l$$

and it is this last term which is missing in H. This means that we have to regard $-i\hbar\nabla_j\Psi$ as a one form and therefore Hilbert space itself should be extended to all covariant n tensors. That is, the underlying vector space is given by $T_\infty\mathcal{M} = \oplus_{n=0}^{\infty}T_{n,c}\mathcal{M}$ where $T_{n,c}\mathcal{M}$ stands for the vectorspace of complex tensors with n covariant indices. A scalar product can be defined by

$$\sum_{n,m\geq 0}\int dx\overline{\Psi}_{a_1\ldots a_n}T^{a_1\ldots a_n b_1\ldots b_m}\Phi_{b_1\ldots b_m}h^{\frac{1}{2}}$$

where $T^{a_1\ldots a_n b_1\ldots b_m}$ is a tensor constructed from the spatial metric h^{jk} alone such that the total expression is positive definite. Now, it is obvious that in case

[2]This Hamiltonian can be derived from the action principle

$$S = \int dt h_{\alpha\beta}(t, x(t))\frac{dx^\alpha(t)}{dt}\frac{dx^\beta(t)}{dt}$$

or in reparametrization invariant form

$$S = \int d\tau \frac{h_{\alpha\beta}(t,x)\dot{x}^\alpha\dot{x}^\beta}{\dot{t}}$$

where $\dot{t}$ denotes the derivative of t with respect to τ.

no coupling exists between the different n as well dynamically as kinematically that everything reduces to our previous setting with a time dependent scalar product and non-Hermitian momentum operators. Note that now, it becomes possible to define self adjoint momentum operators with respect to *physical* directions n in space; the relevant operators being given by

$$n.p = -i\hbar n^{\mu}\nabla_{\mu} - \frac{i\hbar}{2}\left(\nabla_{\mu}n^{\mu}\right).$$

The reader may check that those are indeed symmetric and densely defined with respect to the scalar product defined by $T^{a_1\ldots a_n b_1\ldots b_m} = \delta_{nm}\prod_j h^{a_j b_j}$. It is of crucial importance to notice that

$$\left[-i\hbar\nabla_{\mu}, -i\hbar\nabla_{\nu}\right]\Psi_{\kappa} = -\hbar^2 R_{\mu\nu\kappa}{}^{\alpha}\Psi_{\alpha}$$

and likewise so for higher order covariant tensors. This means that the Heisenberg relation is only *at best* valid on the scalar sector; indeed, nothing could have stopped us from defining the momentum operators as

$$-i\hbar e^{-\alpha R}\nabla_{\mu}$$

in the first place. All of this makes it much more difficult to integrate the time flow as no spectral theorem applies and very different techniques will have to be developed. Note that this entire setting has nothing to do with curved space (time) but follows from the mere demand of covariance and lifting the limitation of linear transformations between inertial systems. The Poisson bracket is generally covariant in the sense that coordinate transformations do constitute symplectic transformations as the reader may easily verify; bringing this covariance to the quantum sector requires one to reinterpret the meaning of the right hand side of the Poisson brackets as well as to consider quantum corrections on the classical dynamical laws concerning the so called observables of the theory. Indeed, for our above classical Hamiltonian, one obtains that the classical equations of motion are

$$\dot{x}^j = 2h^{jk}p_k, \; \dot{p}_j = -\partial_j h^{kl}p_k p_l$$

the second of which the left nor the right hand transform covariantly since the time derivative does not commute with $\frac{\partial y^k(t)}{\partial x^j(t)}$. However, both non-covariant terms are equal so that

$$\dot{p}_j = -\partial_j h^{kl}p_k p_l$$

is a basis independent statement. In our *covariant* quantum theory however, the second equation

$$\frac{Dp_j}{dt} = i\left[H, -i\nabla_j\right]$$

is manifestly covariant on both sides - at the reference time where the Schrodinger and Heisenberg picture coincide - and gives

$$\frac{Dp_j}{dt}\Psi = -\left[\hbar^2 h^{kl}\nabla_k\nabla_l, \nabla_j\right]\Psi = \hbar^2 h^{kl}R_{jkl}{}^{s}\partial_s\Psi$$

and more complicated expressions for higher order tensors. Hence, the new expression only depends upon the first order derivatives of Ψ whereas, in the old framework, this would have been the second order derivatives due to the noncovariant term $\partial_j h^{kl}$ which now vanishes identically. The first equation of motion, $\dot{x}^j = 2h^{jk}p_k$ remains identical at the reference time as an easy computation reveals. This strongly suggests one to revisit the *classical* Hamiltonian theory and develop a *covariant* formalism by viewing p_j as a covariant one tensor from which one can build higher order tensors. An expression such as $x^j p_j$ must then be interpreted as a scalar by regarding the x^j as the coordinate expressions of a contra variant one tensor. Therefore, the correct derivative to apply to x^j is the covariant derivative defined by the spatial metric. This suggests one to define the covariant Poisson bracket

$$\{f, g\}_c = \sum_k \left(\nabla_j^k(f) \frac{\delta}{\delta p_j^k}(g) - \frac{\delta}{\delta p_j^k}(f) \nabla_j^k(g) \right)$$

where the index k sums over all different particles and f, g can be scalar functions of the type

$$\sum_n T^{a_1 \ldots a_n}(x_j^k) p_{a_1}^{k_1} \ldots p_{a_n}^{k_n}$$

or a general tensor in which we have suppressed the tensor indices. Since this Poisson bracket maps tensors to tensors of the same type and the Hamiltonian is a scalar, the time derivative on the left hand side must be the covariant derivative

$$\frac{D}{dt} p_j = \dot{p}_j - \Gamma_{jk}^r \dot{x}^k p_r$$

with as result that we only have equations between manifestly covariant properties. Applied to our Hamiltonian above, this results in the system

$$\frac{dx^j}{dt} = \{x^j, H\}_c = 2h^{jk}p_k, \; \frac{D}{dt} p_j = \{p_j, H\}_c = 0$$

and we shall show now that this idea can be consistently applied to any Hamiltonian of second order in the momenta where the kinetic term is of the metric form. Given

$$H = h^{jk} p_j p_k + p_k A^k + B$$

the standard equations of motion are

$$\dot{x}^j = h^{jk} 2p_k + A^j, \; \dot{p}_j = -\partial_j h^{kl} p_k p_l - p_k \partial_j A^k - \partial_j B$$

and the last equation is equivalent to

$$\frac{D}{dt} p_j = -p_k \nabla_j A^k - \partial_j B = \{p_k, H\}_c$$

which we needed to show. What I propose now, is that it is the covariant bracket which needs to be quantized instead of the Poisson bracket; the former however does not satisfy the Jacobi identity anymore as

$$\{f,\{g,h\}_c\}_c + \{g,\{h,f\}_c\}_c + \{h,\{f,g\}_c\}_c = R_{jk}(h)\frac{\delta}{\delta p_k}(g)\frac{\delta}{\delta p_j}(f) +$$

$$R_{jk}(g)\frac{\delta}{\delta p_k}(f)\frac{\delta}{\delta p_j}(h) + R_{jk}(f)\frac{\delta}{\delta p_k}(h)\frac{\delta}{\delta p_j}(g)$$

which constitute the Riemann curvature tensor corrections. Hence, it seems that the Jacobi identity is something which pertains to flat space(time) which suggests that in a general curved space the commutator will have to be replaced by something else, at least if we take Dirac's suggestion seriously. Since this material is, as far as I know, a new addition to the literature we leave its full implications to be investigated in the future; also, we shall come back to this when dealing with quantum field theory.

In my opinion, we stress a very important point here which is that the formulation of physical laws should be such that, in principle, one could do without coordinates all together. General relativity is such theory as one formulates the basic functional, that is the action principle, in a way which does not depend upon coordinates; Regge calculus provides a generalization of this principle towards piecewise linear manifolds. Likewise did I want to convey the attitude in chapter two that potentialities do not really depend upon the coordinate system at hand and can be used in any framework of discrete spacetime too so that effectively they should transform as a scalar or a density like we suggested above. Nevertheless, there will be always those who would like to regard coordinates as fundamental in the description of the theory and for them alone can our potentialities be attached to a coordinate system, which would imply that one cannot speak in terms of "properties" anymore but one deals with "representations of properties". I deem this stance to be very unlikely but let us examine nevertheless its consequences; the constraints at hand are that the measure

$$|\Psi(x,t)|^2 dx$$

or

$$|\Psi(x,t)|^2 h^{\frac{1}{2}} dx$$

needs to be preserved. In the second case, one only can only make the transformation

$$\Psi'(t,x') = e^{i\theta(t;x;x')}\Psi(t,x(x'))$$

and we will show now that this is insufficient to compensate for the non-covariant terms in the Hamiltonian so that the entire enterprise is misguided. Note that we are extremely liberal here and allow for θ to explicitly depend upon t even if the latter has nothing to do with the coordinate transformation; applying two coordinate transformations in a row should imply that

$$\theta(t;x(x'');x'(x'')) + \theta(t;x';x'') = \theta(t;x;x'')$$

meaning one must have an additive group representation in the sense that

$$\theta(g) \circ h + \theta(h) = \theta(g \circ h)$$

where $x' = h(x'')$, $x = g(x')$ and θ sends a injective coordinate transformation to a function. We will argue now from different sides: suppose for a moment that the correct momentum operator is given by $-i\hbar\partial_j$ even if it is not Hermitian. The "canonical" Hermitian momentum operator is given by

$$-i\hbar\partial_j - \frac{i\hbar\partial_j h^{\frac{1}{2}}}{2h^{\frac{1}{2}}} = -i\hbar\partial_j - \frac{i\hbar h^{rs}\partial_j\left(h_{rs}\right)}{4}$$

and the latter contains the terms necessary for obtaining a covariant Hamiltonian. Closer inspection, however, shows that it produces also lots of higher order non-covariant terms which cannot be eliminated and moreover, it does not transform covariantly under spatial coordinate transformations. Therefore, we shall restrict to the usual momentum and posit the Hamiltonian to be

$$H = -\frac{\hbar^2}{2m}h^{jk}\partial_j\partial_k.$$

It is now a straightforward exercise to show that a transformation of the type $\Psi'(t,x') = e^{i\theta(t;x;x')}\Psi(t,x(x'))$ cannot compensate for the non-covariant terms in $H' = -\frac{\hbar^2}{2m}h'^{jk}\partial'_j\partial'_k$. Indeed, the non covariant terms induce the following equation

$$i\hbar\partial_t\left(e^{i\theta(t;x;x')}\right)\Psi = -\frac{\hbar^2}{2m}h'^{jk}\partial'_j\partial'_k\left(e^{i\theta(t;x;x')}\right)\Psi -$$
$$\frac{\hbar^2}{m}h'^{jk}\partial'_j\left(e^{i\theta(t;x;x')}\right)\frac{\partial x^l}{\partial x'^k}\partial_l\Psi \;-\; \frac{\hbar^2}{2m}e^{i\theta(t;x;x')}h'^{jk}\frac{\partial^2 x^l}{\partial x'^j\partial x'^k}\partial_l\Psi$$

which can be split up in two equations, one for Ψ and another for $\partial_l\Psi$. The latter can be solved explicitly in terms of the derivatives of $\theta(t;x;x')$ and produces

$$i\partial'_j\theta(t;x;x') = -\frac{1}{2}h'^{lk}\frac{\partial^2 x^s}{\partial x'^l\partial x'^k}\frac{\partial x'^r}{\partial x^s}h'_{jr}$$

which is the necessary contradiction since $\theta(t;x;x')$ has to be real. This shows, in the context of this simple example, that one cannot eliminate the non-covariant terms by means of a measure preserving transformation. One might hope that adding a term

$$-\alpha\frac{\hbar^2}{2m}\partial_j\left(h^{jk}\right)\partial_k$$

would nevertheless allow for some representation of the group of coordinate transformations; a short computation, however, shows that the issue of a complex $\theta(t;x;x')$ remains unchanged proving, once again, that the non-covariant terms cannot be compensated by a measure preserving transformation. I think it is safe to say that such line of thought is dead and that this paragraph shows that Dirac quantization is in conflict with the principle of general covariance, at least this is so for point particles in a general background.

4.3 Non-covariance of Quantum Field Theory.

Quantum Field theory is by no means covariant given that the Hamiltonian has no suitable transformation properties under general coordinate transformations which results in an observer dependence of the vacuum state. The algebraic formulation of the theory "hides" this fact given that it only concentrates upon the covariant field equations and operator algebra $\phi(f)$ where f is a Schwartz function:

$$\begin{array}{rcl} \phi(f)^{\star} & = & \phi(\overline{f}) \\ \phi(\alpha f + \beta g) & = & \alpha\phi(f) + \beta\phi(g) \\ \phi(\left(g^{\mu\nu}\nabla_{\mu}\partial_{\nu} + m^2\right) f) & = & 0 \\ [\phi(f), \phi(g)] & = & i\hbar(f, \Delta g)1 \end{array}$$

and $\Delta(x, y) = G_R(x, y) - G_R(y, x)$ is the so called Pauli-Jordan bi-function. This constitutes the starting point of a recent, much more abstract approach developed by Fredenhagen, Brunetti and Verch and there are a few things one should understand regarding this formulation of physics: (a) it *appears* to be manifestly covariant (b) it works in any globally hyperbolic spacetime, for extensions towards more general cases, see the work of Kay and Fewster (c) it is unclear what an accurate replacement for the measurement axiom is given that one works with "local states". Also, a bit of reflection shows that the commutator contains expressions such as $\phi(f)\phi(g)$ where the support of f is in the past of the support of g; such expressions are physically meaningless and it is somewhat unsatisfying that the basic formulation of the theory hinges upon such construction. One could weaken the fourth axiom to

$$[\phi(f), \phi(g)] = 0$$

for the support of f and g spacelike to one and another. Generically, this would not change the theory with the possible exception that "1" may be replaced by any constant Hermitian operator and as such, Planck's constant isn't fixed and also the classical theory can be found in this way. As mentioned before, the theory is not covariant and the entire physics becomes coordinate dependent given that different global reference systems may give rise to distinct or no particle notions at all.

4.4 General covariance or geometrization at the core of all physical laws and entities.

So far, we have shown that quantum theory is not generally covariant and that general relativity is; moreover, we have argued that general covariance is required if we want the wavefunction to have a "genuine" probability interpretation. More in general, the case for geometric laws of physics is a deep one since the latter embodies the more primitive idea that all laws are determined by means of relationships. In case you need a class of coordinate systems without any geometrical meaning whatsoever, then you impose an absolute reference frame with respect to which everything is defined. Such kind of reference frames are meaningless given that all our measurements are relational and the physical laws should therefore only pertain to these relationships and nothing else. Funny enough, this is a kind of "Heisenbergian" argument for Einstein's point of view; Einstein however went further than that and proclaimed that every entity in your theory should be dynamical and this also applies to spacetime itself. This principle is called background independence and we shall come back to it in chapter eleven; the main theory of this book assumes a background to be given and does not care about determining it in a dynamical way from classical degrees of freedom in the universe. We simply assume that this has been done and that our quantum theory does not cause for a backreaction; this is a sensible point of view and we shall comment on that too further down chapter eleven.

So far our comments regarding the geometrization of physical laws; in the next chapter, we will speak about the geometrization of elementary particles. Therefore, in the theory which we shall develop from now on, every constituent is geometrical in nature which leads to a certain class of physical theories which are all viable candidates as a theory of nature.

Chapter 5

Fourier transform and generalized Heisenberg operators.

Prior to reading this chapter, the mathematically trained reader is supposed to have digested the introduction as well as chapter two on the general philosophy of physics and in particular the process view on spacetime. This view breaks so radically with Einstein's vision that I have decided to keep the presentation of this chapter in the "middle". That is, it will be presented from an Einsteinian point of view but when the reader notices we are getting into trouble, I will employ the process view to make everything precise. This has, as a downside, that the presentation of the final picture will not be as formal as possible and is somewhat "hidden" in comments regarding the *need* for a process view on spacetime. This, situation, however, shall be rectified in chapter six where we will present an extension of the same material but then from a radical process point of view. I think this is the best way of presenting things as it allows the reader to get gradually accostumed to the new way of thinking and constitutes therefore the better choice from a historical point of view. The intention of this chapter is to define a spin-0 particle wave in any curved spacetime and provide one with some examples as well as a *universal* probability interpretation. The reason for me to turn this into a separate chapter is that it allows me to treat the theory in full detail and generality and it paves the ground for the treatment of particles with higher spin also. However, in the latter case, more physical input is required especially regarding the physical polarization vectors of spin-$\frac{1}{2}, 1$ particles. To obtain those, we need the tensorial structure of the propagator in contrast for the case of spin-0 particles where this structure is trivial. Another motivation for this relies on the fact that the development of a Fourier theory in a general curved spacetime is of independent mathematical interest and we shall look at it from this point of view too. In what follows, we shall adopt an Einsteinian view on spacetime and consider the latter to be eternally given and fixed; therefore, take a generic, time-orientable spacetime $(\mathcal{M}, g)$ and select a base point x, k^a a Lorentz vector at x defined with respect to $e_a(x)$ and y any other point in $\mathcal{M}$. Let $\gamma(s)$ be a curve from x to y and denote by $k^\mu(s)$ the parallel transport of $k^\mu(x) = k^a e^\mu_a(x)$ along γ, then we can define a potential $\phi_\gamma(x, k^a, y)$ by means of the differential equation

$$\frac{d}{ds}\phi_\gamma(x, k^a, \gamma(s)) = i\dot{\gamma}^\mu(s)k_\mu(s)\phi_\gamma(x, k^a, \gamma(s))$$

with boundary condition $\phi_\gamma(x, k^a, x) = 1$. Then, one easily calculates that in Minkowski spacetime, the potential is independent from the choice of γ and is given by the following group representation

$$\phi(x, k^a, y) = e^{ik_a(y^a - x^a)}$$

where the formula is with respect to global inertial coordinates defined by the vierbein $e_a(x)$. Minkowski is special in many ways: (a) every two events are connected by a unique geodesic (b) the ϕ_γ are path independent and define a group representation. Neither (a) nor (b) are true in a general curved spacetime which means we have to select for a preferred class of paths: the natural choice being that the *information* about the birth of a particle at x travels freely, meaning on geodesics which implies that we should sum over *all* distinct geodesics between x and y. This inspires one to consider the following mapping

$$\tilde{\phi} : T^\star\mathcal{M} \times T^\star\mathcal{M} \to U(1) : (x, k^a, w^a) \to \tilde{\phi}(x, k^a, w^a)$$

where $\tilde{\phi}(x, k^a, w^a)$ is defined as before by means of integrating the potential over the unique geodesic emanating from x with tangent vector w^a and affine parameter length one. One has then that

$$\phi(x, k^a, y) = \sum_{w:\exp_x(w)=y} \tilde{\phi}(x, k^a, w^a)$$

and although $\tilde{\phi}$ is more fundamental, we will sometimes switch between $\tilde{\phi}$ and ϕ by assuming that they are the same meaning that every two points in spacetime can be connected by a unique geodesic: this last assumption will be abbreviated to GS standing for "geodesic simplicity". In a general spacetime,

$$\tilde{\phi}(x, k^a, w^b) = e^{ik^a w_a} = e^{-ik^a e^\mu_a(x)\sigma_{,\mu}(x, \exp_x(w))}$$

where we assume in the last equality GS to hold and

$$\sigma(x, y) = \frac{1}{2}\epsilon L^2(x, y)$$

is Synge's function where $\epsilon = -1$ if x and y are connected by a spacelike geodesic and 1 if they are connected by a timelike geodesic and $L(x, y)$ denotes the geodesic length. Covariant derivatives of $\sigma(x, y)$ with respect to x will be denoted by unprimed indices μ, ν whereas their counterparts with respect to y

are denoted with primed indices. It is clear that as usual the standard Fourier identities hold between the two tangent spaces at x, that is

$$\int_{T^{\star}\mathcal{M}_x} \frac{dk^a}{(2\pi)^4} \overline{e^{ik^a w_a}} e^{ik^a v_a} = \delta^4(w^a - v^a)$$

and

$$\int_{T^{\star}\mathcal{M}_x} \frac{dw^a}{(2\pi)^4} \overline{e^{ik_a w^a}} e^{il_a w^a} = \delta^4(k^a - l^a)$$

being the inverse Fourier transform. Under the hypothesis of GS, the first integral reduces to

$$\int_{T^{\star}\mathcal{M}_x} \frac{dk^a}{(2\pi)^4} \overline{e^{-ik^a e^{\mu}_a(x)\sigma_{,\mu}(x,y)}} e^{-ik^a e^{\mu}_a(x)\sigma_{,\mu}(x,z)} = \frac{\delta^4(y,z)}{\sqrt{-g(y)}\Delta(x,y)}$$

and the second one under the additional assumption of geodesic completeness (GC) becomes

$$\int_{\mathcal{M}} \frac{d^4y}{(2\pi)^4} \sqrt{-g(y)}\Delta(x,y) \overline{e^{-ik^a e^{\mu}_a(x)\sigma_{,\mu}(x,y)}} e^{-il^a e^{\mu}_a(x)\sigma_{,\mu}(x,y)} = \delta^4(k^a - l^a).$$

Here,

$$\Delta(x,y) = \frac{|det(\sigma_{,\mu\nu'}(x,y))|}{\sqrt{-g(x)}\sqrt{-g(y)}}$$

is the absolute value of the Van Vleck-Morette determinant. Still working under the GS assumption, one recognizes the presence of a global coordinate system given by $\sigma_{,\mu}(x,y)$ which transforms as a co vector under coordinate transformations at x; contracting with $e^{a\mu}(x)$, one obtains local Lorentz coordinates $\sigma^a(x,y)$ and momentum operators $i\frac{\partial}{\partial \sigma^b(x,y)}$ which transform as a local Lorentz covector such that

$$i\frac{\partial}{\partial \sigma^b(x,y)}\phi(x,k^a,y) = k_b \phi(x,k^a,y)$$

meaning our generalized exponentials are eigenfunctions of the relative momentum operators. Also,

$$-\eta^{ab}\frac{\partial}{\partial \sigma^a(x,y)}\frac{\partial}{\partial \sigma^b(x,y)}\phi(x,k^a,y) = k^2\phi(x,k^a,y)$$

meaning that the above operator is to be preferred over the generalized d'Alembertian. In Minkowski spacetime, something special happens as

$$\sigma^b(x,y) = x^b - y^b$$

and one can substitute $i\frac{\partial}{\partial \sigma^b(x,y)}$ by $i\frac{\partial}{\partial x^b}$ or $-i\frac{\partial}{\partial y^b}$. In other words, the x,y coordinates factorize and one can identify all pictures in this way and obtain *one* Heisenberg pair only. Indeed, I have stressed in the introduction that the philosophy of Minkowski is misleading due to its translational invariance and the reader should appreciate that the latter just falls out from our formalism. Also, it is now clear that a generalized Heisenberg picture demands the condition of geodesic simplicity whereas there is no good physical reason why this should be the case: our geometric framework is far more interesting than that.

5.1 Elementary particles and a universal probability interpretation.

In this section, we shall work towards a theory for a single free spinless particle in a general curved spacetime, the extension towards multiple particles of higher spin being worked out in the next chapters. Our first, preliminary, postulate of relativistic quantum theory consists in the statement that an *idealized* free, spin-0 particle of mass m and future pointing momentum k^a, created at x, is given by the Fourier wave $\phi(x, k^a, y)$ where $k^2 = \eta_{ab}k^a k^b = m^2$ is Einstein's energy-momentum relationship or the *mass-shell* condition. We have deduced that in a geodesically simple universe

$$(\eta^{ab}\frac{\partial}{\partial\sigma^a(x,y)}\frac{\partial}{\partial\sigma^b(x,y)} + m^2)\phi(x, k^a, y) = 0$$

which is the correct generalization to a GS spacetime of the Klein-Gordon equation in flat spacetime. In the literature however, one proposes the equation

$$(g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2)\psi(x) = 0$$

which leads to a conserved current

$$j^\mu(\overline{\psi}, \phi)(x) = -i(\overline{\psi}(x)\nabla^\mu\phi(x) - \phi(x)\nabla^\mu\overline{\psi}(x))$$

on the space of solutions ψ, ϕ to the Klein-Gordon equation. Standard arguments in the old fashioned quantum theory then suggest that the correct probability interpretation is given by the charge of this current which determines the bilinear form

$$\langle\psi|\phi\rangle = -i\int_\Sigma d^3x\sqrt{h(x)}n_\mu(\overline{\psi}(x)\nabla^\mu\phi(x) - \phi(x)\nabla^\mu\overline{\psi}(x))$$

where $h(x)$ is the determinant of the induced metric on the Cauchy hypersurface Σ and n_μ is the normal vector to it. So, this reasoning only holds in globally hyperbolic spacetime. There are two problems with this scalar product: (a) it is of indefinite signature meaning there are as many positive as negative norm states in a nondegenerate basis and (b) the probability density is not positive restricted to some positive norm solution meaning it cannot serve as the probability density associated to a generalized position operator. (a) is well known and reflects that the theory is not unique or covariant given that distinct splits of the total vector space of solutions in a positive and negative norm subspace determine different theories. (b) on the other hand is not well known and even true in Minkowski spacetime; indeed, the density for a superposition of two plane waves

$$\alpha e^{ik_a x^a} + \beta e^{il_a x^a}$$

on an inertial hypersurface Σ reads

$$\overline{(\alpha e^{ik_a x^a} + \beta e^{il_a x^a})}(k^0\alpha e^{ik_a x^a} + l^0\beta e^{il_a x^a}) + cc = 2k^0|\alpha|^2 + 2l^0|\beta|^2 + 2\mathrm{Re}(\alpha\overline{\beta}e^{i(k_a - l_a)x^a})(k^0 + l^0)$$

where Re denotes the real part. By adjusting the phase of α we get at some value of x the expression

$$2k^0\left|\alpha\right|^2 + 2l^0\left|\beta\right|^2 - 2\left|\alpha\right|\left|\beta\right|(k^0 + l^0)$$

which can easily be made smaller than zero. Now, in Minkowski, unlike in any other spacetime, it is still possible to save the day as one can look for a canonical Heisenberg conjugate of the dynamical momentum operators and interpret those as position operators. It must be clear to the reader that the only possible position density is given by

$$\left|(T\phi)(x)\right|^2$$

where T constitutes a linear transformation of ϕ. To find this operator in Minkowski spacetime and generalize it to our setting later on, note that the correct scalar product between plane waves is given by

$$\langle e^{ik_a x^a}|e^{il_b x^b}\rangle_x = (2\pi)^3 k^0 \delta(\vec{k} - \vec{l})$$

where the right hand side is Lorentz invariant given that the left hand side must be. Therefore, we obtain that with

$$\psi(x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int d^3k\widehat{\psi}(k)e^{ik_a x^a}$$

and

$$\phi(x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int d^3k\widehat{\phi}(k)e^{ik_a x^a}$$

that

$$\langle\psi|\phi\rangle_x = \int d^3k k^0\overline{\widehat{\psi}(k)}\widehat{\phi}(k) = \int_{\mathbb{R}^3} d^3x\overline{(T\psi)}(t,x)(T\phi)(t,x)$$

where

$$(T\psi)(x) = \frac{1}{(2\pi)^{\frac{3}{2}}}\int d^3k\sqrt{k^0}\widehat{\psi}(k)e^{ik_a x^a}.$$

There is a canonical Lorentz invariant wave function associated to a particle being born at x which is given by a dimensionless multiple of

$$\delta_x(y) = \frac{1}{(2\pi)^3 m}\int \frac{d^3k}{k^0}e^{ik_a(y^a - x^a)}$$

which is, as the notation suggests, a relativistic replacement of the $\delta^3(\vec{y} - \vec{x})$ function of dimension mass instead of mass^3. Indeed, as the reader may verify later on, we have that

$$\langle W(z,y)|\delta_x(y)\rangle_y = \delta_x(z) = \int d^3y\, T_y\delta_x(y)\overline{T_y W(z,y)}$$

where

$$W(x,y) = \int \frac{d^3k}{(2\pi)^3 k^0}e^{ik_a(y^a - x^a)} = m\delta_x(y)$$

is the propagator and T_y means the T operator with respect to the y variable. These facts indicate the correct probability interpretation in a general curved spacetime given that the dimension of $(T\psi)$ is given by mass$^{\frac{3}{2}}$ which means that the dimension of ψ is given by mass. In our setting, particle notions depend upon the place where they are born and therefore also depend upon two points x, y which means one can consider two operators T_x and T_y applied to it, but here T_x has a slightly distinct meaning than before. Indeed, T_x means that the T operation is applied with respect to the Fourier waves $\phi(x, k^a, y)$ and the reader should keep this in mind. In a general curved spacetime, we do not dispose of analogues of $\langle | \rangle_x$ given that the derivative of Synge's function does not factorize; we can, however, generalize the T mappings and spatial scalar products in a canonical way. More specifically, regarding any particle state

$$\psi_x(y) = \frac{1}{(2\pi)^{\frac{3}{2}}} \int d^3k \, \widehat{\psi}(k)\phi(x, k^a, y)$$

we define

$$(T_{x,e_0}\psi_x)(y) = \sum_{w:\exp_x(w)=y} \frac{1}{(2\pi)^{\frac{3}{2}}} \int d^3k \, \sqrt{k^{0'}_{\star w}}\widehat{\psi}(k)\tilde{\phi}(x, k^a, w)$$

where $k^{0'}_{\star w}$ is the component of $k_{\star w}$ with respect to e_0 at y and the *weight* for a particle to cross a spatial, but not necessarily achronal, cross section Σ, whose intersection with $J^-(x) \setminus J^+(x)$ is empty, is given by

$$w_\Sigma(\psi_x) = \int_\Sigma d^3y \sqrt{h(y)} \, |(T_{x,e_0 \perp \Sigma}\psi_x)(y)|^2$$

whereas the propagation, seen as a process of annihilation and recreation, is given by

$$P_\Sigma(\psi_x)(z) = \int_\Sigma d^3y \sqrt{h(y)} T_{x,e_0\perp\Sigma}\psi_x(y) \, \overline{T_{z,e_0\perp\Sigma}W(z, y)}.$$

Here, again, we assume Σ to reside outside of $J^-(x) \setminus J^+(x)$ where x is the point of birth meaning that if Σ intersects the past of x, it must do so from the future which can only happen if one considers space times with closed timelike curves. Without going too much into detail here, in the spacetime process philosophy of chapter two, we assume that Σ itself has come to birth *after* x has been born which is somewhat difficult to express in our notation since the relations $J^\pm(x)$ are *perennials*, they exist once and forever given that the entire spacetime has been given. Once, by means of a sequence of processes, a closed timelike curve has been formed we will always interpret x as being in the past of Σ; the reader notices that when a closed timelike curve is formed, the entire "propagation" picture behind the wave ψ_x dissapears as new geodesics are included in the very *definition* of the wave. There is no analogue of the propagation process in standard quantum mechanics where unitarity and the unique choice of a Cauchy hypersurface Σ garantuee that

$$P'_\Sigma(\psi)(x) = \psi(x)$$

where P' is defined by means of the Klein-Gordon product explained above. In fact, if you think about this, the latter equation is not very natural given that propagation through a hypersuface is associated to a process where knowledge about the state of the particle has been *gained* and therefore, there is no reason why this should come without a cost. The latter translates itself into the loss of "unitarity" in our framework. Likewise do we define the weight of detection on a world tube W_Σ of spatial hypersurfaces Σ_t which correspond to the "time evolution" (towards the future) of a spatial surface Σ, given that the particle is annihilated at some event y such that $y \in W_\Sigma$. In our philosophy of strong measurements, as explained in chapter two, we insist that a measurement corresponds to a process of *renewal* which is always associated to an annihilation: basically, the particle leaves one quantum world and enters another one. Obviously, the point of annihilation is born after Σ and the size of Σ is always very small, certainly below the micron scale. Hence, we define

$$d_{W_\Sigma}(\psi, y; \delta) = \int_{\Sigma_{t-\delta}} d^3z \sqrt{h(z)} \left| T_{y,e_0 \perp \Sigma} \widetilde{\psi}_y(z) \right|^2$$

where $\widetilde{\psi}_{y|_{\Sigma'}} = \chi_{\Sigma_{t-\delta}} \psi_{|\Sigma'}$, $\Sigma_{t-\delta} \subset \Sigma'$ and $y \in \Sigma_t$; Σ' is a complete spacelike hypersurface for y meaning that every geodesic emanating from y remains to the future of Σ' or crosses it; $\chi_{\Sigma_{t-\delta}}$ is the characteristic function on $\Sigma_{t-\delta}$. Moreover, Σ' contains *actual space* at time δ prior to the happening of the annihilation process as seen by the measurement apparatus and it is, moreover, assumed that $\Sigma_{t-\delta}$ can be reached by means of a geodesic starting at y at the instant y is born. Under those conditions, and possibly some slight technical details, one should be able to show that $\widetilde{\psi}_y$ is unique. This expression gives the probability for some spot or trace to be found in Σ_t, whereas the calculation refers to a past state $\Sigma_{t-\delta}$. So, the determination of the probability that a trace of the particle's impact is found in a certain region is fixed at the moment of annihilation, which means that detection is by no means a simple mechanism. It is a bit like a wound on your skin which appears some time after you have been hurt. The reader must obtain the intuition that our process view on spacetime itself is *mandatory* for a probability interpretation to be defined, which is logical since quantum theory itself is a theory of processes. Now, we are all set for our relative probability interpretation to be defined: the relative amplitude for a particle to be detected into world tubes W_{Σ_i} on the slices Σ_{i,t_i} where $W_{\Sigma_1} \cap W_{\Sigma_2} = \emptyset$ given annihilation events $y_i \in \Sigma_{i,t_i}$, is determined by

$$\frac{d_{\Sigma_1}(\psi, y_1; \delta)}{d_{\Sigma_2}(\psi, y_2; \delta)}$$

where ψ is the spacetime state of the particle and y_1, y_2 happen at the same time in our process view. There is a lot of new physics in here: for example, the probability that a "wound" is found in the region Σ_t depends on the amount of some processing "time" δ associated to the apparatus. This is still a very simple model and more complex detection processes can be set up, depending upon the nature of the machine.

There remains another relative amplitude to be defined which expresses the amplitude between the processes of a particle crossing a hypersurface Σ_i and then being detected by an apparatus with world tube W_{Σ_f} at the time it is annihilated at y. It is given by

$$\frac{d_{W_{\Sigma_f}}(\psi_x, y; \delta)}{w_{\Sigma_i}(\psi_x)}$$

and it is a quantity which is really never considered in standard quantum mechanics. In the next section, we shall further formalize the remarks of this section and generalize it to multi-particle theories. We now turn our head towards some interesting example confirming that our theory is the right one.

5.2 Some interesting example.

What we will show in this section is that while maintaining flatness but imposing a non-trivial topology, leading to periodicity conditions on the wave vectors associated to the plane waves defined by the d'Alembertian operator, arises automatically in our framework due to an infinite winding of geodesics. Let us study the example of the timelike cylinder $\mathbb{R} \times S^1$ with coordinates (t, θ) where θ has to be taken modulo $L > 0$ and see if only the discretized modes $k^1 = \frac{2\pi n}{L}$ for some $n \in \mathbb{Z}$ play a part in the propagator. The reader has to be capable of figuring out that

$$\phi(x, k^a, y) = e^{i(\sqrt{(k^1)^2+m^2}\delta t - k^1 \delta\theta)} \left[\sum_{n\in\mathbb{Z}} e^{ik^1 L n}\right]$$

where

$$y - x = (\delta t, \delta\theta)$$

in the global flat coordinate system. This function is clearly invariant under the translation $\delta\theta \to \delta\theta \pm L$ and it is therefore well defined on the cylinder. Forming now a wave packet at x

$$\psi_x(y) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int dk^1 \widehat{\psi}(k^1)\phi(x, k^a, y) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int dk^1 \widehat{\psi}(k^1) e^{i(\sqrt{(k^1)^2+m^2}\delta t - k^1 \delta\theta)} \left[\sum_{n\in\mathbb{Z}} e^{ik^1 L n}\right]$$

and taking the Fourier transform with

$$\frac{1}{\sqrt{L}} e^{i\frac{2\pi p \delta\theta}{L}}$$

gives

$$\psi_x(y) = \frac{1}{L} \sum_{p\in\mathbb{Z}} \left(\int_0^L \psi_x(y) e^{i\frac{2\pi p\delta\theta}{L}} d(\delta\theta) \right) e^{-i\frac{2\pi p\delta\theta}{L}}$$

and it is easy to calculate that

$$\frac{1}{L}\int_0^L \psi_x(y)e^{i\frac{2\pi p\delta\theta}{L}}d(\delta\theta) = \frac{1}{(2\pi)^{\frac{1}{2}}L}\int dk^1 \int_{-\infty}^{+\infty} d(\delta\theta)e^{i\left(\sqrt{(k^1)^2+m^2}\delta t+\left(\frac{2\pi p\delta\theta}{L}-k^1\right)\delta\theta\right)}\widehat{\psi}(k^1).$$

The latter equals

$$\frac{(2\pi)^{\frac{1}{2}}}{L}e^{i\sqrt{\left(\frac{2\pi p\delta\theta}{L}\right)^2+m^2}\delta t}\widehat{\psi}\left(\frac{2\pi p\delta\theta}{L}\right)$$

which results in the ordinary Fourier transform

$$\psi_x(y) = \frac{(2\pi)^{\frac{1}{2}}}{L}\sum_{p\in\mathbb{Z}}\widehat{\psi}(\frac{2\pi p\delta\theta}{L})e^{i\left(\sqrt{\left(\frac{2\pi p\delta\theta}{L}\right)^2+m^2}\delta t-\frac{2\pi p\delta\theta}{L}\right)}.$$

So, the winding of geodesics kills of all modes which do not satisfy the global boundary conditions. A similar result of course holds for the propagator and the reader may enjoy making that exercise. This example obviously generalizes to higher dimensional cylinders over the spatial d-dimensional torus $\mathbb{T}^d$. This concludes the shortest chapter of this book, but nevertheless an important one as it indicates very clearly the line of thought to be followed in the chapters to come.

Chapter 6

Spin, two point functions, probability and particle statistics.

In Minkowski space time, we showed in the previous chapter and the introduction that the correct two point function for a particle of zero spin is given by

$$W(x, y) = \int \frac{d^3k}{(2\pi)^3\, 2E_k} e^{ik_a(y^a - x^a)}$$

where the signature of the metric is $(+ - --)$ and $k^0 = E_k = \sqrt{\vec{k}^2 + m^2}$. Another way to write it is

$$W(x, y) = \int \frac{d^4k}{(2\pi)^3} e^{ik_a(y^a - x^a)} \delta(k^2 - m^2)\theta(k^0)$$

where $\theta(x) = 1$ if $x \geq 0$ and 0 otherwise. The delta and theta function just express Einstein's energy-momentum relationship and the fact that the four momentum is future pointing whereas the factor $e^{ik_a(y^a - x^a)}$ has been explained by means of a geometric process dictated by the principle of Lorentz invariance. $W(x, y)$ is supposed to be a relational quantity in the sense that it relates the creation of a particle at x to the annihilation theirof at y so, therefore, we looked for such a way to define the exponential function. In the previous section, we dictated that *information* about the birth of a free particle has to travel freely, meaning on geodesics; it is our first intention to explain the ramifications of this choice. Therefore, define the two point function in a general time-orientable curved space time by means of

$$W_\gamma(x, y) = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0)\phi_\gamma(x, k^a, y)$$

where, again, we do not fix γ to be a (collection of) geodesic(s). This definition is clearly Lorentz invariant, as it should, and from the equality

$$\overline{\phi_\gamma(x, k^a, y)} = \phi_\gamma(y, k_\star^{a'}, x)$$

and the fact that the mapping $\star(x, y); T^\star\mathcal{M}_x \to T^\star\mathcal{M}_y : k^a \to k_\star^{a'}$ is an orthochronous Lorentz transformation, it follows that

$$\overline{W(x, y)} = W(y, x)$$

as it should. This result does not depend upon the path γ joining x to y; the following demand however leaves in general just one option open:

$$W(x, y) = W(y, x)$$

for all $x \sim y$ where $\sim$ stands for being spacelike related. This is our demand of quantum causality, it says that the amplitude for propagation of a particle between two spacelike separated points x and y does not depend upon the order of the points. We now show that if γ is a geodesic between x and y, then this demand is automatically satisfied. By definition this geodesic must be a spacelike geodesic (it may be possible for timelike separated points to be joined by a spacelike geodesic such as occurs on the timelike cylinder); hence

$$\phi(x, k^a, y) = e^{ik_a w^a}$$

where $w^a w_a = 2\sigma(x, y)$, w^a is tangent to the geodesic at x and $\sigma(x, y)$ is Synge's function. Equivalently,

$$\phi(x, k^a, y) = e^{-i\sigma(x,y)_{,\mu} e_a^\mu(x) k^a}$$

as the reader may show or $w^a = -e^{a\mu}(x)\sigma_{,\mu}(x, y)$. To prove that the associated two point function satisfies indeed quantum causality, consider the reflection around w^a, the latter is a Lorentz transformation, preserving the sign of k^0 if k^a is a causal vector and maps $k^a w_a$ to $-k^a w_a$; hence, $W(x, y) = \overline{W(x, y)}$ which proves our assertion. In the case of general paths, the reader may easily see that this reflection of k^a does not need to flip the sign of $w^\mu(s)k_\mu(s)$ as this quantity is not preserved under general transport; the very preservation requires the geodesic equation to be fullfilled. One can now wonder to what extend the Klein-Gordon equation still plays a roll; consider that $W(x, y) \equiv W(\sigma_{,\mu}(x, y))$ satisfies

$$\left(\Box' + m^2\right) W(x, y) = -ig^{\alpha'\beta'}\sigma_{,\mu\beta'\alpha'}\frac{\partial}{\partial\sigma_{,\mu}}W(x, y) + m^2 W(x, y) - g^{\alpha'\beta'}\sigma_{,\mu\alpha'}\sigma_{,\nu\beta'}\frac{\partial^2}{\partial\sigma_{,\mu}\partial\sigma_{,\nu}}W(x, y)$$

where primed indices refer to y and unprimed to x and all derivatives of σ are covariant derivatives. The reader now notices that in the coincidence limit $y \to x$, we have that the left and right hand side reduce to zero where we use Synge's rule $[\sigma_{,\mu\beta'}] = -g_{\mu\beta}$ and $[\sigma_{,\mu\alpha'\beta'}] = 0$ where the square brackets indicate that the limit $y \to x$ is taken. Before we proceed, let us stress that our point of view is relational in the sense that it is the way we have build the two point function, the point of view of field operators was absolute in the sense that propagation is a derived concept of composite entities whereas here, the bifunction is fundamental. Notice also that the above formula gives our covariantization of the flat space time equation and as anticipated in the previous chapter, the right hand side is in general not zero; we will come to other, more substantial deviations later on. Our two point function is natural in the sense that it only depends upon the geodesics joining the two points which is as "local" as one may get. There is a useful information interpretation of our formula which is that the information of the creation of a particle travels on geodesics possibly exceeding the local speed of light: therefore, the interacting theory will be constructed as a theory of interacting information currents.

6.1 Spin-0 extended.

So, we have now uncovered why our paths along which information travels have to be geodesics and why we have to sum over all of them by means of simplifying assumptions which allowed for the use of Synge's function. Also, in the last chapter, we invented a relative probability interpretation which we introduced for the first time in chapter three. Given the somewhat more general character of our setup, we will introduce some extra notation needed for future reference and a tool for the reader to verify the aforementioned properties of the propagator. In particular, we need to change $k_{\star}^{a'}$, being a Lorentz vector at y, to $k_{\star\,w}^{a'} = \Lambda(x,w)_a^{a'} k^a$ being a Lorentz vector at $y = \exp_x(w)$ determined by the parallel transporter $\Lambda(x,w)$ which is defined by dragging a generic vector over the geodesic connecting x with y with tangent vector at x given by w. Using this concept, the reader should be able to prove all desirable properties of the two point function

$$W(x,y) = \int \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\theta(k^0)\phi(x,k^a,y).$$

Before we come to the general analysis of spin, let me take the condition of internal temporality into account discussed briefly in the introduction and stating that to any process going "backwards" in time, there corresponds an amplitude of a process going forwards in time. Here, the formula that $W(x,y) = W(y,x)$ for $x \sim y$ is of great help since it suggests the property of Bose statistics; that is, if we interchange the properties of two identical spin zero particles, then the total amplitude remains the same. These remarks *could* lead to the following definition of the Feynman propagator:

$$\Delta_F(x,y) = W(x,y)$$

if $y \in J^+(x) \setminus J^-(x)$,

$$\Delta_F(x,y) = W(y,x)$$

if $y \in J^-(x) \setminus J^+(x)$,

$$\Delta_F(x,y) = \frac{1}{2}\left(W(x,y) + W(y,x)\right)$$

if $y \in J^+(x) \cap J^-(x)$ and $\Delta_F(x,y) = W(x,y) = W(y,x)$ for $x \sim y$. The reader easily verifies that

$$\Delta_F(x,y) = \Delta_F(y,x)$$

as it should be: it is the Feynman propagator which shall be used in defining the interaction theory. It is important to notice that we have extended the usual definition to the case that $J^+(x) \cap J^-(x)$ is non-empty, the standard Feynman propagator only being defined on globally hyperbolic space times. The reader notices that basically we have no choice but to frame the definition in this way if we start from the two point function W alone, at least if we want to preserve bose statistics. However, the novel part is extremely non-local and therefore of a somewhat "mysterious" physical nature, as we shall argue for later on. Therefore, this prescription shall be modified in chapter seven; it will turn out that the Feynman propagator cannot be directly expressed in terms of the two point function in that case. It is obvious that the singularity structure of our two point function is of Hadamard type and therefore identical to the one of the standard Minkowski vacuum; this leads to infinite renormalizations which one would preferably avoid and this matter will be thoroughly discussed in the next chapter. One might consider the following regularization of $W(x,y)$:

$$W(x,y) = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3}\, \delta(k^2 - m^2)\theta(k^0) e^{-\kappa R_{\mu\nu}(x)k^\mu k^\nu - \kappa R_{\mu'\nu'}(y) k_\star^{\mu'} k_\star^{\nu'}} \phi(x, k^a, y)$$

if and only if $x \in J^\pm(y)$ and for $x \sim y$ our expression remains unchanged. Here, we use the fact that $k^\mu_{\star(x,y)\star(y,x)} = k^\mu$ to deduce that W still satisfies $\overline{W(x,y)} = W(y,x)$. Here we of course allude to an energy condition of the type $R_{\mu\nu}(x)V^\mu V^\nu > 0$ for any non-spacelike vector and $\kappa > 0$ so that the above integral converges and hence we remove the singularity structure of the two point function. Physically, this is very appealing since one would expect the gravitational field to give an ultraviolet regulator for quantum physics which is exactly what the above formula tells you. Minkowski space time with its associated infinite renormalization difficulties may then be seen as a singular limit where the gravitational energy condition trivializes. Let me also comment on how the Heisenberg commutation relations are hidden in our formalism: $W(x,x)$ is an integral over all on shell-momenta, each with an equal amplitude (in the non-modified version) of 1 which basically means that if you nail the particle at x, the momentum is going to be democratically uncertain. This is precisely the content of Heisenberg's commutation relation, the gravitationally modified two point function hence imposes corrections to that backbone of quantum theory. I think these gravitational modifications are certainly worthwhile studying as they constitute natural candidates regarding the equivalence principle. Also, our "equation of motion" for $\phi(x, k^a, y)$ can be thought of as a covariant substitute for the Schrodinger equation. To violate Bose statistics, it is sufficient for information to travel on different paths than geodesics which might be the case when an external force, such as the one associated to an observer, intervenes.

6.2 General theory of spin.

This section contains the most deviating ideas from conventional quantum mechanics in *flat* space-time where conventional wisdom determines the notion of spin from an operational point of view attached to *global* Poincaré transformations. This is clearly not a viable point of view as a global Poincaré symmetry is not a symmetry of nature. In our language, we have a globally determined coordinate system $\sigma_a(x, y)$ determined by local Lorentz indices a at x in case space time is geodesically simple and complete. This might suggest to walk the same route and look for an operational, active, unitary, representation of the local Lorentz group for which our Fourier waves would consists spin zero-particles. However, there is no a-priori reason to do that given the limitations it imposes on space time; more generally, one might want to consider the action of the local Lorentz group on the tangent vectors at x. Such a point of view would not even work for the free theory of one particle born at x and annihilated at y given that both actions are defined with respect to different Lorentz groups and live in different Hilbert spaces, which shows that the definition of interactions in such framework would become an impossible thing from the usual point of field theory. Moreover, in our framework, the action on the tangent vectors w at x does not necessarily project down to one on space time as the action of a Lorentz transformation on two vectors, which determine the same space time point y by means of the exponential map $\exp_x$, might result in different points on space time. So, our definition of a "local" particle born at x is not commensurable with the application of an *active* local Lorentz transformation at x: the reader might want to give examples where the action of a local Lorentz transformation would be ill defined as the point would fall outside of space time.

By this, I do not intend to say that it is a-priori impossible to try to define a whole *new* quantum theory in which interactions are defined in an entirely different way and the operational formalism has to be extended in order to incorporate actions of the local Lorentz group but alas, a universal attempt of mine in that direction resulted in internal inconsistencies. So, I have *tried* to walk that route and generalized the operational definition of spin to the tangent bundle by relying upon ultra-local particle notions living *on* tangent space, defined in a *different* way than what we have done so far for spinless particles in this book, and I have recieved the lesson that such idea does not work. In our theory, we simply stick to the reality that a *space time* action of the local Lorentz group is ill defined and therefore spin cannot be retrieved from such formalism in the usual way. Therefore, spin has to be associated to *bundle* actions of the local Lorentz group requiring finite dimensional representations of the local Lorentz group which must therefore be non-unitary but which can always be chosen in such a way that the physically relevant, compact, part is represented in a unitary way. Indeed, it is an open question in the standard approach towards spin *why* massless particles shouldn't come in a continuous spectrum of internal degrees of freedom, something which does not appear in nature, but which is explained in our approach given that all representations are finite dimensional. This leads us to the statement that only degrees of freedom associated to a maximal *compact* part of the symmetry group can be physically measured which does not imply that the non-compact part does not influence the physics, in constrast to what is usually thought in the operational approach. More precisely, let us unravel spin for the fundamental spin-$\frac{1}{2}$ and spin one representations: all others following from the latter by means of tensorial products. Starting at spin one, let k^a be the four momentum at x and denote by $k^{\perp}$ the complexification of the space of vectors perpendicular to k; then, in case k is timelike, $k^{\perp}$ is a three dimensional Euclidean space, with inner product defined by $-\eta_{ab}$, and carrying an irreducible, unitary spin one representation of the little group[1] of k, which is $SO(3)$. In case k is null, $k^{\perp}$ is a three dimensional degenerate inner product space, with inner product induced by $-\eta_{ab}$, carrying an irreducible representation of the little group of k, which is the Euclidean group in two dimensions $E(2)$, from which only the rotation part, with respect to any timelike vectorfield e_0 and spatial axis e_3, is unitarily represented. The action of the translation part of $E(2)$ is given by

$$e_i \rightarrow e_i + \alpha_i k$$

and the rotations are given by

$$e_i \rightarrow O_i^j e_j$$

where O represents the rotation around e_3 and $k = e_0 + e_3$. Therefore, in contrast to the standard theory, *helicity* of[2] a particle is not necessarily a Lorentz invariant concept given that the translations play a role too. The only way to make helicity into a Lorentz invariant concept is to make sure that the k-mode decouples dynamically from your theory; we will show that this only happens in some limit where the theory gets ill defined, which is standard quantum field theory on Minkowski. Hence, our version of quantum electrodynamics will always contain a "ghost" particle of zero helicity, which *could* potentially be observed if different geodesics connect two points, as well as the usual Faddeev-Popov ghosts added to the theory to decouple that "ghost" particle in the unphysical limit of standard quantum field theory on Minkowski. It will turn out however that we shall totally eliminate the k-mode from observation as this is needed to ensure positive probabilities. So far for the fundamental representation of the Lorentz group and its tensor products: the lesson we received is that the internal degrees of freedom are associated to irreducible representations of the little group determined by the hermitian projection operator $P_m(k) = \delta_b^a - \frac{k^a k_b}{m^2}$ in

[1]The little group of k^a is defined as the subgroup of the Lorentz group leaving k^a invariant.

[2]Helicity of a massless particle is the half integer number j such that there exists an eigenvector $|j\rangle$ corresponding to the eigenvalue $e^{2\pi j}$ of the rotation operator O determined by the e_0, e_3 plane.

the massive case and the non-hermitian projection operator $P_0(k,l) = \delta^a_b - l^a k_b$, where $k_a l^a = 1$, in the massless case. It is important to realize that the last operator depends upon the vierbein or a supplementary vector l_a such that

$$\Lambda P_0(k,l)\Lambda^{-1} = P_0(k,\Lambda l)$$

for any Λ in the little group of k. This is in contrast to the massive case where $P_m(k)$ commutes with the little group of k. We will now turn to the fundamental Dirac representation of the universal cover of the Lorentz group.

The Dirac representation is defined by means of the γ^a matrices satisfying

$$\gamma^a\gamma^b + \gamma^b\gamma^a = 2\eta^{ab}1$$

and $(\gamma^a)^\dagger = \eta^{aa}\gamma^a$ with a special role for γ^0 since

$$\gamma^0(\gamma^a)^\dagger\gamma^0 = \gamma^a.$$

The generators of spin rotations $\mathcal{J}^{ab}$ is given by

$$\mathcal{J}^{ab} = \frac{-i}{4}\gamma^{[a}\gamma^{b]}$$

and as we all know, the Dirac representation contains precisely two irreducible spin-$\frac{1}{2}$ unitary representations of the little group of k for $m > 0$, which is in this case $SU(2)$, where the relationship between the fundamental $SL(2,\mathbb{C})$ action on spinors on one side, and the action of the Lorentz group on tangent space on the other, is standard and given by

$$\sigma_a\Lambda^a_b v^b = \Lambda^{\frac{1}{2}}(v^c\sigma_c)(\Lambda^{\frac{1}{2}})^\dagger$$

where the $\sigma_a = (1,\sigma_i)$ constitute the usual Pauli matrices. We are now interested in finding the "self-adjoint" canonical projection operators on those subspaces. Here, it is necessary that

$$\Lambda^{\frac{1}{2}}P_\pm(k)\Lambda^{-\frac{1}{2}} = P_\pm(k)$$

for Λ in the little group of k. Using the covariance properties of the γ^a matrices

$$\Lambda^a_b\Lambda^{\frac{1}{2}}\gamma^b\Lambda^{-\frac{1}{2}} = \gamma^a$$

we obtain that

$$P_\pm(k) = \frac{1}{2m}\left(\pm k_a\gamma^a + m1\right)$$

satisfying

$$P_+(k)P_-(k) = 0$$

and the "hermiticity" properties with respect to the indefinite scalar product

$$\langle v|w\rangle = \overline{v}^T\gamma^0 w.$$

Now, it remains to find a preferred basis for those two dimensional subspaces: for this purpose, we introduce commuting operators with $P_{\pm}(k)$ which are defined by means of an infinitesimal rotation in a two plane perpendicular to k; more in particular, let m, n denote two unit spacelike vectors perpendicular to k and one and another, then a generator of rotations in the n, m plane is given by by

$$R(n,m) = n_{[a}m_{b]}J^{ab}$$

which constitutes an hermitian operator with respect to the indefinite scalar product and defines two hermitian projection operators

$$P_{\pm}(n,m) = \frac{1}{2}\left(\mp 4R(n,m) + 1\right)$$

satisfying

$$P_{+}(n,m)P_{-}(n,m) = 0.$$

Therefore, we can *define* four canonical, normalized, wave vectors $u_{n,m,k;\pm}, v_{n,m,k;\pm}$ as solutions to

$$P_{+}(k)P_{\pm}(n,m)u_{n,m,k;\pm} = u_{n,m,k;\pm}$$

and

$$P_{-}(k)P_{\pm}(n,m)v_{n,m,k;\pm} = v_{n,m,k;\pm}.$$

We study these vectors now in somewhat more detail; under a combined Lorentz and spin transformation, we have that

$$u_{\Lambda n,\Lambda m,\Lambda k;\pm} = \Lambda^{\frac{1}{2}}u_{n,m,k;\pm}$$

and likewise for $v_{n,m,k;\pm}$. We now choose a Lorentz frame such that $k = me_0, n = e_1, m = e_2$; in that case $P_{\pm}(e_0)$ and $P_{\pm}(e_1, e_2)$ are *also* hermitian operators with respect to the standard Euclidean inner product so that the $u_{e_1,e_2,me_0;\pm}, v_{e_1,e_2,me_0;\pm}$ constitute both an orthonormal basis with respect to the Lorentzian as well as the Euclidean inner product. In particular, we have that

$$\frac{1}{4}\left(\gamma^0 + 1\right)\left(\pm i\gamma^1\gamma^2 + 1\right)u_{e_1,e_2,me_0;\pm} = u_{e_1,e_2,me_0;\pm}$$

which reduces to

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \pm\sigma_3 + 1 & 0 \\ 0 & \pm\sigma_3 + 1 \end{pmatrix}u_{e_1,e_2,me_0;\pm} = 4u_{e_1,e_2,me_0;\pm}$$

and therefore

$$u_{e_1,e_2,me_0;\pm} = \frac{1}{\sqrt{2}}\begin{pmatrix} \chi_{\pm} \\ \chi_{\pm} \end{pmatrix}$$

and likewise

$$v_{e_1,e_2,me_0;\pm} = \frac{1}{\sqrt{2}}\begin{pmatrix} \chi_{\pm} \\ -\chi_{\pm} \end{pmatrix}$$

where $\sigma_3\chi_\pm = \pm\chi_\pm$ and $\chi^\dagger_\pm\chi_\pm = 1$. The reader might want to explicitely verify all orthogonality properties; we now come to the important conclusion that

$$\overline{u_{n,m,k;\alpha}}^T\gamma^0 u_{n,m,k;\beta} = \delta_{\alpha\beta},\ \overline{v_{n,m,k;\alpha}}^T\gamma^0 v_{n,m,k;\beta} = -\delta_{\alpha\beta},\ \overline{u_{n,m,k;\alpha}}^T\gamma^0 v_{n,m,k;\beta} = 0.$$

The reader should notice that changing n, m but keeping k fixed determines precisely the same subspaces. Hence, *assuming* that nature forbids superpositions of states

$$au_{n,m,k;\alpha} + bv_{n,m,k;\beta}$$

one obtains a canonical, *Lorentz invariant* probability interpretation by means of the "Wick rotation"

$$\langle u_k|u'_k\rangle_p = \overline{u}^T_k\gamma^0 u'_k,\ \langle v_k|v'_k\rangle_p = -\overline{v}^T_k\gamma^0 v'_k,\ \langle u_k|v_k\rangle_p = 0$$

and we shall work out this idea further on in the next section. The generalization towards tensor products is again obvious.

6.3 Spin-$\frac{1}{2}$ particles.

We shall now treat the theory of spin-$\frac{1}{2}$ particles in full detail in a similar way as we did for spin-0 particles. First, we construct a canonical Schrodinger equation to define the propagator as well as fundamental Fourier waves (particle notions) and then proceed with the generalization of the probability interpretation explained in the previous chapter. Again, we completely abandon the "quantum field" viewpoint here and derive the entire theory from a novel implementation of well known physical principles. That is, we aim further and try do derive well known results of the free theory in flat Minkowski without ever speaking about Hamiltonians, field operators, action principles and so on. So, what I propose is a nouvelle cuisine for quantum theory: a purely geometrical framework with a realist ontology. Since we work in a general curved space time, we need a Lorentz connection $\omega^a_{\mu\,b}$ and the reader may verify that the associated spin connection is given by

$$\omega^k_{\mu j} = i\omega_{\mu ab}(\mathcal{J}^{ab})^k_j$$

where the $k, j : 0\ldots 3$ denote spinor indices and the generator of spin rotations J^{ab} has been introduced before. Therefore, the spin covariant derivative looks like

$$\nabla^s_\mu = \nabla_\mu + \omega^a_{\mu b} + i\omega_{\mu ab}(\mathcal{J}^{ab})^k_l$$

where $\omega^a_{\mu b}$ is given by

$$\omega^a_{\mu b} = -e^\nu_b\nabla_\mu e^a_\nu$$

and one may directly verify the antisymmetry property

$$\omega_{\mu ab} = -\omega_{\mu ba}.$$

Coming back to the main line of our story, we would like to introduce a function $\phi_m(x, k^a, y)^i_{j'}$ where primed indices again refer to y and m is the mass of the particle such that

$$W(x,y)^i_{j'} = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0)\phi_m(x, k^a, y)^i_{j'}$$

denotes some "propagator". Upper indices refer to spin properties of a vector while lower indices to those of a covector and moreover, annihilation and creation always go in a vector-covector pair. We *agree* that particle creation corresponds to a covector in the propagator while antiparticle creation corresponds to a vector. So, the above propagator *might* signify the amplitude for an anti-particle to be created at x, with spin component i, and be annihilated at y with spin component j'. Likewise, we should have an amplitude $\psi_m(x, k^a, y)^{j'}_i$ to denote the "propagation" of a particle from x, with spin i towards y with spin j'. Now, one *might* also interpret the amplitude $\phi_m(x, k^a, y)^i_{j'}$ as a number associated to the birth of an anti-particle at x and a particle at y, where at the moment y is born as an event, the anti-particle does not belong to the universe anymore. Such an interpretation is forbidden in quantum field theory on Minkowski given that it would violate charge conservation. In our framework on a general space time however, there is no conserved current, and therefore no a-priori reason why an anti-particle cannot be born out of a particle. We shall see that such effect really does exist, but it is very small and depends upon the strength of the gravitational field. To fix the propagator, we will proceed in the same way as for the particle of zero spin, arguing what the coincidence limit $\phi_m(x, k^a, x)$ should look like and then solve for the entire space time by using the Schrodinger equation associated to (geodesic) paths γ:

$$\frac{D'^s}{dt}\phi(x, k^a, \gamma(t))^i_{j'} = i\dot{\gamma}^\mu(t)k_\mu(t)\phi(x, k^a, \gamma(t))^i_{j'}.$$

Indeed, the latter is our replacement for the Dirac equation and we will study its solution later on. Let us start by the most straightforward principles of which the first does not necessarily need to be satisfied in a general curved space time but it is for sure true in Minkowski due to spatial homogeneity. That is, the coincidence limit $\phi_m(x, k^a, x)^i_j$ does not depend upon x and it transforms in the adjoint representation of $SL(2, \mathbb{C})$ meaning that

$$\phi_m(x, (\Lambda k)^a, x) = \Lambda^{\frac{1}{2}}\phi_m(x, k^a, x)\Lambda^{-\frac{1}{2}}.$$

The latter requirement, taken together with our generalized Schrodinger equation, ensures that the definition of the propagator shall be independent of the Lorentz frame chosen. Both conditions, taken together, imply that our only building blocks are $k_a\gamma^a$ and $m1$ and since we only work with on shell momenta, $\phi_m(x, k^a, x)$ may be chosen of the form $\alpha(k_a\gamma^a + \beta m1)$ where α and β are complex numbers: the mass dimension should be zero so that the limit of zero mass gives a nonvanishing result. Now, we arrive at our third and most important principle which says that the creation and annihilation of both a particle and antiparticle with the same four momentum should give a vanishing amplitude on shell when summing over all internal spin degrees of freedom, that is:

$$\phi_m(x, k^a, x)\psi_m(x, k^a, x) = \psi_m(x, k^a, x)\phi_m(x, k^a, x) \sim (k^2 - m^2).$$

This gives that $\phi_m(x, k^a, x) = \alpha(k_a\gamma^a \pm m1)$ and $\psi_m(x, k^a, x) = \alpha'(k_a\gamma^a \mp m1)$. Finally, we have our fourth condition which I call the positive energy condition, which says that

$$\frac{1}{4}\mathrm{Tr}(\gamma^0\phi_m(x, k^a, x)) = k^0 = \frac{1}{4}\mathrm{Tr}(\gamma^0\psi_m(x, k^a, x))$$

which states that the energy of a particle equals the zero'th component of its momentum vector. This further limits $\alpha = \alpha' = 1$; so we are left with

$$\phi_m(x, k^a, x) = (k_a\gamma^a \pm m1) = \pm 2mP_\pm(k), \quad \psi_m(x, k^a, x) = (k_a\gamma^a \mp m1) = \mp 2mP_\mp(k)$$

and we now agree that the particle propagator $\psi_m(x, k^a, x)$ should come with positive mass meaning that our particle wave vectors are given by $u_{n,m,k;\pm}$ whereas the anti-particle wave-covectors are given by $\overline{v_{n,m,k;\pm}}^T\gamma^0$ as we shall discuss in more detail later on. This ends our discussion of the coincidence limit; our novel principles have brought us to matrices which equal $\pm 2mP_\pm(k)$ giving the propagator a dimension of mass^3 in contrast to the propagator for a spin-0 particle.

Now, we come to the integration of the Schrodinger equation: the latter is easy and natural and before giving its solution, denote by $(\Lambda^{\frac{1}{2}}(x, w))_i^{j'}$ the *spin* holonomy attached to the geodesic from x to $y = \exp_x(w)$ determined by tangent vector w and similarly for $(\Lambda(x, w))_a^{b'}$ the associated Lorentz holonomy. Thus given our initial conditions, the solutions to the "equation of motion" read

$$\tilde{\phi}_m(x, k^a, w)^i_{j'} = (k_a(\gamma^a)^i_r - m\delta^i_r)(\Lambda^{-\frac{1}{2}}(x, w))^r_{j'}\tilde{\phi}(x, k^a, w)$$

and

$$\tilde{\psi}_m(x, k^a, w)^{j'}_i = (\Lambda^{\frac{1}{2}}(x, w))^{j'}_r(k_a(\gamma^a)^r_i + m\delta^r_i)\tilde{\phi}(x, k^a, w).$$

We will now prove a remarkable property which shows that quantum causality, as it is usually understood, holds for this propagator. Indeed, the very structure of our formulae suggests that there may be a relationship between $\tilde{\psi}_m(x, k^a, w)$ and $\tilde{\phi}_m(y, k^{a'}_{\star w}, -w_{\star w})$ where, as before, $k^{a'}_{\star w} = (\Lambda(x, w))^{a'}_b k^b$. Indeed, a small calculation reveals that

$$\begin{aligned}
\tilde{\phi}_m(y, k^{a'}_{\star w}, -w_{\star w})^{j'}_i &= (k_b((\Lambda(x, w))^{-1})^b_{a'}(\gamma^{a'})^{j'}_{k'} - m\delta^{j'}_{k'})(\Lambda(x, w)^{\frac{1}{2}})^{k'}_i\tilde{\phi}(y, k^{a'}_\star, -w_{\star w}) \\
&= (\Lambda^{\frac{1}{2}}(x, w))^{j'}_l\left(k_b(\gamma^b)^l_i - m\delta^l_i\right)\overline{\tilde{\phi}(x, k^a, w)}
\end{aligned}$$

where we have used on the first line that $\Lambda^{\frac{1}{2}}(x, w) = (\Lambda^{\frac{1}{2}}(y, -w_{\star w}))^{-1}$; in the second line, we used covariance of the gamma matrices under joint spin and Lorentz transformations as well as the previous established formula for $\tilde{\phi}(x, k^a, w)$. Now, the way in which this formula becomes useful is by means of the particle and antiparticle propagators:

$$W_p(x,y)_i^{j'} = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0)\psi_m(x, k^a, y)_i^{j'}$$

and

$$W_a(x,y)_{j'}^{i} = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0)\phi_m(x, k^a, y)_{j'}^{i}$$

where, as before,

$$\psi_m(x, k^a, y)_i^{j'} = \sum_{w:\exp_x(w)=y} \tilde{\psi}_m(x, k^a, w)_i^{j'}$$

and likewise for $\phi_m(x, k^a, y)$. Indeed,

$$\begin{aligned} W_a(y,x)_i^{j'} &= \sum_{w:\exp_x(w)=y} \int_{T^\star \mathcal{M}_y} \frac{d^4k_{\star w}}{(2\pi)^3} \delta(k_{\star w}^2 - m^2)\theta(k_{\star w}^0)\tilde{\phi}_m(y, k_{\star w}^{a'}, -w_{\star w})_i^{j'} \\ &= \sum_{w:\exp_x(w)=y} (\Lambda^{\frac{1}{2}}(x,w))_l^{j'} \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0) \left(k_b(\gamma^b)_i^l - m\delta_i^l\right) \overline{\tilde{\phi}(x, k^a, w)} \end{aligned}$$

and we concentrate now on points $x \sim y$ which are exclusively connected by spacelike geodesics. In that case, we could write

$$\tilde{\phi}(x, k^a, w) = e^{ik_a w^a}$$

where w^a is the spacelike tangent at x to the geodesic connecting x with y. Choosing now for each term a different Lorentz frame at x such that the vector w is parallel to the three axis e_3; we perform, as before, a reflection around w given by $k^3 \to -k^3$ to obtain

$$(\Lambda^{\frac{1}{2}}(x,w))_l^{j'} \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0) \left(k_b(\gamma^b)_i^l - 2k_3(\gamma^3)_i^l - m\delta_i^l\right) e^{ik_3 w^3}$$

where $e^{ik_3 w^3} = \tilde{\phi}(x, k^a, w)$. Summing this formula with the corresponding part of $W_p(x,y)_i^{j'}$ in the same frame gives

$$(\Lambda^{\frac{1}{2}}(x,w))_l^{j'} \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0) \left(2 \sum_{j=0\ldots 2} k_j(\gamma^j)_i^l\right) e^{ik_3 w^3}$$

which is immediately seen, due to the antisymmetry of some part of the integrand under $k_1, k_2 \to -k_1, -k_2$, to reduce to

$$(\Lambda^{\frac{1}{2}}(x,w))_l^{j'} (\gamma^0)_i^l i \int_{T^\star \mathcal{M}_x} \frac{d^3k}{(2\pi)^3} e^{ik_3 w^3}$$

where the last integral equals $\delta^3(w^a)$ which proves that

$$W_p(x,y)_i^{j'} + W_a(y,x)_i^{j'} = 0$$

in any local Lorentz frame. This constitutes a proof of the well known statement that the amplitude for a particle with spin i to travel form x to y and be annihilated with spin j' equals the amplitude for an antiparticle with spin j' to travel from y to x where it is annihilated with spin i. The very minus sign reveals that spin-$\frac{1}{2}$ particles are fermions, meaning that exchanging two particles comes with a minus sign; this constitutes the proof of the spin statistics theorem in our setting at least for spin-0 and spin-$\frac{1}{2}$ particles. As before, we can now define the Feynman propagator for particle propagation $\Delta_{F,p}(x,y)_i^{j'} = W_p(x,y)_i^{j'}(x,y)$ if $y \in J^+(x) \setminus J^-(x)$, $-W_a(y,x)_i^{j'}$ if $y \in J^-(x) \setminus J^+(x)$, $\frac{1}{2}\left(W_p(x,y)_i^{j'}(x,y) - W_a(y,x)_i^{j'}\right)$ if $y \in J^+(x) \cap J^-(x)$ and $W_p(x,y)_i^{j'} = -W_a(y,x)_i^{j'}$ if $x \sim y$. We also could define a Feynman propagator for anti-particle propagation as $\Delta_{F,a}(x,y)^i_{j'}$ as before by replacing p with a and the reader immediately notices that $\Delta_{F,a}(x,y)^i_{j'} = -\Delta_{F,p}(y,x)^i_{j'}$ as is the case in Minkowski quantum field theory. The same comment as for the spin-0 case applies to the determination of $\Delta_{F,p}(x,y)_i^{j'}$ in case $y \in J^+(x) \cap J^-(x)$; we shall adequately correct this expression in the next chapter. This concludes our discussion of the free Fermi theory and the reader notices that all salient features of the standard Minkowski theory have been saved. We can now, as in the previous case suggest gravitational modifications of the two point function for causally related points such that causality remains valid but the singularity structure of the propagator changes. The way to do this is exactly identical to the one suggested before for the scalar two point function and therefore, we do not have to discuss this further on here. Evidently, our propagator does not satisfy the Dirac equation anymore and the reader is invited to investigate if the latter would still hold in the coincidence limit $y \to x$ just as the Klein Gordon equation did for the scalar two point function.

We end this section by giving all relevant details of the probability interpretation. Basically, we proceed here in the same way as in the previous chapter regarding the spin zero particle; it is obvious that the correct particle notions at x are given by

$$u_{n,m,k;\pm}(x,y) = \sum_{w:\exp_x(w)=y} \Lambda^{\frac{1}{2}}(x,w) u_{n,m,k;\pm} e^{ik_a w^a}$$

where $n, m, k \in T^\star\mathcal{M}$ and for $v_{n,m,k;\pm}(x,y)$, the latter is given by

$$v_{n,m,k;\pm}(x,y) = \sum_{w:\exp_x(w)=y} \Lambda^{\frac{1}{2}}(x,w) v_{n,m,k;\pm} e^{-ik_a w^a}.$$

The reason for the minus sign in the exponential is due to the fact that it is the *covector*

$$\overline{v_{n,m,k;\pm}(x,y)}^T \gamma^0$$

which represents an anti-particle. Therefore, superpositions of u and v waves are physically meaningless which expresses itself mathematically by the fact that v waves get annihilated by the particle propagator in Minkowski. Unlike in Minkowski, it is possible in a general curved space time that the propagator, with respect to a spatial hypersurface Σ to the "future", in the process sense, of the point of creation x contains $v_{n,m,k;\alpha}(z,z')$ modes with respect to the point of *creation* z. Indeed, this is a novel effect in our physics which we shall explain intuitively now, but which shall be made clear later on; usually, the process of propagation is associated with a particle being created at x and being annihilated at y; but then, why would we not write down something like $W(x,y)^{ij'}$? The point is that one cannot find a suitable mathematical object akin to $(\gamma^a)^i_j$ with two contravariant spinor indices. But then, nothing forbids us to "associate" $W_p(x,y)^{j'}_i$ to an amplitude correlating the birth of a particle at x with an anti-particle being *created* at y, together with a measurement apparatus showing a "reaction" because of this. Typically, one would expect the state of such measurement apparatus to change in way as to have approximate charge conservation during the entire process: that is, the measurement appartus increases its particle number by two. This is new physics and I have described schemes in this direction already in previous publications; it is somehow amusing to see that this issue comes back under a different guise. Of course, this interpretation can also hold for "virtual" process in interactions given that the integration of the interaction vertex over (a portion of) space time allows for u,v couplings to occur. This is to be expected given that "propagation" acting on a wave function constitutes a process of annihilation and recreation, which is different than a process of creation and annihilation, and has therefore a similar status than measurement has. The correct candidate for a scalar product in Minkowski being given by

$$\begin{aligned}
\langle u_{n,m,k;\alpha}(x,y)|u_{n',m',k';\beta}(x,y)\rangle &= (2\pi)^3\frac{k^0}{m}\delta(\vec{k}-\vec{k}')\langle u_{n,m,k;\alpha}|u_{n',m',k;\beta}\rangle_p \\
\langle v_{n,m,k;\alpha}(x,y)|v_{n',m',k';\beta}(x,y)\rangle &= (2\pi)^3\frac{k^0}{m}\delta(\vec{k}-\vec{k}')\langle v_{n,m,k;\alpha}|v_{n',m',k;\beta}\rangle_p \\
\langle u_{n,m,k;\alpha}(x,y)|v_{n',m',k';\beta}(x,y)\rangle &= 0
\end{aligned}$$

and as before, we look for a space time realization of this. The reader should make the following exercise prior to proceeding, which is that

$$\begin{aligned}
u^\dagger_{n,m,k;\alpha}u_{n',m',k;\beta} &= \frac{k^0}{m}\langle u_{n,m,k;\alpha}|u_{n',m',k;\beta}\rangle_p \\
v^\dagger_{n,m,k;\alpha}v_{n',m',k;\beta} &= \frac{k^0}{m}\langle v_{n,m,k;\alpha}|v_{n',m',k;\beta}\rangle_p \\
u^\dagger_{n,m,k;\alpha}v_{n,m,S(k);\beta} &= 0
\end{aligned}$$

where in the last line $S(k)$ denotes the spatial reflection with respect to the given vierbein. These relations constitute the replacement, with respect to the standard Euclidean product defined by the vierbein, of the orthogonality relations of the $o_{n,m,k;\alpha}$ vectors with respect to the indefinite product. The reader should now use these to show that

$$\langle o_{n,m,k;\alpha}(x,y)|p_{n',m',k';\beta}(x,y)\rangle = \int_{\Sigma} d^3\vec{y}\; o_{n,m,k;\alpha}(x,y)^{\dagger} p_{n',m',k';\beta}(x,y)$$

where Σ is an inertial Cauchy surface in Minkowski with a unit normal given by e_0. One immediately verifies that in Minkowski

$$\int_{\Sigma} d^3\vec{y}\; W_p(y,z)\gamma^0 u_{n,m,k;\alpha}(x,y) = u_{n,m,k;\alpha}(x,z)$$

and likewise so for $\overline{v_{n,m,k;\alpha}(x,y)}^T \gamma^0$ and $W_a(y,z)$.

These formula suggest the appropriate definition of propagation, detection and traversing on a general curved space time. Here, we will only treat the detection process since the remainder is analogous to the previous chapter and I insist to highlight the difference in interpretation hinted at above given that "detection" may give rise to *two* different situations which need to be treated orthogonally: (a) annihilation of a particle (anti-particle) or (b) *creation* of an anti-particle (particle). As usual, we consider a world tube W_{Σ} cut out by an "irreducible" component of the measurement apparatus and consider the slice Σ_t of the actual now at the moment the point z of annihilation or creation is born. Then, as before, we argue that we can match the particle spinor wave, born at x, on Σ', where $\Sigma_{t-\delta} \subset \Sigma'$ and the latter is a complete spacelike hypersurface with respect to z, with a unique spinor wave defined with respect to z. Now, we split the latter into two pieces, one of the form $u(z,y)$ and another one of the form $v(z,y)$. The respective weights we are looking for then are

$$d_{W_{\Sigma},(p,a)}(\Psi,z;\delta) = \int_{\Sigma_{t-\delta}} d^3y\, \sqrt{h(y)}u(z,y)^{\dagger}u(z,y)$$

for particle annihilation at z and

$$d_{W_{\Sigma},(p,c)}(\Psi,z;\delta) = \int_{\Sigma_{t-\delta}} d^3y\, \sqrt{h(y)}v(z,y)^{\dagger}v(z,y)$$

for anti-particle creation at z. Here, the scalar products are defined with respect to a vierbein such that e_0 is the unit normal to $\Sigma_{t-\delta}$: the residual local $SU(2)$ symmetry leaving the scalar product invariant. We finish this chapter by giving a treatment for spin-one particles.

6.4 Spin 1 "gauge" particles.

In contrast to what one may expect, the two point function for massless spin-1 particles is extremely easy to guess, even when they carry another charge such as is the case for non-abelian gauge theories. We do not speak anymore in terms of gauge transformations which were necessitated by the quantum field viewpoint but we derive the main formula for the two point function and the Feynman propagator from two simple demands. The reader should appreciate the plain simplicity of the construction as the computation of the two point function for non-abelian gauge fields in standard quantum field theory is a matter of laborious work, the proof that gauge particles satisfy bosonic statistics being evident. Hence, we are interested in computing a quantitity

$$W_{\mu\nu'}^{\alpha\beta'}(x,y) = \int_{T^\star\mathcal{M}_x} \frac{d^4k}{(2\pi)^3}\delta(k^2)\theta(k^0)\psi(x,k^a,y)_{\mu\nu'}^{\alpha\beta'}$$

and again, we derive the correct form of the two point function. Note here that our group transformations are global transformations and therefore do *not* depend upon the space time point; so, the indices α, β' stands for the adjoint representation of the compact simple Lie group whose algebra is defined by

$$[t_\alpha, t_\beta] = i f_{\alpha\beta}^{\gamma} t_\gamma$$

where $f_{\alpha\beta\gamma} = f_{\alpha\beta}^{\delta} g_{\delta\gamma}$ is totally antisymmetric and the positive definite invariant Cartan metric is given by $g_{\alpha\beta}$. The fact that we do not make any distinction between covariant and contravariant vectors is due to the possibility to raise and lower indices with both metrics $g_{\mu\nu}$ and $g_{\alpha\beta}$. Let us study the coincidence limit $y \to x$ of $\psi(x,k^a,y)_{\mu\nu'}^{\alpha\beta'}$ first. Since there is no mass parameter, the only object of mass dimension zero which we can write down is a multiple of $g_{\mu\nu}g^{\alpha\beta}$, the only other term one can write down on shell has mass dimension squared and is given by a multiple of $k_\mu k_\nu g^{\alpha\beta}$. So here, we make our first law, $\psi(x,k^a,y)_{\mu\nu'}^{\alpha\beta'}$ has mass dimension zero and we can absorb any positive, real constant in the definition of the Cartan metric; so we obtain that

$$\psi(x,k^a,x)_{\mu\nu}^{\alpha\beta} = -g_{\mu\nu}g^{\alpha\beta}$$

where the minus sign originates from the fact that the vectors of helicity ± 1 should come with a plus sign. Writing out our Schrodinger equation is extremely easy

$$\frac{D'}{dt}\psi(x,k^a,\gamma(t))_{\mu\nu'}^{\alpha\beta'} = i\left[\dot\gamma(k)\right](t)\psi(x,k^a,\gamma(t))_{\mu\nu'}^{\alpha\beta'}$$

and when $\gamma(t)$ is a geodesic, the solution is given by

$$\psi(x,k^a,y)_{\mu\nu'}^{\alpha\beta'} = -g_{\mu\nu'}(x,y)\phi(x,k^a,y)g^{\alpha\beta'}$$

where $g_{\mu\nu'}(x,y)$ denotes the parallel transport of the metric along the geodesic. The latter can be written as a composition of the Van Vleck matrix with Synge's function and since the metric is covariantly constant one has that $g_{\mu\nu'}(x,y) = g_{\nu'\mu}(y,x)$. In case multiple geodesics join x and y, we obtain that

$$W_{\mu\nu'}^{\alpha\beta'}(x,y) = -\sum_{w:\exp_x(w)=y} g_{\mu\nu'}(x,w)g^{\alpha\beta'}W(x,w)$$

where $W(x,w) = \int \frac{d^3k}{(2\pi)^3}\delta(k^2 - m^2)\theta(k^0)e^{ik_a w^a}$, which shows that the two point[3] function for spin-1 particles transforming under a global, compact symmetry group is determined by the two point function of the scalar theory, a transporter and the Cartan metric. From our previous results and the symmetry of the transporter as well as the Cartan metric it follows that

$$W^{\alpha\beta'}_{\mu\nu'}(x,y) = W^{\beta'\alpha}_{\nu'\mu}(y,x)$$

for $x \sim y$ so that our theory satifies quantum causality and has bosonic exchange properties. Clearly, massless spin-1 particles are their own antiparticles as there exists only one two point function and not two. Let us better understand the magic which happened here: instead of following the quantization procedure of a theory with a local gauge symmetry and impose a gauge, we simply took the transformation group of the quantum numbers to be a *global* one. This is a meaningful point of view since those numbers themselves do not correspond to any force field, they are attributes of particles which is something different. It is possible to introduce classical gauge fields and introduce a dynamical gauge bundle so that we have to use the holonomies associated to this gauge field. This would be new physics and I hold it entirely possible that the future may lead us there; for now, we obtain on one sheet of paper a result which can be found in every textbook and which requires a long introduction to derive. As mentioned in the previous section, the structure constants $f_{\alpha\beta\gamma}$ and Cartan metric $g_{\alpha\beta}$ will be used to build interactions, everything is perfectly consistent with quantum chromo dynamics and quantum electro dynamics. The Feynman propagator $\Delta^{\alpha\beta'}_{F\,\mu\nu'}(x,y)$ has precisely the same prescription as is the case for spin-0 particles, which concludes the discussion for spin-1 particles. We now come to the discussion of Faddeev-Popov ghosts; first, let us ask ourselves why we insist upon spin-1 particles to transform in the adjoint representation and spin-$\frac{1}{2}$ in the defining one. The general reason is that it allows us to write down intertwiners of the kind

$$(\gamma^a)^i_j e^\mu_a(x)(t_\alpha)^m_n$$

and as the reader may verify, this is the only way to couple spin-1 and spin-$\frac{1}{2}$ particles. This leaves us with the question of coupling spin-0 particles to spin-1, the relevant intertwiner is given by

$$f_{\alpha\beta\gamma}\nabla^\mu$$

where the derivative acts on the gauge boson propator only and therefore these spin-0 particles should transform as a vector in the adjoint representation; moreover they should have fermionic exchange properties since $f_{\alpha\beta\gamma}$ is totally antisymmetric. Such particles could be coupled to ordinary spin-$\frac{1}{2}$ matter though by means of the intertwiner

$$f_{\alpha\beta\gamma}(t^\alpha)^m_n\delta^i_j$$

[3]The fact that we need the Cartan metric for the construction of the two point function is precisely the reason why the Lie group had to be compact and simple in the first place.

and it is very easy to derive the unique propagator, using complex numbers only and having the correct transformation properties,

$$W^{\alpha\beta}(x,y) = g^{\alpha\beta} W(x,y)$$

which suggests that we made an error since the associated particles behave like massless bosons instead of fermions. It is, however, very easy to correct for this deficit by constructing propagators with the Grassmann numbers; the resulting expressions being

$$W_p(x,y) = \theta(x)\overline{\theta(y)} g^{\alpha\beta} W(x,y)$$

and

$$W_a(x,y) = \overline{\theta(x)}\theta(y) g^{\alpha\beta} W(x,y).$$

We now finish this chapter by discussing particle notions as well as the probability interpretation.

From our Schrodinger equation, it follows that particles born at x are determined by

$$\phi_{m,k;\nu'}(x,y) = \sum_{w:\exp_x(w)=y} (\Lambda(x,w)^{-1})^{\mu}_{\nu'} m_\mu e^{ik_a w^a}$$

where $k^\mu m_\mu = 0$. Here, the reader notices that the longitudonal modes k_μ can come to life if there are different geodesics connecting x with y since

$$-g_{\mu'\nu'}(y)\Lambda(x,w)^{\mu'}_{\mu} k^\mu \Lambda(x,v)^{\nu'}_{\nu} k^\nu < 0.$$

This is fairly problematic as the resulting norm becomes of indefinite nature, something we should wish to avoid. We have a similar problem as in the case of spin-$\frac{1}{2}$ particles since there is no reason *why* the restriction of a particle wave, born at x, to some Σ' could be written as the restriction of a particle wave annihilated at z. In the case of spin-$\frac{1}{2}$ particles, we could understand this situation by means of the anti-particles *created* at z but there is no such luxury at hand here. We shall be ruthless here and eliminate the null modes as well as our problem of non-compatible particle notions by employing the $SO(3)$ class of vierbeins associated to the surface Σ'. Indeed, the latter determines a preferred timelike vector field e_0 and therefore a preferred notion of helicity vectors belonging to $T^\star\Sigma'$; more precisely, we define the equivalent of the T_{x,e_0} mapping in the spin-0 case by means of

$$T_{x,e_0}(\Lambda(x,w)^{-1})^{\mu}_{\nu'} m_\mu e^{ik_a w^a} = \sqrt{k^{0'}} P_{k^{a'}=\Lambda(x,w)^{a'}_{b} k^b;e_0}((\Lambda(x,w)^{-1})^{\mu}_{\nu'} m_\mu) e^{ik_a w^a}$$

where $k^{\mu'} m_{\mu'} = 0$, $k^{0'}$ is the component of $k^{a'}$ with respect to e_0 and $P_{k^{a'};e_0}$ projects a covector on the space orthogonal to $k_{\mu'}$ and $e_{0\mu'}$. Here,

$$(\Lambda(x,w)^{-1})^{\mu}_{\nu'} m_\mu e^{ik_a w^a}$$

is to be seen as a covector-valued function in the complexification of $T_\star \mathcal{M}_y$, where $y = \exp_x(w)$, in the vector variable $w^a \in T^\star \mathcal{M}_x$. Hence we define,

$$P_{x,e_0}\left(\sum_{w:\exp_x(w)=y} \int d^3k \widehat{\psi}(k) \sum_{m^i : m^i_\mu k^\mu = 0, i=1\ldots 3} (\Lambda(x,w)^{-1})^\mu_{\nu'} m_\mu e^{ik_a w^a}\right) =$$

$$\sum_{w:\exp_x(w)=y} \int d^3k\, \widehat{\psi}(k) \sum_{m^i : m^i_\mu k^\mu = 0, i=1\ldots 3} P_{\Lambda(x,w)^{a'}_b k^b; e_0}((\Lambda(x,w)^{-1})^\mu_{\nu'} m_\mu) e^{ik_a w^a}$$

and we suggest now that under reasonable conditions the function space

$$\mathcal{S}_x(\Sigma') = \{y \in \Sigma' \to P_{x,e_0 \perp \Sigma'} \Phi_x(y) | \Phi_x \text{ is a spin one wave defined at } x\}$$

is independent of x. That is,

$$\mathcal{S}_x(\Sigma') = \mathcal{S}_z(\Sigma').$$

Under these conditions, we find some Φ_z such that $y \in \Sigma' \to P_{z,e_0 \perp \Sigma'} \Phi_z(y)$ equals $y \in \Sigma' \to P_{x,e_0 \perp \Sigma'} \Phi_x(y)$. The corresponding probability interpretation then being given by the scalar product

$$\langle \Psi_z(y) | \Phi_z(y) \rangle = -\int_{\Sigma'} d^3y\, \sqrt{h(y)} g^{\mu'\nu'}(y)\, \overline{(T_{z,e_0 \perp \Sigma'} \Psi_z(y))_{\mu'}} (T_{z,e_0 \perp \Sigma'} \Phi_z(y))_{\nu'}$$

which coincides with the version in Minkowski. The reader may check from here that the formula for propagation is given by

$$(P_\Sigma(\Psi_x))_{\alpha''}(z) = -\int_\Sigma d^3y\, \sqrt{h(y)} g^{\mu'\nu'}(y)\, (T_{x,e_0 \perp \Sigma} \Psi_x(y))_{\mu'} \overline{(T_{z,e_0 \perp \Sigma} W(z,y)_{\alpha''\nu'})}$$

and we shall have more to say about this in chapter eight; the treatment of gravitons also being postponed to that chapter.

Chapter 7

Old problems requiring new physics.

In this chapter, we shall work our way towards an appropriate definition of the interacting theory, everything we said so far relating to the free theory. In the best quantum field theory books, one formally derives constraints on the possible interactions which leads one to the field picture and the *dogma* of relativistic causality. The latter, which says that physically realistic observables, located at spatially seperated events, must commute is however totally unnecessary: the commuting of the field operators for bosonic particles, which is mandatory for a Lorentz covariant scattering matrix, is by no means a sign that all realistic observables should commute. In particular, it would imply that the projection operator on the distributional state of a particle created at x is not an observable and neither is the propagator. However, the field picture has many more problems given that its defining constituents are not well defined as is the case in our approach so far which matches field theory exactly on a Minkowski background. So far, we have argued that the "correct" two point function for a spin-0 particle in a general curved background space time is given by

$$W(x,y) = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3} \delta(k^2 - m^2)\theta(k^0)\phi(x, k^a, y)$$

where

$$\phi(x, k^a, y) = \sum_{w^a \in T\mathcal{M}_x : \exp_x(w) = y} e^{ik_a w^a}$$

where the exponential map is defined as usual. In Minkowski space time, this expression is given by

$$W(x,y) = \int_{T^\star \mathcal{M}_x} \frac{d^3k}{2(2\pi)^3\sqrt{\vec{k}^2 + m^2}} e^{ik_a(y^a - x^a)}$$

which may be computed further by making a distinction between the spacelike, null, and timelike case. For spacelike $y^a - x^a$, one may choose the Lorentz frame such that $y^a - x^a = \sqrt{(y-x)^2} e_3$ resulting in

$$W(x,y) = \int_{T^\star \mathcal{M}_x} \frac{d^3k}{2(2\pi)^3\sqrt{\vec{k}^2+m^2}} e^{ik^3\sqrt{(y-x)^2}} = \frac{1}{8\pi^2}\int_0^\infty r\,dr \int_{-\infty}^{+\infty} dk \frac{1}{\sqrt{k^2+r^2+m^2}} e^{ik\sqrt{(y-x)^2}}$$
$$= \frac{\infty}{4\pi}\delta(\sqrt{(y-x)^2})$$

where we performed the r integration prior to the k integration which does not only give the wrong answer but also shows that the original "integral" cannot be computed by appealing to Fubini's theorem in this coordinate system and therefore, the Lebesgue integral does not exist. Indeed, *no* momentum integral in standard field theory exists in the sense of Lebesgue as one considers integration of widely fluctuating functions which do not go sufficiently fast to zero at infinity so that the positive and negative, real and imaginary parts of the integrand do not give finite integrals by themselves. In fact, there does not exist a straightforward way how to define this expression. It does exist as a bi-distribution however:

$$W(f,g) = \int_{\mathcal{M}} dx \int_{T^\star\mathcal{M}_x} \frac{d^3k}{2(2\pi)^3\sqrt{\vec{k}^2+m^2}} \int_{\mathcal{M}} dy\, e^{ik_a(y^a-x^a)} f(x)g(y)$$

or

$$W(f,g) = \int_{\mathbb{R}^3} \frac{d^3k}{2(2\pi)^3\sqrt{\vec{k}^2+m^2}} \int_{\mathcal{M}\times\mathcal{M}} dx\,dy\, e^{ik_a(y^a-x^a)} f(x)g(y)$$

since all tangent spaces are isomorphic and both definitions agree for smooth test functions f, g of compact support, where the integrals are taken in the order indicated in the above expressions. In the literature $W(x,y)$ is often presented as a smooth function $\tilde{W}(x,y)$ with a delta distribution on the light-cone; this *representation* however holds only when contractions with Schwartz functions f, g are made, in either

$$W(f,g) = \int_{\mathcal{M}\times\mathcal{M}} f(x)g(y)\tilde{W}(x,y)$$

and the reader may easily find out that $\tilde{W}(x,y)$ is given by special Bessel functions. Indeed, for x, y spacelike, we have that

$$\tilde{W}(x,y) := \frac{m}{\sqrt{(x-y)^2}4\pi^2}\int_0^\infty \frac{dk}{\sqrt{k^2+1}} k\,\sin(km\sqrt{(x-y)^2})e^{-\epsilon k^2} = \frac{m}{\sqrt{(x-y)^2}4\pi^2} K_1(m\sqrt{(x-y)^2})$$

as a formal expression. Indeed, it is fairly easy to check by means of partial integration that $K_1(z)$ satisfies Bessels equation

$$z^2\ddot{K}_1(z) + z\dot{K}_1(z) - (z^2+1)K_1(z) = 0$$

with appropriate boundary conditions. However, $\tilde{W}(x,y)$ is not absolutely integrable given that it does not vanish at infinity (it remains constant on spacelike hyperbolae). Therefore, one cannot extend the definition of $\tilde{W}(x,y)$ from

Schwartz functions to smooth L^2 functions of non-compact support as one would expect of realistic wave packages. However, it is worthwhile to mention that $K_1(z)$ diverges as $\frac{1}{z}$ at $z = 0$ and goes to zero as e^{-z} at $z = +\infty$. Indeed, coming back to the formal integral representation of $K_1(z)$ one may consider the effect of smoothening out with a Schwarz function of compact support as cutting off the integral at high momenta so that only the lower momenta count; this cutoff can be computed by means of a square contour in the complex plane which goes form 0 to R to $R+i\frac{\pi}{2}$ to $i\frac{\pi}{2}$ to 0 in the variable α where $k = \sinh(\alpha)$. The large vertical integral oscillates in a bounded way for large R but becomes irrelevant in the limit for R to infinity when smeared out with test functions while the vertical integral from 0 to $\frac{\pi}{2}$ is irrelevant. In this way, it can be shown that the Schwartz kernel $K_1(z)$ corresponds to the integral

$$K_1(z) = \int_0^\infty dt \, \cosh(t) e^{-\cosh(t) z}$$

and it is easy to see that this expression diverges as $\frac{1}{z}$ if z approaches zero. Hence, $K_1(z)$ is not uniformly bounded and therefore the best kind of duality one may set up is one of L^1_{loc} which are the absolutely integrable functions of compact support disjoint from the lightcone. To construct interactions, we need to calculate the Feynman propagator, which has been defined in full generality before, and has a formal integral representation on Minkowski as

$$\Delta_F(x, y) = \int \frac{d^4k}{(2\pi)^4} \frac{e^{ik_a(y^a - x^a)}}{k^2 - m^2 + i\epsilon}$$

where ϵ is a positive infinitesimal which may be taken to zero after all computations have been performed. Hence, integrals of the kind

$$\int dx \, dv \, dy \, dz \Delta_F(v, x) \Delta_F(w, x) \Delta_F(y, x) \Delta_F(z, x) f(v, w, y, z)$$

are well defined since all *logical* orders[1] of integration give the same result for entire complex analytic f with exponential falloff on the real section towards infinity. I am not aware if such special functions are really needed to obtain this result but it certainly allows one to appeal to the residue theorem for complex analytic functions in order to compute the result[2]. Even stronger, the above integral exists in a distributional sense for ordinary multidimensional plane waves as the reader may easily compute. Loops, however, are not well digested since one cannot give direct meaning to

$$\int_{\mathcal{M} \times \mathcal{M}} dx \, dy \, \Delta_F(x, y)^2 f(x, y)$$

[1] By logical, we mean any order which gives a well defined, finite, result.

[2] It would be interesting to have a result regarding the existence of the above integral if f were merely a Schwartz function.

with $f(x,y)$ an absolutely integrable function, not necessarily of compact support[3]. Alternatively, one might suggest that the correct expression to compute is given by

$$\int_{\mathcal{M}\times\mathcal{M}} dx\, dy\, \tilde{\Delta}_F(x,y)^2 f(x,y)$$

where $\tilde{\Delta}_F$ is the smooth distribution constructed before. Taking for $f(x,y) = e^{i(p+q)x+i(k+l)y}$, one notices that for $x \sim y$ the integral reduces to

$$\delta^4(k+l+p+q)\int_{(y-x)\,\text{spacelike}} d(y-x)\,\frac{m^2K_1^2(y-x)}{16\pi^4(y-x)^2}e^{i(k+l)(y-x)}$$

and by an appropriate change of variables

$$\begin{aligned}
t &= r\sinh\alpha\\
x &= r\cosh\alpha\,\sin\theta\sin\psi\\
y &= r\cosh\alpha\sin\theta\cos\psi\\
z &= r\cosh\alpha\cos\theta
\end{aligned}$$

which reduces the metric to

$$ds^2 = -dr^2 + r^2 d\alpha^2 - r^2\cosh^2\alpha d\theta^2 - r^2\cosh^2\alpha\sin^2\theta d\psi^2$$

and the volume form to

$$r^3\cosh^2\alpha\sin\theta d\alpha d\theta d\psi dr$$

one obtains that the latter integral reduces to

$$I(k+l) = \int dr\, d\alpha\, \cosh\alpha\,\frac{m^2K_1^2(r)}{4\pi^3 t_0\sinh\alpha_0}e^{it_0\cosh\alpha_0\, r\sinh\alpha}\sin(t_0\sinh\alpha_0 r\cosh\alpha)$$

[3]This follows easily from

$$\int d^4k d^4l \frac{\widehat{f}(-k-l,k+l)}{(k^2-m^2+i\epsilon)(l^2-m^2+i\epsilon)} = \int d^4r\, d^4k \frac{\widehat{f}(-r,r)}{(k^2-m^2+i\epsilon)((r-k)^2-m^2+i\epsilon)}$$

and for S-matrix elements $f(x,y) = e^{i(p+q).x}e^{i(r+s).y}$ where p,q are the on-shell incoming momenta and r,s the on-shell outgoing momenta so that

$$\widehat{f}(k,l) = \delta^4(k+p+q)\delta^4(l+r+s)$$

and therefore $\widehat{f}(-k-l,k+l) = \delta^4(k+l+r+s)\delta^4(p+q+r+s)$. It is easy to see that for generic absolutely integrable and differentiable $\widehat{f}$, the above integral is ill defined as

$$\int \frac{d^4k}{(k^2-m^2+i\epsilon)((r-k)^2-m^2+i\epsilon)}$$

is for generic r. This is most easily seen by application of the residue theorem and noticing that one is left with integrals of the kind

$$\int \frac{d^3k}{|k|^2+m^2}$$

which are linearly divergent.

where $k + l = t_0(\cosh \alpha_0, 0, 0, \sinh \alpha_0)$. It is clear, again, that this integral does not exist in the Lebesgue sense but one might wish to regard it as a distribution in $k + l$ where k, l are on-shell. As before, we may extract a kernel $\tilde{I}(k + l)$ in the dual sense and equate the integral to that expression. However, in general, one superposes wave packages of such on-shell plane waves which do not have compact support in momentum space and therefore, even this method will fall short in the end although it can be consistenly applied on a much higher level than is usually argued for in standard QFT textbooks. The lightcone will give trouble since there we do have a $\delta((y - x)^2)$ distribution in the formula for $\tilde{\Delta}_F(x, y)$ and the square of that is of course ill defined; one might, however, wish to ignore these contributions and effectively "cut out" the null cone. However, such procedure seems to be hard to motivate from a physical point of view and we will proceed in a way which makes the propagator well defined in the Lebesgue sense so that W and $\tilde{W}$ coincide and are smooth functions. In a general renormalization procedure, one takes "particular sums" of such nonsensical integrals, performs an associated ad-hoc analytic continuation, and makes the result finite by means of a redefinition of the bare parameters with an infinite amount. This happens, for example, in ϕ^4 field theory regarding corrections to the bare propagator; apart from the fact that this procedure is entirely arbitrary (but motivated by "physical intuition"), distinct "regularizations" might give different answers and this should not be the case for a physical theory unless there is a very good physical reason to prefer a particular regularization scheme over another. Moreover, this procedure splits theories into two categories: those to which *some* procedure of this kind can be applied, called the renormalizable theories, and those to which it cannot, the nonrenormalizable ones. The shear arbitrareness of the *infinite* renormalization procedure as well as the lack of a deep physical motivation behind it resulted in my thesis that interacting QFT on Minkowski does not exist and that gravitation had to play a fundamental role in making each Feynman diagram finite to the dismay of many field theorists I know of.

The reader notices that we had to twist ourselves into many small corners in order to give meaning to the two point function, the Feynman propagator and *some* "interaction" integrals. The results in the literature regarding renormalizability are alas much weaker than the kind of results we alluded to above; there, it is only shown that S-matrix elements in the distributional basis of plane Fourier waves can be given a distributional, perturbative, meaning due to renormalization. Nothing is said about physical, more general wave packages and not a single non-perturbative result is achieved. By this, I do not want to say that the results of 't Hooft and Veltman in the 1970's are virtually meaningless; they constituted a big step forwards in a time where everybody was concentrated upon Minkowski space time and the scattering matrix orginated by Wheeler. From a modern point of view, they do however fall short by many margins and better mathematicians such as Connes and Marcolli have tried to dig deeper in the mathematics behind renormalization. However, they seem to suggest that such a thing would only work for some noncommutative geometry, something I deeply disagree with and, moreover, might be in conflict with nature. We will now argue now that all these "dual" points of view are rather nonsensical from a physical point of view and that the propagator has to exist in a stronger sense than the dual one, that is the usual Lebesgue sense. It is here that gravitation by means of some positive energy condition becomes of primordial importance. Our fundamental formula for the two point function in a general curved space time has rather the same shortcomings than the standard Minkowski one; in chapter six, we therefore suggested to gravitationally deform it so that the resulting integrals become well defined in the standard Lebesgue sense. The particular proposal made at that point is however not entirely complete and we shall discuss here a better, albeit still incomplete, one in the remainder of this section. We want to keep the definition of $\phi(x, k^a, y)$ as a sum over geodesic(s) but we will provide every exponential $e^{ik_a w^a}$ with an exponential surpression factor which is *local* at x and y; these factors may be interpreted as a kind of "resistance" space time offers to the sending and receiving of geodesic signals. If w^a is causal, then this surpression factor *might* be defined by

$$\alpha(x, k^a, w^b) = R_{\alpha\beta}(x)k^\alpha k^\beta + R_{\alpha'\beta'}(y)k^{\alpha'}_{\star\, w^b}k^{\beta'}_{\star\, w^b} + \gamma(k_a w^a)^2$$

where $R_{\alpha\beta}$ is the Ricci tensor, $\star\, w^b : T^\star\mathcal{M}_x \to T^\star\mathcal{M}_y : k^a e^\mu_a(x) \to k^{a'}_{\star w^b} e^{\mu'}_{a'}(y)$ denotes parallel transport along the geodesic defined by $w^b e^\alpha_b(x)$. The latter induces an orthochronous Lorentz transformation and (un)primed indices do refer to y (x). Here, we require the weak energy condition that $R_{\alpha\beta}V^\alpha V^\beta > 0$ for all timelike vectors V^α. This certainly does the job for a timelike w^a, however for a null w^a this formula may be insufficient to get convergence. In case w^b is spacelike, then denote by $R(w^b)^\alpha_\beta$ the reflection around w^b: the latter is an idempotent isometry on the future pointing causal vectors. One could now define

$$\alpha(x, k^a, w^b) = R_{\alpha\beta}(x)k^\alpha k^\beta + R_{\alpha'\beta'}(y)k^{\alpha'}_{\star\, w^b}k^{\beta'}_{\star\, w^b} +$$

$$R_{\alpha\beta}(x)R(w^b)^\alpha_\kappa k^\kappa R(w^b)^\beta_\gamma k^\gamma + R_{\alpha'\beta'}(y)R(w^{b'}_{\star\, w^b})^{\alpha'}_{\kappa'}k^{\kappa'}_{\star\, w^b}R(w^{b'}_{\star\, w^b})^{\beta'}_{\gamma'}k^{\gamma'}_{\star\, w^b} + \gamma(k_a w^a)^2$$

and by using that $R(\lambda w^b)^\alpha_\beta$ is independent of λ for $\lambda \neq 0$ (a reflection is defined by an axis, not an orientation), we have that

$$\alpha(x, k^a, w^b) = \alpha(y, k^{a'}_{\star\, w^b}, -w^{b'}_{\star\, w^b})$$

and

$$\alpha(x, k^a, w^b) = \alpha(x, R(w^b)^a_b k^b, w^c).$$

The distinction between the spacelike and causal case is obvious since null w^a do not canonically define a reflection and the reflection around timelike vectors swaps the future and past lightcones. We define now

$$\phi_\mu(x, k^a, y) = \sum_{w^a \in T^\star\mathcal{M}_x : \exp_x(w) = y} e^{ik_a w^a} e^{-\mu\alpha(x, k^a, w^b)}$$

and as before

$$W_\mu(x,y) = \int_{T^\star \mathcal{M}_x} \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\theta(k^0)\phi_\mu(x,k^a,y).$$

From the above properties and similar reasoning as in the previous chapters we obtain that

$$\overline{W_\mu(x,y)} = W_\mu(y,x)$$

and

$$W_\mu(x,y) = W_\mu(y,x)$$

for $x \sim y$. It is kind of obvious that this propagator on a de-Sitter space time is *not* finite for $\mu, \lambda > 0$ given that the Ricci tensor is proportional to the metric and therefore all curvature terms are constant. More precisely, for timelike w^a we do have exponential surpression due to the $(k_a w^a)^2$ term, but the latter does not do a proper job in case w^a is spacelike. Thus, in a maximally symmetric space time, where the Riemann tensor is fully equivalent to the metric itself, there is no way to get a theory out satisfying our finiteness criteria unless one simply ignores spatial propagation which would endanger the spin-statistics theorem. One can easily save the day by relying on geometries which do locally define a *preferred* timelike unit vectorfield V^μ; as is well known, such geometries are *generic* and may even be algebraically special; Wylleman has recently given an explicit construction hereof. Hence, one could simply replace the $(k_a w^a)^2$ term by a $(k_\mu V^\mu)^2$ or $(k_\mu V^\mu)^2 + (R(w)^\mu_\nu k^\nu V_\mu)^2$ term, in case w is spacelike, which would provide one with the necessary falloff and symmetry properties independent of w^a. The physical message here is plain and simple, in the non-relativistic theory, one had that the two point function is well defined[4] and finite unlike in the Minkowski case; to restore these salient properties, we need a physical arrow of time which is realized by generic matter distributions. All maximally symmetric space times are pathological in the sense that no realistic matter propagates on them; now, people would argue that such timelike vectorfield is not observed in nature as it might suggest a violation of "Lorentz invariance" although everything is formally locally Lorentz covariant. Such attitude is of course rather nonsensical given that we have not specified yet how the two point functions relate to observable quantities and moreover, the surpression terms in the amplitude are *local* and therefore do not influence the propagation part of the definition which resulted in the Fourier basis functions. All our surpression terms do is to incoorporate a kind of "resistance" of the space time fabric to the creation and annihilation of a signal of a particular type just like a liquid offers resistance to the creation and annihilation by pointlike motion of waves but little if nothing to the propagation theirof. The gravitational field is such an eather and Minkowski's idealization is just fictituous; I have no idea wether it is sensible to say that these surpression terms have to be small in some sense as, again, they do not pertain to the propagation aspect of the signal but

[4]The propagator is certainly well defined when applying a momentum cutoff and sending the cutoff towards infinity; I did not check if it exists in the Lebesgue sense.

merely to the creation and annihilation thereof. It is still possible to work in a spatially homogeneous and isotropic cosmology, such as the one given by the usual Friedmann universes: therefore, in a later chapter, we will compute the quantum theory on such a big bang type of universe. It will turn out, however, that a little friction on the propagation of the signal is also required in order to tame the divergencies of the lightcone and we shall adress that issue in a while.

The reader might infer at this point that the local weight factors seem rather ad-hoc, an attitude which I can agree with to some extend. Let us first comment that every more general framework for physics always allows for more possibilities: we can know the principles of nature but not its representation! Einstein's theory enlarged our vision on the universe by many orders of magnetude and likewise does our principle of general Lorentz covariance regarding the possible quantum laws. The *local* weights attached to the creation and annihilation process are however of a different nature: one interpretation is that they are attached to an action occuring *outside* the framework of four dimensional space time. Constales suggested to me that one could regard space time as being made out of atoms and that $\sqrt{\mu}(V^a k_a)^2$ could be interpreted as a kind of self energy the wave has relative to the space time gas. Here, the length scale $\sqrt{\mu}$ could serve as an inverse temperature and one could therefore uphold a thermodynamic interpretation. This is certainly an interesting point of view but in my eyes no "weakening" of our continuum formalism: I accepted already for a long time that our theories come with motivated representations and that there are in general no real reasons to prefer one representation over another. This is the beauty of science, we are never able to tell to the fullest extend how things *are* but we can gain insights about what kind of ideas are necessary and moreover, we are able to refute certain wordimages. Religious types on the other hand claim to know the secret to come to God or they believe that they can find a God given deterministic theory behind quantum mechanics. Even worse, some of them are so delirious that they believe that the number 137 should be "explained" by some mystical mechanism. It is the duty of any decent university to dismiss a professor or researcher once he starts to behave in such a way, even when they think of that person as a kind of local hero.

7.1 First steps with modified propagators.

We will now start to investigate, by means of a couple of examples, the consequences of the local surpression terms added above. The reader immediately notices that we have a different prescription for causal geodesics than for spacelike geodesics so that ultimately one may expect discontinuities on the lightcone. Indeed, everywhere else, our regularization scheme leads to a C^∞ scalar propagator with uniformly bounded covariant derivatives, except at the lightcone where we will have to perform a supplementary regularization. This issue is *not* important for theories regarding interacting spin-0 particles or quantum electrodynamics, the theory of charged spin-$\frac{1}{2}$ particles interacting by means of a massless spin-1 particle since here, the derivatives of the propagators do not play any role. They become only important in non-abelian gauge theory or the theory of gravitons and I have therefore omitted an explicit regularization of the lightcone in those publications. I have promised however to fill this gap in a book publication and I shall grasp the occasion to do so. The reader should very well understand that the lightcone regularization falls within our principle of Lorentz covariance so that there is nothing strange about it; however, it is always interesting to see how it becomes necessary by means of more "elementary" computations. I have decided to present this chapter at a pedestrian level, showing step by step by examples what one needs in order to obtain a well defined theory: ultimately, the reader should understand the presence of the supplementary parameters as a mere possibility allowed by nature and there is *no* reason why supplementary constraints should be imposed upon the theory. On the contrary, such limitations often make the theory ill defined and this is what we want to avoid.

To start with, let us study our regularization scheme in Minkowski space time where ∂_t has to be associated to the timelike vectorfield V^μ defined by some physical observer making the quantum particle feel an eather due to him or herself, and see if our integral has all desired properties. As is evident from the previous discussion, the only problem with the two point function really resides near the null cone and for this purpose it is sufficient to take the massless limit $m \to 0$. With these reservations

$$W_\mu(x,x') = \frac{1}{2(2\pi)^3} \int \frac{d^3\vec{k}}{\left|\vec{k}\right|} e^{i\left(|\vec{k}|(t'-t)+\vec{k}.(\vec{x}'-\vec{x})\right)} e^{-2\mu\left|\vec{k}\right|^2}$$

for points x, x' which are causally related. We will not explicitely calculate the regularization for spacelike separated events and leave this as an exercise for the reader. One may further calculate the propagator to be

$$\begin{aligned}
W_\mu(x,x') &= \frac{1}{(2\pi)^3\,|\vec{x}'-\vec{x}|} \int_0^\infty dk \sin(k\,|\vec{x}'-\vec{x}|) e^{ik(t'-t)-2\mu k^2} \\
&= \frac{1}{2i(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|} e^{-\frac{(t'-t+|\vec{x}'-\vec{x}|)^2}{2\mu}} \int_0^\infty dk e^{-\left(k-i\frac{(t'-t+|\vec{x}'-\vec{x}|)}{\sqrt{2\mu}}\right)^2} - \\
&\quad \frac{1}{2i(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|} e^{-\frac{(t'-t-|\vec{x}'-\vec{x}|)^2}{2\mu}} \int_0^\infty dk e^{-\left(k-i\frac{(t'-t-|\vec{x}'-\vec{x}|)}{\sqrt{2\mu}}\right)^2}
\end{aligned}$$

and to study the limit $\mu \to 0$ is a rather subtle issue since, albeit the real part of both integrals equals $\frac{\sqrt{\pi}}{2}$ independent of the arguments $t' - t \pm \left|\vec{x'} - \vec{x}\right|$, the complex part is diverging and cannot be computed exactly. More precisely, we note that both integrals are of the form

$$I(c) = \int_0^\infty dk e^{-(k-ic)^2}$$

and the integrand is complex analytic in k and c. For real c, we may compute the integral by considering the limit of a contour in the complex plane from 0 to R to $R+ic$ to ic and finally back to 0. As usual, the integral over the large vertical part vanishes in the limit for R to infinity while the remainder gives

$$I(c) = \int_0^\infty dk e^{-k^2} + i \int_0^c dk e^{k^2}.$$

This shows that the imaginary part of $W_\mu(x,x')$ equals

$$\frac{\sqrt{\pi}}{4(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|}\left(e^{-\frac{(t'-t-|\vec{x}'-\vec{x}|)^2}{2\mu}} - e^{-\frac{(t'-t+|\vec{x}'-\vec{x}|)^2}{2\mu}}\right)$$

which converges in the limit for μ to zero to the usual delta functions on the lightcone. The real part however is given by

$$\frac{1}{2(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|}\left(e^{-\frac{(t'-t+|\vec{x}'-\vec{x}|)^2}{2\mu}}\int_0^{\frac{t'-t+|\vec{x}'-\vec{x}|}{\sqrt{2\mu}}} dk e^{k^2} - e^{-\frac{(t'-t-|\vec{x}'-\vec{x}|)^2}{2\mu}}\int_0^{\frac{t'-t-|\vec{x}'-\vec{x}|}{\sqrt{2\mu}}} dk e^{k^2}\right)$$

and the task remains to get insight into the large c behavior of

$$\int_0^c dk e^{k^2}.$$

A crude estimate

$$\frac{\sqrt{\pi}}{2}e^{\frac{c^2}{2}} \leq \int_0^c e^{k^2}\,dk \leq \frac{\sqrt{\pi}}{2}e^{c^2}$$

may be shown immediately by means of

$$\left(\int_0^c e^{k^2}dk\right)^2 \leq \frac{\pi}{2}\int_0^{\sqrt{2}c} dr\, r e^{r^2} = \frac{\pi}{4}(e^{2c^2}-1)$$

and likewise for the lower bound. However, this is not good enough and for $c>0$ one can, by means of analytic methods, obtain that

$$\int_0^c e^{k^2}dk = \frac{1}{g(c)c}\left(e^{c^2}-1\right)$$

where $1 \leq g(c) \leq 2$ and $g(0)=1$ and $g(+\infty)=2$ which is precisely what we need. Hence, the real part of the two point function behaves as

$$\begin{aligned}\mathcal{R}e\, W_\mu(x,x') &= \frac{1}{2(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|}\frac{1}{c_+(x,x',\mu)g(c_+(x,x',\mu))}\left(1-e^{-c_+(x,x',\mu)^2}\right) \\ &\quad - \frac{1}{2(2\pi)^3\sqrt{2\mu}\,|\vec{x}'-\vec{x}|}\frac{1}{c_-(x,x',\mu)g(c_-(x,x',\mu))}\left(1-e^{-c_-(x,x',\mu)^2}\right)\end{aligned}$$

and

$$c_\pm(x,x',\mu)) = \frac{t'-t\pm|\vec{x}'-\vec{x}|}{\sqrt{2\mu}}.$$

It is easy to see that for x' in the lightcone of x, one has that the limit of μ to zero of $\mathcal{R}e\, W_\mu(x,x')$ vanishes and the same holds when x' is null related. The convergence of the right hand side towards space time infinity for x' causally related to x is only slow since, along a branch of $t' - t - |\vec{x}' - \vec{x}| = c$, it goes proportional to

$$\frac{1}{|\vec{x}' - \vec{x}|}$$

which is not quadratically integrable in $\vec{x}'$. Similar results hold when x' is spacelike related to x albeit the computation is somewhat more difficult there due to the reflection symmetry. It is obviously so that in Minkowski space time, it will never be possible to get the integral

$$\int |\Delta_{F,\mu}(x,y)|^2 \, dxdy = \int |W_\mu(x,y)|^2 \, dxdy$$

finite due to the translation symmetry. However, this is not something we should be ambitious of as such integrals have nothing to do with real physics. We shall examine now wether this weak asymptotic behavior is sufficient to get finite loop diagrams by studying some cases which usually give infinite results. Before we proceed, let us notice that, under the agreement that the coincidence limit is defined by the causal presciption, we have

$$W_\mu(x,x) = \frac{1}{4\pi^2} \int_0^\infty dk \, k e^{-2\mu k^2} = \frac{1}{8\pi^2 \mu}$$

which is a finite number usually much larger than one since μ is taken to be small. Therefore, the simpelest one vertex correction to the propagator from x to y reads

$$O_\mu(x,z) = \frac{-i\zeta}{8\pi^2\mu} \int_{\mathcal{M}} \Delta_{F,\mu}(x,y) \Delta_{F,\mu}(y,z) dy$$

where $\zeta > 0$ is the coupling constant of the theory. We will now isolate a, fairly special, subintegral which diverges to infinity: consider the geometrical situation where z is in the future of x and y in the future of z. These three points determine a plane and consider now the set of space time point y' such that

$$c_-(x,y',\mu) = c_-(x,y,\mu)$$

and

$$c_-(z,y',\mu) = c_-(z,y,\mu).$$

The set of y' constitutes a two dimensional manifold as it is the intersection of two three dimensional manifolds and the product $\Delta_{F,\mu}(x,y')\Delta_{F,\mu}(y',z)$ behaves as

$$\sim \quad \frac{1}{8(2\pi)^6 \mu} \frac{1}{|\vec{y}' - \vec{x}| \, |\vec{y}' - \vec{z}|} \left(-\frac{1}{c_-(x,y,\mu) g(c_-(x,y,\mu))} \left(1 - e^{-c_-(x,y,\mu)^2}\right) + i\frac{\sqrt{\pi}}{2} e^{-c_-(x,y,\mu)^2} \right)$$
$$\left(-\frac{1}{c_-(z,y,\mu) g(c_-(z,y,\mu))} \left(1 - e^{-c_-(z,y,\mu)^2}\right) + i\frac{\sqrt{\pi}}{2} e^{-c_-(z,y,\mu)^2} \right)$$

for sufficiently large $|\vec{y}'|$. It is clear that the integration of this expression over the one dimensional manifold defined as the intersection of the previous two dimensional manifold with the plane formed by x, y, z diverges linearly (since we have to take into account a $|\vec{y}'|^2$ coming from the measure). It must be clear to the reader that this pathological behavior of well chosen subintegrals is going to cause general trouble, which necessitates friction on the propagation of the signal to obtain improved convergence properties. However, this friction should be momentum independent as is dictated by Lorentz invariance. Obviously, we might just have excluded loop diagrams of this type since no propagation from x to x should ever happen but one might envision problems with other diagrams containing two interaction vertices with one loop and four externel legs. Naively, a logarithmic divergency might occur there.

Before we proceed, let us think about potential trouble regarding the general definition as well as generic features of the regularization scheme *outside* the lightcone. The reader must have wondered what should happen to our definition when there exists a continuum of geodesics joining x to y such as is the case in a closed Friedmann universe where the spatial metric is the one of a three sphere embedded in a flat four dimensional Euclidean space. In that case, one obtains, just like on a two sphere endowed with the standard Riemannian metric, that a continuum of geodesics joins a point with its antipodal point and only if a vector w in the tangent space of x has the property that $\exp_x(w)$ equals the antipodal point, do we have that $\exp_x$ does not define a local diffeomorphism in a neighborhood of $w \in T^\star\mathcal{M}_x$. In all other cases $\exp_x$ does behave like a local diffeomorphism on $T^\star\mathcal{M}_x$ and, therefore, the occurance of such anomaly is really of measure zero where the measure we speak about could be the one on tangent space $T^\star\mathcal{M}_x$ as well as on space time itself. Hence, the prescription is to simply ignore these points and to declare the propagator not defined on them as a function. This is not too bad given that one would expect, physically, such thing to happen: there is just too much information flowing to the antipodal point and our prescription, which is again of distributional nature on the sphere, does not know how to deal with it. This example brings us to an important property, namely that it is desirable to consider only those tangent vectors w in $T^\star\mathcal{M}_x$ such that $\exp_x$ is a local diffeomorphism in a neighborhood of w. This is important when we want derivatives of the propagator to be defined and we will come back to this later on. Finally, let us study what happens regarding the regularization scheme for spacelike geodesics defined by $w \in T^\star\mathcal{M}_x$ when w approaches the null cone in x so that the corresponding sequence of points $\exp_x(w)$ converges to a point on the null cone. We will show that the w contribution to the propagator vanishes in this limit which means in Minkowski that the entire propagator vanishes. Therefore, we obtain a jump on the lightcone given our previous computations. To get an idea why this is true, consider four dimensional Minkowski space time, x as the origin, and a sequence of spacelike vectors $w_\alpha = (\tanh(\alpha), 1, 0, 0)$, where we are interested in the limit for α to

$+\infty$. The reflection around w is defined by

$$R(w) = 1 + 2\frac{1}{1-\tanh^2(\alpha)}(\tanh(\alpha), 1, 0, 0)^T(\tanh(\alpha), -1, 0, 0)$$

which in matrix form reads

$$R(w) = \frac{1}{1-\tanh^2(\alpha)}\begin{pmatrix} 1+\tanh^2(\alpha) & -2\tanh(\alpha) & 0 & 0 \\ 2\tanh(\alpha) & -(\tanh^2(\alpha)+1) & 0 & 0 \\ 0 & 0 & (1-\tanh^2(\alpha)) & 0 \\ 0 & 0 & 0 & (1-\tanh^2(\alpha)) \end{pmatrix}$$

as the reader may verify. Therefore,

$$(V^a k_a)^2 + (V^a(R(w)k)_a)^2 = (k^0)^2 + \left(\frac{1}{1-\tanh^2(\alpha)}((1+\tanh^2(\alpha))k^0 - 2\tanh(\alpha)k^1)\right)^2$$

and this expression diverges to $+\infty$ for any k^a in the limit for $\alpha \to +\infty$ in case $m > 0$. Only in case $m = 0$ and $k^0 = k^1$ do we obtain a finite answer $(k^0)^2$ but the set of such wavevectors has measure zero so that we may conclude that $e^{-2\mu((V^a k_a)^2 + (V^a(R(w)k)_a)^2)} \to 0$ in the limit for $\alpha \to \infty$ almost everywhere for $m \geq 0$. This shows that the regularized propagator for spacelike separated points has vanishing limit towards the lightcone in Minkowski space time. I leave it up to the reader to show that this is generically the case, at least for every w mode. Hence, we have shown the need for an extra regularization scheme near the lightcone.

We shall now argue what kind of "extensions" one can make regarding the Schrodinger equations we have written down to determine the generalized Fourier waves while keeping in mind the nature of the surpression terms we have to build in. In principle, one can write down an infinite number of terms commensurable with local Lorentz covariance: this is not really a surprise given that general covariance allows for a similar extension of Newton's gravitational theory. We can constrain, however, the extend to which this principle should be applied by restricting to data which is locally determined by the second derivatives of the metric tensor in the same way as Einstein's theory follows uniquely from action principles in local densities containing at most two derivatives of the metric tensor but there is no such a-priori need to do so. So, we shall mainly discuss a few examples of "deformations" which I deem interesting but the reader may invent plenty more of them. We start by giving the example of an "energy" term which could be added to the Schrodinger equation and which respects Lorentz covariance on the propagation part. As mentioned already, we assume that our geometry provides for a unit timelike vectorfield V^μ causing friction in the creation and annihilation of particles at definite space time points: as is well known, a unit timelike vectorfield determines a unique Riemannian metric tensor $h_{\mu\nu}(x)$ as

$$h_{\mu\nu} = 2V_\mu V_\nu - g_{\mu\nu}$$

given our signature convention $(+---)$. The reader should keep in mind that all indices are raised and lowered with the Lorentzian metric and associated vierbein; so $h_{ab} = e^{\mu}_{a} e^{\nu}_{b} h_{\mu\nu}$ with the standard vielbein e^{μ}_{a}. With these lessons in mind, we can now write down another covariant energy term given by

$$\sqrt{h_{ab}(x_{w^c}(s))w^a(s)w^b(s)}$$

where $w^{\mu}(s) = \frac{dx^{\mu}(s)}{ds}$. So, our differential equation becomes

$$\frac{d}{ds}\tilde{\phi}_{\kappa}(x, k^a, w^b, s) = \left(iw^{\mu}(s)k_{\mu}(s) - \kappa\sqrt{h_{\mu\nu}(x_{w^b}(s))w^{\mu}(s)w^{\nu}(s)} \right) \tilde{\phi}_{\kappa}(x, k^a, w^b, s)$$

giving rise to the solution

$$\tilde{\phi}_{\kappa}(x, k^a, w^b) = e^{ik^a w_a} e^{-\kappa \int_0^1 \sqrt{h_{\mu\nu}(x_{w^b}(s))w^{\mu}(s)w^{\nu}(s)}} ds.$$

In our case of Minkowski space time, and some vielbein with $e_0 = \partial_t$, $h_{ab} = \delta_{ab}$ and

$$\phi_{\kappa}(x, k^a, y) = e^{ik^a(y_a - x_a)} e^{-\kappa|y-x|}.$$

For sake of convergence, it is assumed that the real part of κ is greater than zero. It turns out that this surpression mechanism is interesting as integrals of the kind

$$\int \Delta_{F,\mu,\kappa}(x, y)\Delta_{F,\mu,\kappa}(y, z)$$

are in the same "function class" as $\Delta_{F,\mu,\kappa}$ meaning they have similar falloff properties towards infinity so that the proof of perturbative renormalizability of the theory becomes self evident as we shall see later on[5]. Roughly speaking, all cases are covered if integrals of the kind

$$\int dy e^{-\kappa|x-y|-\rho|y-z|}$$

where $\kappa, \rho > 0$ belong to the same function class as $e^{-\zeta|x-z|}$ for some other $\zeta > 0$. From a simple triangle inequality estimate, one obtains that

$$\frac{1}{2}|x - z| + \left| y - \frac{x + z}{2} \right| \leq |x - y| + |z - y|$$

for $\left|y - \frac{x+z}{2}\right| \geq |x - z|$. This splits the integral into two parts as follows

$$e^{-\frac{1}{2}\min\{\kappa,\rho\}|x-z|} \int_{\left|y-\frac{x+z}{2}\right| \geq |x-z|} e^{-\min\{\kappa,\rho\}\left|y-\frac{x+z}{2}\right|} dy + e^{-\min\{\kappa,\rho\}|x-z|} \int_{\left|y-\frac{x+z}{2}\right| \leq |x-z|} dy$$

and this may further be bounded by

$$2\pi^2 \left(\frac{6}{(\min\{\kappa, \rho\})^4} e^{-\frac{1}{2}\min\{\kappa,\rho\}|x-z|} + \frac{1}{4} e^{-\min\{\kappa,\rho\}|x-z|} |x - z|^4 \right)$$

[5]Although the proof of convergence of the series is more involved as we will figure out later.

where $2\pi^2$ equals the volume of the three dimensional unit sphere with radius one. These functions obiously belong to the same class as $x^n e^{-\kappa x} \leq a e^{-\zeta x}$ for some $a > 0$ and $0 < \zeta < \kappa$ for all n. The same technique can be applied to an arbitrary number of points $x, z, \ldots$ in the integral as the reader may easily verify for himself. The bound above is slightly inconvenient because of the division of $\min\{\kappa, \rho\}$ by a factor of two in the exponential; this can however be repaired by noticing that

$$|x - z| + \left| y - \frac{x + z}{2} \right| \leq |x - y| + |z - y|$$

for $\left| y - \frac{x+z}{2} \right| \geq \frac{3}{2} |x - z|$ which would only change the coefficients in the polynomial. This shows that our idea is an interesting one regarding theories which do not need derivatives of the propagators given that the behavior on the light cone is still anomalous.

We will restrict the situation to be considered to geometries such that for every $x, y \in \mathcal{M}$ there exist open environments $\mathcal{V}_x, \mathcal{W}_y$ of x and y respectively, as well as disjoint opens $\mathcal{O}_{x,w} \subset T\mathcal{M}$, containing (x, w) for each $w : \exp_x(w) = y$, projecting down to $\mathcal{V}_x$ such that

$$\phi(x', k^{a'}, y') = \sum_{w' : exp_{x'}(w') = y'} \tilde{\phi}(x', k^{a'}, w')$$

for all $x' \in \mathcal{V}_x$ and $y' \in \mathcal{W}_y$ where every w' lies in exactly one $\mathcal{O}_{x,w}$ and vice versa. In other words, we assume $\exp_x$ to be a local diffeomorphism around each w such that $\exp_x(w) = y$ and moreover, the different $w's$ are "regularly" separated so that we can find open neigborhoods around them diffeomorphically mapping to $\mathcal{W}_y$ without intersecting. It would be worthwhile to investigate this condition in closer detail but I suspect it to be true generically; from this, one can locally construct Synge functions $\sigma(x', y'; w)$ which are defined for each w on $\mathcal{V}_x \times \mathcal{W}_y$ so that $\tilde{\phi}(x', k^{a'}, w')$ contains the exponential $e^{-ik^{a'}\sigma_{a'}(x', y'; w)}$ where $w' \in \mathcal{O}_{x,w}$. Now, we modify for example the latter by a prefactor of

$$e^{-\frac{1}{L^2 \sigma^2(x', y'; w)}}$$

which the reader immediately reconizes as the insertion of a factor

$$-\frac{2}{L^2 \sigma^3(x', \gamma(s); w)} \sigma_{,\alpha'}(x', \gamma(s); w) \gamma^{\alpha'}(s)$$

in the Schrodinger equation for the potential. Now, it is a well known property of the function $e^{-\frac{1}{z^2}}$ that the limit for $z \to 0$ of $z^{-n} e^{-\frac{1}{z^2}}$ vanishes for any $n \in \mathbb{N}$. Therefore, under reasonable *uniform* boundedness properties with respect to $h_{\mu\nu}$, of the covariant differentials, given by $g_{\mu\nu}$, of $\sigma(x, y; w)$ regarding w, the reader should be able to verify that not only

$$\lim_{w' \to \mathcal{N}_{x'}} \tilde{\phi}_{\mu,\kappa,L;\alpha\ldots\beta'\ldots}(x', k^{a'}, w') = 0$$

where $\mathcal{N}_{x'}$ denotes the lightcone at x' in $T^\star\mathcal{M}_{x'}$, but also that

$$\phi_{\mu,\kappa,L;\alpha\ldots\beta'\ldots}(x', k^{a'}, y')$$

is well defined as a function. Again, I suspect this to be true for generic space times and we shall make use of these results when developing non-abelian gauge theory and as well as the graviton theory. L is a large cosmological *mass*-scale associated to macroscopic physics so that $\frac{1}{L^2\sigma^2}$ quickly comes close to zero for relatively small σ; this is a pretty interesting remark as it shows how macroscopic scales intertwine with microscopic physics. Conventionally, one might choose $L^2 = \frac{1}{\mu}$ but we don't have to.

We finish this discussion by providing for the correct definition of the Feynman propagator *directly*

$$\Delta_{F,\mu,\kappa,L}(x,y) = \sum_{w:\exp_x(w)=y \text{ and w is in the future lightcone of x}} W_{\mu,\kappa,L}(x,w) + \sum_{w':\exp_y(w')=x \text{ and w' is in the future lightcone of y}} W_{\mu,\kappa,L}(y,w') + \sum_{w:\exp_x(w)=y \text{ and w is spacelike at x}} W_{\mu,\kappa,L}(x,w).$$

The reader must understand how this definition differs from the previous one and that $\Delta_{F,\mu,\kappa,L}(x,y)$ is everywhere differentiable.

7.2 Physical remarks regarding the construction.

The reader not familiar with Feynman diagrams will understand that the kind of integrals considered in this chapter are mandatory for any spin-0 theory to be well defined. We have not given any attention so far to the regularization of the spin-$\frac{1}{2}$ or spin-1 propagator since that would merely have obfuscated the presentation and would not have brought any essential point on the table. The reasons for doing so are, however, somewhat different: I am not aware of any known physical theory which contains the derivatives of the Fermi-propagator so that the fine details of the spin-0 regularization scheme near the lightcone seem somewhat unnecessary to implement albeit everything proceeds in a straightforward way. The regularization scheme for any integer spin propagator, on the other hand, is identical to the spin-0 case so that there is nothing lost in presenting that case only. We shall come back to estimates for regularized propagators of higher spin in a later chapter since, at this point, the global constraints imposed on such regularization would seem to be rather ad-hoc and not so important. It is however in chapter nine, while studying an alternative vacuum cosmology of the hyperbolic type that the relevant constraints will become physically clear and intuitive. For this reason, I have decided to postppone the issue of regularization for higher spin propagators to chapter ten, where they will serve as the basis for a very general proof of finiteness of Feynman diagrams.

So, we are not completely done yet and the cosmological $SO(3)$-class of common vierbeins for $g_{\mu\nu}$ and $h_{\mu\nu}$ shall have an important role to play in that construction as the reader can guess immediately. Upper bounds on spin components have to happen in such cosmological class of reference frames as arbitrary local Lorentz transformations are in the position to violate any inequality. Therefore, we shall just finish this chapter by making some further physical comments; our regularization scheme near the lightcone imposed that the latter is a forbidden place for a particle to be found, even a massless one. Whereas in the traditional theory, the lightcone comes with a delta singularity, the latter has been smeared in a certain band in the spacelike and timelike region near the lightcone where the width is measured in the cosmological class of reference frames determined by a preferred timelike vectorfield. This is a sensible thing to do as the lightcone is a kind of "unbreakable" wall in the classical theory which has now been softened in the quantum theory. Instead of taking the negative attitude that the regularization scheme has many liberties and therefore, the canonical character of our theory is destroyed, one should cherish the very fact that our computations show that such regularization is necessary. Moreover, it falls within the class of Lorentz covariant theories and therefore, this is the very best we can do, no further determination can reasonably be expected. Only religious bigots with no understanding of physics whatsoever could keep on complaining about this very point but to them I say: go and study some elementary relativity my "friend". Therefore, let me stress that the only aspect of our construction which appears to call for a "deeper" picture regards the insertion of the weight factors associated to the creation and annihilation process of particles. Again, our construction showed that such idea is *necessary* but it might find a "prettier" origin in a different representation of the same physics. Here, we must make the deep remark that the property of particle statistics hinges upon the reflection symmetry of the weight factors associated to the creation and annihilation processes; therefore, we *need* a new principle of nature from which statistics follows, which is the one that such processes are indeed reflection symmetric. This was to be expected as the entire, standard, argumentation behind particle statistics hinges upon properties of flat space time and there is no a-priori reason for standard or any kind of statistics to hold in curved space time. In this book, we shall not study theories of that kind as they would require very novel and deep ideas regarding its very formulation.

Chapter 8

Interactions for (non-abelian) gauge theory and gravitons.

This chapter will show why the integrals in the previous chapter were important: we will systematically explain all the necessary ingredients prior to defining the interacting theory. We shall work in the utmost generality and explain the genesis of gauge invariance, a principle necessitated by the operational Minkowski theory from a different point of view. Therefore, gauge invariance comes in a different guise and indeed, our derivation is very different but gives completely isomorphic results in the aforementioned unphysical limit. The intention of this chapter is to give a *formal* definition meaning that all questions regarding the "well-definedness" of the theory are postponed until chapters ten and eleven, where we shall provide for a rather general answer. Indeed, our theory is given by a so-called perturbation series and there need to be shown two things: (a) finiteness of and appropriate bounds on the constituents of the series (b) convergence of the series in some well chosen domain of the interaction parameters. In the literature on standard quantum field theory, regarding point (a), one has a control over finiteness (after an illegitimate infinite substraction) for the so-called renormalizable theories but no bound whatsoever so that adressing (b) is far out of reach. We shall progress systematically in this chapter by including particles of higher spin one at the time; also, we shall define simplified scattering amplitudes first, which resemble expressions found in the literature. Only at a later stage do we define the real physical amplitudes and weights associated to more complex processes. All proofs in chapters ten and eleven refer to the simplified situation; the same results regarding the more complex physical amplitudes are however quickly obtained by means of the same methods.

8.1 Interactions for spin-0 particles.

In order to describe realistic theories such as QED and QCD, we should include spin degrees of freedom by means of the gamma-matrices; however, we will content ourselves for now with the description of relativistic ϕ^4 theory. How should an interaction theory be constructed? We have so far defined the two point function and the associated Feynman propagator for a free particle born or created at a space time event x; in the definition of the two point function, geodesic paths were allowed to travel into the relativistic past whereas this is explicitely forbidden in the definition of the Feynman propagator. Indeed, the amplitude associated to $\Delta_F(x,y)$ is one calculated with repect to a process going forwards in time even if $y \in J^-(x)$. This does not imply that there do not exist processes going backwards on space time in some sense but that associated to those is the amplitude of another processes going forwards on space time. This is our condition of internal temporality discussed in the introduction to this book. Often, researchers in physics interpret the Feynman propagator as stating that no process can go to the relativistic past on space time, however, such strong interpretation is not mandatory at all and we shall posit a way of looking at things which is more flexible. So, the Feynman propagator is the correct physical object at hand and corresponds to information travelling on geodesics regarding the birth or creation of a particle at x and being annihilated at y. Only if there exists a future oriented geodesic directed from y towards x, does the amplitude of the process of travelling from x towards y backwards in time, agree with the standard amplitude defined by the process of travelling from y towards x forwards in time. This is an interpretation which is *necessary* to be consistent with the birth or creation of the particle at x, on one hand, and the condition of internal temporality on the other. So what is interaction? It is nothing but the process of scattering of information currents at an intermediate space time point z called an internal vertex of the Feynman diagram. Depending upon the type of interaction vertices do we have different theories and we shall see that local symmetry properties impose severe constraints on the possible types of interaction vertices. Every vertex comes with a coupling constant λ which in non-gravitational theories has been assumed to be dimensionless; given that the mass dimension of every interaction vertex has to be four, we conclude that the only possible such theory is one with interaction vertices having four legs corresponding to one endpoint of the Feynman propagator. We are almost there now, it turns out to be that in standard quantum mechanics, λ has to be multiplied by $-i$ which consitutues one aspect of "unitarity"; furthermore, it is logical that the contribution of each diagram has to be divided by its symmetry factor and that all diagrams have to be summed over. More abstract, a "Feynman diagram" is a multi-graph with interaction four vertices (in our case) having an undetermined position in space time. Each diagram D has a symmetry factor $s(D)$ given by the number or symmetries of the oriented multigraph keeping the end-points fixed[1] and we demand each interaction vertex to be connected to an

[1]A symmetry of an oriented multigraph is a map α from the oriented multigraph to itself determined by its action on the vertices and edges such that all coincidence relations are

IN or OUT boundary vertex. Hence, we want to calculate an amplitude such as $\langle \mathrm{OUT}\, y_1, \ldots, y_m | \mathrm{IN}\, x_1, \ldots, x_n \rangle$ regarding the creation process of n particles at x_i and the subsequent detection (not necessarily annihilation as we know from our discussion regarding Fermions) of m particles at y_j where the y_j come later than the x_i in the process sense. Here, the order of the space time labels x_i, y_j is irrelevant given that our particles obey bosonic statistics; with those reservations

$$\langle \mathrm{OUT}\, y_1, \ldots, y_m | \mathrm{IN}\, x_1, \ldots, x_n \rangle = \sum_D \frac{(-i\lambda)^V}{s(D)} \left(\prod_{j=1}^{V} \int_{\mathcal{M}} dz_j \sqrt{g(z_j)} \right) \prod_{\text{edges } (\alpha_i, \alpha_j)} \Delta_F(\alpha_i, \alpha_j)$$

where $\alpha_k \in \{z_l, x_i, y_j\}$ and V stands for the number of internal vertices of the diagram. Moreover, the IN vertices are never directly connected to themselves and the same holds for the OUT vertices; this definition also holds in case IN or OUT are empty. In case IN and OUT are empty, the amplitude equals one; this constitutes the *definition* of the theory and we notice that the only unexplained factor so far concerns the domain of integration $\mathcal{M}$. We shall give a physical motivation for this definition in the comments section of this chapter: there is very little, if almost nothing, ad-hoc about it as the reader will understand.

We will now specify three distinct choices of $\mathcal{M}$ one can, in principle, make; the reader familiar with quantum field theory will recognize that the issue we are discussing here is related to an "instantaneous" notion of vacuum state as well as the issue of global Poincaré covariance in the interacting theory. What follows has been discussed already, in one form or another, in chapter two but we shall summarize some relating thoughts at this moment. There, we discussed the notion of growth of a four dimensional universe as well as the notion of an actual NOW. This NOW has nothing to do with some Newtonian character of the interactions but reveals the healthy point of view that all interactions from IN to OUT cannot travel to the *realized* past of IN and nor to the potential future of OUT. Therefore, we have to complement the setup exlained so far with an initial S_I and final S_F spatial hypersurface associated to the IN state and OUT state, meaning that they contain x_i and y_j respectively and are disjoint. Associated to two hypersurfaces, one can define the sandwished region $R(S_I, S_F)$ as the set of events x such that every curve emanating from x either remains within $R(S_I, S_F)$ or leaves it by crossing $S_I \cup S_F$; hereby, it is assumed that any inextendible past oriented causal curve leaves $R(S_I, S_f)$ at S_I and any inextendible future oriented causal curve leaves $R(S_I, S_F)$ at S_F. Note that this definition is framed as such that closed timelike curves are allowed for given that we did not demand the hypersurfaces to be achronal; moreover, S_I, S_F are chosen such that $R(S_I, S_F)$ is nonempty. When the domain of integration of the interaction vertices is given by $R(S_I, S_F)$, we say that our quantum theory is of TYPE I. In a *classical* theory of the universe, one can speak about

preserved. The latter means that if E is an edge between v and w then $\alpha(E)$ is an edge between $\alpha(v)$ and $\alpha(w)$ but possibly the orientation between v and w has been reversed.

the realized past as a classical space time to the past of S_I; this is *not* so in a quantum theory where the past consists out of measurements made and those do not constitute a classical space time at all since space time is rather unknown when no measurement occurs. In that regard, for classical space time theories, we defined a quantum theory to be of type II when all events past to S_F have to be taken into account in the computation of the transition amplitude $\langle \mathrm{OUT}\, y_j, j = 1 \ldots m | \mathrm{IN}\, x_i, i = 1 \ldots n \rangle$. In a sense, this would mean that the recorded space time history plays a role in the behavior of elementary particles when evolving to the future: this is not a silly idea but one remniscent of Einstein causality. Type I is the most logical one in the sense that elementary particles do not care about the future nor about the past and all computations have to occur within $R(S_I, S_F)$. Type III is the opposite of Type II meaning that the potential (deterministic) future of S_I beyond S_F plays a role in the determination of the relevant amplitudes; the computations in quantum field theory are of Type II and III in the sense that the entire space time is taken into account. My personal guess is that nature works according to a Type I principle but, in the general analysis of subsequent chapters, we will leave the matter entirely open. The type alluded to in standard quantum field theory is mixed, the integrals go between $-\infty$ and $+\infty$ as they should for an S-matrix; this concludes the definition for an interacting spin-0 theory with dimensionless coupling constants. We now turn our heads towards the right setup for a spin-1 theory, again with no dimensionful coupling constants, a well as a spin-2 graviton theory.

8.2 Interactions for general (non-abelian) gauge theories.

In this section, we describe some part of the known relevant physics regarding interactions between spin-$\frac{1}{2}$ particles by means of massless spin-1 bosons. In doing so, we assume that the theory has some global symmetry group giving rise to charges for the fermionic as well as the bosonic particles in case the group is non-abelian. Standard non-abelian gauge theory is constructed in a way where the transformation laws of the gauge potential, or particle polarization, $A^\alpha_\mu(x)$ are *induced* from the transformation laws of the multiplets on representation space. This means, in particular, that all interactions are constructed from the basic object

$$\mathbf{A}_\mu = A^\alpha_\mu (t_\alpha)^m_n$$

by means of Lie-algebra operations as well as the trace operation between two Lie-algebra elements, where the t_α constitute the generators of the Lie-algebra

$$[t_\alpha, t_\beta] = i f^{\gamma}_{\ \alpha\beta} t_\gamma$$

and $\mathrm{Tr}(t_\alpha t_\beta) = g_{\alpha\beta}$. Here,

$$f_{\gamma\alpha\beta} = g_{\gamma\delta} f^{\delta}_{\ \alpha\beta}$$

is totally anti-symmetric in its three covariant indices and $g_{\alpha\beta}$ is positive definite. The former condition follows from the latter as one may show and the latter is required for a finite dimensional positive probability interpretation. Moreover, we do not take into account interactions requiring a length scale which implies all our interaction vertices are of mass dimension four. Moreover, by the very definition of interaction, the respective vertices need to be tri- or four-valent since gauge fields contribute a mass dimension of 1, while spinorial particles a mass dimension of $\frac{3}{2}$. All these considerations leave us with the following intertwiners

$$\begin{aligned} f_{\alpha\beta\gamma}\left(\nabla_{\kappa}A^{\alpha}_{\mu}\right)A^{\beta}_{\nu}A^{\gamma}_{\lambda}g^{\kappa\nu}g^{\nu\lambda} &= -i\mathrm{Tr}\left(\nabla_{\kappa}\mathbf{A}_{\mu}\left[\mathbf{A}_{\nu},\mathbf{A}_{\lambda}\right]\right)g^{\kappa\nu}g^{\mu\lambda} \\ f_{\alpha\beta\gamma}f^{\alpha}_{\ \beta'\gamma'}A^{\beta}_{\mu}A^{\gamma}_{\nu}A^{\beta'}_{\mu'}A^{\gamma'}_{\nu'}g^{\mu\nu}g^{\mu'\nu'} &= -\mathrm{Tr}\ \left(\left[\mathbf{A}_{\mu},\mathbf{A}_{\nu}\right]\left[\mathbf{A}_{\mu'},\mathbf{A}_{\nu'}\right]\right)g^{\mu\mu'}g^{\nu\nu'} \end{aligned}$$

concerning the self interaction of the gauge particles[2]. There remain the following two vertices

$$(\mathbf{A}_{\nu})^{m}_{n}(\gamma^{a})^{i}_{j}e^{\nu}_{a}(x)\Psi_{im}\overline{\Psi}^{jn},\ \ f_{\alpha\beta\gamma}v^{\beta}\overline{v}^{\gamma}\nabla^{\mu}A^{\alpha}_{\mu}$$

where the last vertex is constructed from

$$\mathbf{v} = v^{\alpha}t_{\alpha}$$

as

$$-i\mathrm{Tr}\left(\left[\mathbf{v},\overline{\mathbf{v}}\right]\nabla^{\mu}\mathbf{A}_{\mu}\right).$$

Therefore, just out of completeness, we should supplement our theory with a spin zero particle and anti-particle transforming in the adjoint representation of the symmetry group with Fermionic statistics due to the anti-symmetry of the commutator. In chapter six, we argued that the relevant two point functions for such particle had to be given by

$$W^{\alpha\beta}_{a}(x,y) = \overline{\theta}(x)\theta(y)W(x,y)g^{\alpha\beta},\ W^{\alpha\beta}_{p}(x,y) = \theta(x)\overline{\theta}(y)W(x,y)g^{\alpha\beta}$$

and in calculating Feynman diagrams, integration over the Grassmann coordinates should occur. There is however a deeper reason to introduce these ghosts than mere completeness which is that precisely as many "negative" local degrees of freedom are needed to kill the spin zero modes in the propagator

$$W^{\alpha\beta}_{\mu\nu'}(x,y).$$

The associated multiplication terms $\nabla^{\mu}A^{\alpha}_{\mu}$ are then seen as a "gauge condition" eliminating those degrees of freedom (associated to the longitudonal polarization of the massless particle).

Hence, we are left with precisely the same four interaction vertices as in standard non-abelian gauge theory. Moreover, by rescaling the Lie algebra generators $t_{\alpha} \to \lambda t_{\alpha}$, suitably defining the interaction constant $\tilde{g}$ of the theory and by

[2]The only other two remaining options $\mathrm{Tr}\left(\nabla_{\kappa}\mathbf{A}_{\mu}\left[\mathbf{A}_{\nu},\mathbf{A}_{\lambda}\right]\right)g^{\kappa\mu}g^{\nu\lambda}$ and $\mathrm{Tr}\ \left(\left[\mathbf{B}_{\mu\nu\lambda},\mathbf{A}_{\kappa}\right]\right)Z^{\mu\nu\lambda\kappa}$ vanish by means of symmetry. Types such as $\mathrm{Tr}\ \left(\left[\left[\mathbf{A}_{\mu},\mathbf{A}_{\nu}\right],\mathbf{A}_{\lambda}\right]\mathbf{A}_{\kappa}\right)Z^{\mu\nu\lambda\kappa}$ can be expressed in terms of the previous cases.

redefining the Grasmann numbers $\theta \to \lambda' \theta$ we obtain that they are of standard textbook form; that is, derived from a gauge covariant closed two form field strength.

We have just finished the discussion of the structure of the interaction vertices; now, we turn our head towards the definition of the interacting theory akin to what we have accomplished in the previous chapter. As before, we define the interacting theory as a sum over connected Feynman diagrams between in IN and OUT states $|\mathrm{IN}\,(x_1, a_1), \ldots, (x_n, a_n)\rangle$ and $|\mathrm{OUT}\,(y_1, b_1), \ldots, (y_m, b_m)\rangle$ respectively where a_i, b_j is associated to A^{α}_{μ} where α is a group index in the adjoint representation or linked to v_{im} corresponding to a particle in the IN state and an anti-particle in the OUT state, or a contravariant spinor and group index in the defining representation, associated to $\overline{v}^{im}$, with the opposite interpretational conventions. Here, it is understood that all x_i (y_j) belong to non-intersecting spacelike, but not necessarily achronal, hypersurfaces S_I (S_F) such that S_F is in the future of S_I as before. The diagrams we consider are such that any internal vertex is connected to an IN or OUT vertex, no IN (OUT) vertices are connected by a single propagator to an IN (OUT) vertex since otherwise there would exist an IN (OUT) vertex where a particle would arrive (leave) in contrast to the meaning of IN and OUT. What we state is that the correct interpretation is given by putting the IN vertices as first argument in the Feynman propagator and the OUT vertices as last argument; we don't care about a unique interpretation for the internal vertices.

We will proceed by writing things down in case the gauge group is $U(1)$ since that simplifies notation given that there is no charge attached to photon lines and ghost particles are absent; the general case, including ghosts, following immediately from the restricted one. Therefore, as explained before, the only interaction vertex or intertwiner is given by

$$e^{\mu}_{a}(x)(\gamma^{a})^{i}_{j}$$

which has no internal symmetries, so the symmetry factor of a diagram equals always one. An internal vertex with label k is therefore represented by a triple (μ_k, i_k, j_k) where the index j_k is covariant and the remaining two contravariant. Take then the series $(b_m, \ldots, b_1, (\mu_1, i_1, j_1), \ldots, (\mu_V, i_V, j_V), a_1, \ldots, a_n)$ where V represents the number of internal vertices and define the rule that the transposition of a space time index with any other index corresponds to plus one, while the transposition of a spinor index with another spinor index corresponds to minus one. Moreover, only covariant and contravariant spinor indices of different vertices can couple to one and another; then, the reader verifies that the overall sign of a diagram is well defined, taken into account the properties of the Fermi-Feynman propagator, and independent of the labelling of the internal vertices. With all this in mind, we write formally

$$\langle \mathrm{OUT}\,(y_1, b_1), \ldots, (y_m, b_m) | \mathrm{IN}\,(x_1, a_1), \ldots, (x_n, a_n)\rangle = \sum_{D} (-i\lambda)^{V} \epsilon(D)$$

$$\int dz_1 \sqrt{h(z_1)} \ldots \int dz_V \sqrt{h(z_V)} \prod \Delta_{F;\,c_l c_{p(l)}}(\alpha_l, \alpha_{p(l)}) \prod \Delta_{F,p}(\alpha_k, \alpha_{r(k)})^{j_{r(k)}}_{i_k} \prod (\gamma^{c_q})^{i_q}_{j_q}$$

where $\epsilon(D) = \pm 1$ is the sign of the diagram which has been fixed by the consistent choice for the particle Feynman propagator in the Fermi sector and $\alpha \in \{z_k, x_i, y_j\}$. I say formally, since experience has shown that the series does not converge albeit every diagram gives a finite contribution which we will show explicitely in chapter ten where we shall estimate the magnitude of a diagram. Corrections to unitarity should therefore occur and we will comment upon that in chapter eleven.

8.3 Gravitons.

We have not treated spin two particles so far yet neither from the point of abstract spin nor from the side of the two point function or Feynman propagator. We shall not treat this second issue in this chapter and we postpone it until chapter ten when proving finiteness of the respective Feynman diagrams. Therefore, we first treat spin and derive consequently the symmmetry group of the interacting theory. Here, Newton's constant will come into play giving rise to a coupling constant of the dimension of length, the so-called Planck length l_p. Given that a (massless) spin-one particle is described by means of a Lorentz vector, it is natural to look for a tensor product representation of the Lorentz group

$$\Lambda^a_b \Lambda^c_d h^{bd}$$

in order to isolate massless spin two particles invariant under an irreducible representation of the little group $E(2)$ associated to the lightlike momentum vector k. Regarding the entire Lorentz group, there exist two irreducible components, the symmetric and anti-symmetric tensors and the massless spin two particle resides in the symmetric part. Indeed, as is well known, we should look for symmetric states carrying helicity ± 2, There are exactly two of them $e_i \otimes e_i$ where e_i denotes the state of helicity $(-1)^i$ for $i = 1, 2$; furthermore $k \otimes e_i + e_i \otimes k$ denotes a zero norm particle of helicity $(-1)^i$ and likewise so for $l \otimes e_i + e_i \otimes l$. Finally, there are four states of helicity zero: one of positive norm given by $e_1 \otimes e_2 + e_2 \otimes e_1$ where the norm is given by $-\eta_{ab}$, two zero norm particles given by $k \otimes k$ and $l \otimes l$ and finally one of negative norm given by $k \otimes l + l \otimes k$. The little group of k leaves a six dimensional space invariant which is given by the symmetrization of k, e_i a space of two positive norm particles of helicity ± 2, two zero norm particles of helicity ± 1 and finally two particles of helicity zero, one of positive norm and the other of zero norm. In contrast to the probability theory for a massless spin-1 particle, where the longitudonal mode could be ignored because it is of zero norm, there is no reason to ignore the massless helicity zero particle given by $e_1 \otimes e_2 + e_2 \otimes e_1$. Therefore, we conclude that any theory for a massless particle of helicity ± 2 comes with a massless particle of helicity 0, in sharp contrast to the standard view upon spin.

Given this new result, we now come to the determination of the symmetry group of the graviton theory. The big distinction with gauge theory is that the generators of the diffeomorphism Lie-algebra act quasi-locally, instead of ultra-locally, on the "gauge potential" $h_{\mu\nu}$, where we have gotten from h_{ab} to $h_{\mu\nu}$ by means of the vierbein e^a_μ, associated to the Lorentzian space time metric $g_{\mu\nu}$. Indeed, the Lie algebra of the diffeomorphism group is given by the vectorfields $\mathbf{V}$ which are realized by means of the Lie-derivative

$$\delta_{\mathbf{V}} = \mathcal{L}_{\mathbf{V}}.$$

The Lie algebra is preserved given that

$$[\mathcal{L}_{\mathbf{V}}, \mathcal{L}_{\mathbf{W}}] = \mathcal{L}_{[\mathbf{V},\mathbf{W}]}.$$

The Lie derivative on a general tensor field $T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s}$ is given by

$$\mathcal{L}_{\mathbf{V}} T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s} = T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s;\alpha} V^\alpha - T^{\beta \dots \mu_r}_{\nu_1 \dots \nu_s} V^{\mu_1}_{;\beta} - \dots + T^{\mu_1 \dots \mu_r}_{\beta \dots \nu_s} V^{\beta}_{;\nu_1} + \dots$$

where we have used the Levi-Civita connection associated any space time metric. We now come to the definition of what we mean with a generally covariant theory: under the usual action of space time diffeomorphisms, the space time metric $g_{\mu\nu}$ as well as the graviton polarization $h_{\mu\nu}$ transform as

$$g \to g + \mathcal{L}_{\epsilon \mathbf{V}} g, \;\; h \to h + \mathcal{L}_{\epsilon \mathbf{V}} h.$$

Subsequent application gives

$$(g + \mathcal{L}_{\epsilon \mathbf{V}} g) + \mathcal{L}_{\epsilon \mathbf{W}}(g + \mathcal{L}_{\epsilon \mathbf{V}} g) = g + \mathcal{L}_{\epsilon(\mathbf{V}+\mathbf{W})} g + \mathcal{L}_{\epsilon \mathbf{W}} \mathcal{L}_{\epsilon \mathbf{V}} g$$

and the property

$$[\delta_{\epsilon \mathbf{V}}, \delta_{\epsilon \mathbf{W}}] = \delta_{\epsilon^2 [\mathbf{V},\mathbf{W}]}$$

is needed for this to be an action. In order for $g_{\mu\nu}$ to remain *stationary* we therefore form the combination

$$g_{\mu\nu} + l_p h_{\mu\nu}$$

and define

$$\delta'_{\epsilon \mathbf{V}} h = \delta_{\epsilon \mathbf{V}} h + (l_p)^{-1} \delta_{\epsilon \mathbf{V}} g, \;\; \delta'_{\epsilon \mathbf{V}} g = 0$$

where the Plank length has been inserted because the graviton propagator has dimension mass^2. It is readily verified that

$$\delta'_{\epsilon(\mathbf{V}+\mathbf{W})} = \delta'_{\epsilon \mathbf{V}} + \delta'_{\epsilon \mathbf{W}}$$

and

$$[\delta'_{\epsilon \mathbf{V}}, \delta'_{\epsilon \mathbf{W}}] = \delta'_{\epsilon^2 [\mathbf{V},\mathbf{W}]}$$

given that

$$\delta'_{\epsilon \mathbf{V}} \delta'_{\epsilon \mathbf{W}} = \delta_{\epsilon \mathbf{V}} \delta'_{\epsilon \mathbf{W}}.$$

The symmetries of a graviton theory require that internal interaction vertices between gravitons are constructed from scalar densities under the action δ' while interactions with ghost particles are constructed from tensor densities under the action δ. The rationale is the same as the one in non-abelian gauge theory where one adds all *covariant* interaction terms which do not stem from a local gauge symmetric scalar density to the theory and couples them to ghost particles.

As is well known, the interaction vertices and two point function are all we need to define a generally covariant quantum theory; we do not have any problems regarding the definition of a covariant measure.

8.4 Comments.

The picture we arrived at in this chapter is one of colliding *information* currents where, at each instant, the point of collision is uncertain and, therefore, should be integrated over. Moreover, one should sum over all diagrams given that any collision pattern should contribute. Theories do not come in any simpeler form than this and the construction in this chapter is therefore almost self-evident. The only non-obvious part being the imaginary nature of the coupling constants associated to the internal vertices as well as the uniform measure attached to the diagrams. Why should the constant in front of the contribution of a Feynman diagram not depend upon the number of internal vertices as well on the size of the latter? The principle that they don't is called "unitarity" and we shall test this assumption in chapter eleven, where we will study convergence or analycity properties of the series. I realize of course that this constitutes a deviation away from traditional quantum mechanics; but one should simply accept *this* formulation as the proper one and forget about the attempts made by Dirac, Schroedinger and Heisenberg.

So, the content of this chapter is very much like the presentation in chapters five and six: preliminary and in need of closer inspection. In chapter seven, we already saw what was needed to get finite Feynman diagrams out for spin-0 particles; in the next section, we will understand what kind of global constraints on the geometry are needed to make everything well defined for spin-$\frac{1}{2}, 1, 2$ particles as well. Albeit we shall work by means of a simple prototype cosmology, a lot is to be learned from this example and it will provide us with the crucial ingredients and insights.

The reader must wonder, given that we have computed an amplitude between a process of birth of n-particles at separated locations x_i and annihilation of m-particles later on, in the process view, also at separated locations z_j, how to calculate the correct weights serving in a relative probability interpretation. The reader will immediately understand that this concerns a natural extension of the theory laid out in chapters five and six; all particles in the IN state are created in a certain state $\sum_k \otimes_{i=1}^{n} \Psi^k_{x_i}$ where each $\Psi^k_{x_i}$ is defined relative to the space time point of creation x_i. All IN particles are "measured" leaving the source on spacelike hypersurfaces $\Sigma_i \subset R(S_I, S_F)$ which are spatially separated from one and another in the causal relationship restricted to $R(S_I, S_F)$. Moreover, they are measured by m "irreducible" measurement apparati given by world tubes $W_{\Sigma'_j}$ such that $z_j \in W_{\Sigma'_j}$. Hence, we have to calculate the modified amplitudes

$$\langle \mathrm{OUT}\,(y'_j, b_j), j = 1 \ldots m | \mathrm{IN}\,(y_i, a_i), i = 1 \ldots n\rangle_{\mathrm{phys}} =$$

$$\sum_D \prod_{k=1}^{V} \int_{\mathcal{M}} dz_k \sqrt{g(z_k)} \left(\prod_{\text{bosonic IN lines}\, i} T_{\alpha(i), e_0 \perp \Sigma_i} \right) A(D, (y'_j, b_j), (y_i, a_i), (z_k, c^r_k))$$

where r is a degeneracy index allowing an internal vertex to appear more than once and $\alpha(i) \in \{(y'_j, b_j), (z_k, c^r_k)\}$ denoting the endpoint of the IN-line associated to the boson born with parameters (y_i, a_i). Also, $y_i \in \Sigma_i$ and all IN quantum numbers are defined with respect to the $SO(3)$-class of reference frames given be $e_0 \perp \Sigma_i$; moreover, the definition of the $T_{\alpha(i), e_0 \perp \Sigma_i}$ operator has been canonically extended to accomodate for the Feynman propagator Δ_F and the reader should fill in these details. Moreover $y'_j \in \Sigma'_{j;t-\delta}$ where $z_j \in \Sigma'_{j;t} \subset S_F$; A is defined in such a way that if we drop the product

$$\prod_{\text{bosonic IN lines}\, i} T_{\alpha(i), e_0 \perp \Sigma_i}$$

then the expression reduces to the usual one. We are now in a position to define

$$\langle \mathrm{OUT}\,(y'_j, b_j), j = 1 \ldots m | \mathrm{IN} \sum_k \otimes_{i=1}^n \Psi^k_{x_i}, \Sigma_i, i = 1 \ldots n\rangle_{\mathrm{phys}} = \sum_k \left(\prod_{\text{all IN lines}\, i} \int_{\Sigma_i} dy_i \sqrt{h(y_i)} \right)$$

$$\mathrm{PropagationOperator}\left[\otimes_{i=1}^n \Psi^k_{x_i}(y_i); \langle \mathrm{OUT}\,(y'_j, b_j), j = 1 \ldots m | \mathrm{IN}\,(y_i, a_i), i = 1 \ldots n\rangle_{\mathrm{phys}}\right]$$

where this "propagation operator" has been defined before in chapters five and six. Given the definition of the y'_j regarding z_j, it is now clear from previous considerations how to compute the weight of detection at $\Sigma'_{j;t} \subset S_F$; the latter should be computed in the tensor product of the one particle spaces associated to z_j. For fermions, we recall that z_j is not necessarily interpreted as a point of annihilation which concludes the discussion of physical weights.

I realize that I have hidden some details in the notion above, but it should be clear what the expressions mean and how they should be calculated. For example, the definition of "propagation operator" differs for particles of spin-$0, \frac{1}{2}, 1, 2$ in such a way that I cannot use a unified notation. Nevertheless, we have treated the issue of propagation for single particle waves in full detail in all these cases; therefore, it should be clear what the definition is. The only point which might need some clarification is the extension of the T_x operator to the Feynman propagator $\Delta_F(x, y)$ for integer spin particles; the rule is that one has to multiply plane waves starting at x or y with the square root of the component of the four momentum with respect to e_0 as defined in y; also, all projection operators for waves of higher spin have to be executed in y. This finishes this chapter; ultimately, it are these physical quantities one has to compute and not the naive ones defined by

$$\langle \mathrm{OUT}\,(z_j, b_j), j = 1 \ldots m | \mathrm{IN}\,(x_i, a_i), i = 1 \ldots n \rangle.$$

However, the analysis regarding bounds on these amplitudes as performed in chapter ten remains identical and we proceed with the "naive" quantities in the remainder of this book.

Chapter 9

Study of an alternative vacuum cosmology.

We have shown so far in chapter seven on the regularization of the Feynman propagator that Minkowski space time is not suited to define a relativistic quantum theory in, given that it does not determine a dynamical notion of time and therefore does not allow for the necessary friction terms to be defined in a way which does not directly depend upon the observer. Before we proceed, some words of physical significance are in place, in a Schwarzschild and Kerr-Newman rotationally symmetric black hole solution we can speak of a null Killing horizon, which coincides with the union of black hole surfaces defined by Hawking, where our preferred timelike vectorfield, or gravitational arrow of time, becomes null and therefore quantum theory becomes ill defined again. It may be clear that generic perturbations in the initial data, even smooth ones of compact support, will destroy the Killing Horizon *and* most likely, also the strongly future asymptotically predictable character of the space time. Indeed, to my knowledge, the issue of stability regarding the very definition of an event horizon by means of the past of the boundary of the asymptotic future in some conformal space time has not been properly examined. I really do not care much about it, as I have always found this definition rather contrarian and "unphysical" to some extend (given that in quantum gravity the future is not known at all). What our thoughts above reveal is that Kerr-Newman space times also cannot serve as a background for quantum theory as the Lebesgue well-definedness of the propagator goes havoc on the horizon and also within. One might again want to resort to weaker, dual, interpretations as before but it could be that the old problems of Minkowski come back in some different jacket. With those words of caution, we now proceed to the definition of the two-point function on the $k = 0$ or spatially flat Friedmann universe, which in the case of interest serves as an alternative vacuum.

9.1 A cosmological vacuum.

The metric is given by

$$ds^2 = dt^2 - a^2(t)(dx^2 + dy^2 + dz^2)$$

and the Einstein equations with cosmological constant $\Lambda' = 3\Lambda$ and homogeneous isotropic fluid reduce to

$$3\frac{\dot{a}^2}{a^2} = 8\pi\rho + 3\Lambda$$

and

$$\frac{3\ddot{a}}{a} = -4\pi(\rho + 3p) + 3\Lambda.$$

The energy momentum conservation law reads

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0$$

while the geodesic equation equals

$$\frac{d^2t}{ds^2} + \dot{a}a\left|\frac{d\vec{x}}{ds}\right|^2 = 0,\ \frac{d^2\vec{x}}{ds^2} + 2\frac{\dot{a}}{a}\frac{dt}{ds}\frac{d\vec{x}}{ds} = 0.$$

In this section, we shall be interested in the cosmological vacuum defined by $\rho = p = 0$; in that case, the scale factor reads

$$a(t) = \alpha e^{\sqrt{\Lambda}t}$$

with $\alpha > 0$ and the Ricci tensor is given by

$$R_{\alpha\beta} = -3\Lambda g_{\alpha\beta}$$

in other words, our cosmology is an Einstein space. Performing the coordinate transformation $\tilde{t} = \frac{e^{-\sqrt{\Lambda}t}}{\alpha\sqrt{\Lambda}}$ leads to the expression

$$ds^2 = \frac{1}{\tilde{t}^2\Lambda}(d\tilde{t}^2 - dx^2 - dy^2 - dz^2)$$

which shows that our Einstein space is conformally flat. It is also a space of constant negative *sectional* curvature as the Riemann tensor takes on the form

$$R_{\alpha\beta\mu\nu} = -\Lambda\left(g_{\alpha\mu}g_{\beta\nu} - g_{\alpha\nu}g_{\beta\mu}\right)$$

a property which will be most convenient later on when performing our Wick rotation. It is nevertheless not a maximally symmetric space time such as is the case for a de-Sitter space time. Taking $\tilde{t}$ as a time coordinate suggests a big crunch while the t coordinate hints to an exponentially expanding universe.

They both determine the same unit norm timelike vectorfield up to a time orientation, which explains the qualitative difference; in the sequel, we will keep on working in the t, x, y, z instead of in the $\tilde{t}, x, y, z$ system. Further specialization of the geodesic equation leads to

$$\frac{d^2 t}{ds^2} + \sqrt{\Lambda}\alpha^2 e^{2\sqrt{\Lambda}t} \left| \frac{d\vec{x}}{ds} \right|^2 = 0$$

and

$$\frac{d^2 \vec{x}}{ds^2} + 2\sqrt{\Lambda}\frac{dt}{ds}\frac{d\vec{x}}{ds} = 0$$

from which it can be deduced that

$$\left| \frac{d\vec{x}}{ds} \right| = \beta e^{-2\sqrt{\Lambda}t}$$

with $\beta \geq 0$. These equations show that the affine time derivative slows down so that one may wonder wether it is possible to get at $t = +\infty$ in the first place. As we will show, this is the case for future oriented timelike geodesics but *not* so for spacelike geodesics for which the $\frac{dt}{ds} > 0$ part of the solution has a finite future t and s extend. One obtains the Newtonian law

$$\frac{d^2 t}{ds^2} + \sqrt{\Lambda}(\alpha\beta)^2 e^{-2\sqrt{\Lambda}t} = 0$$

which can be integrated to give

$$\frac{e^{-\sqrt{\Lambda}t}}{\sqrt{\frac{\delta}{\alpha^2\beta^2} + e^{-2\sqrt{\Lambda}t}} + \frac{\sqrt{\delta}}{\alpha\beta}} = e^{-\sqrt{\delta\Lambda}(s+\gamma)}$$

where $\alpha, \beta, \delta \geq 0$ and $\gamma \in \mathbb{R}$. This, again, leads to

$$t(s) = -\frac{1}{\sqrt{\Lambda}} \ln \left(\sqrt{\frac{4\delta}{\alpha^2\beta^2}} \frac{e^{-\sqrt{\delta\Lambda}(s+\gamma)}}{1 - e^{-2\sqrt{\delta\Lambda}(s+\gamma)}} \right)$$

and $\gamma > 0$. It is clear that for $s < -\gamma$ the space time is past geodesically incomplete, unless $\gamma = +\infty$, while for s to plus infinity, we obtain again an approximate linear relation between t and s. The geodesic equation for the spatial part then becomes

$$\frac{d^2 \vec{x}}{ds^2} + 2\sqrt{\delta\Lambda}\frac{1 + e^{-2\sqrt{\delta\Lambda}(s+\gamma)}}{1 - e^{-2\sqrt{\delta\Lambda}(s+\gamma)}}\frac{d\vec{x}}{ds} = 0$$

which leads to

$$\frac{d\vec{x}}{ds} = \vec{\beta}\frac{4\delta}{\alpha^2\beta^2}\frac{e^{-2\sqrt{\delta\Lambda}(s+\gamma)}}{(1 - e^{-2\sqrt{\delta\Lambda}(s+\gamma)})^2}$$

where $\left|\vec{\beta}\right|^2 = \beta^2$. This last formula may again be integrated to yield

$$\vec{x}(s) = \vec{r}_0 - 2\vec{\beta}\sqrt{\frac{\delta}{\Lambda}}\frac{1}{\alpha^2\beta^2}\frac{1}{1-e^{-2\sqrt{\delta\Lambda}(s+\gamma)}}$$

where, in the limit for β to 0, $\vec{r}_0$ has to renormalize by an infinite constant too. As it turns out, we have only given a parametrization for future oriented causal geodesics; in terms of the initial values x and $v = (\frac{dx}{ds})_{s=0}$ the original parameters read

$$\begin{aligned}
\vec{\beta} &= \vec{v}e^{2\sqrt{\Lambda}t} \\
e^{-\sqrt{\delta\Lambda}\gamma} &= \frac{1}{\alpha e^{\sqrt{\Lambda}t}\left|\vec{v}\right|}\left(v - \sqrt{v^2 - \alpha^2 e^{2\sqrt{\Lambda}t}\left|\vec{v}\right|^2}\right) \\
\delta &= v^2 - \alpha^2 e^{2\sqrt{\Lambda}t}\left|\vec{v}\right|^2 \\
\vec{r}_0 &= \vec{x} + \frac{\vec{v}}{\sqrt{\Lambda}(v - \sqrt{v^2 - \alpha^2 e^{2\sqrt{\Lambda}t}\left|\vec{v}\right|^2})}
\end{aligned}$$

so in the limit of Λ to zero $\vec{r}_0$ renormalizes $\vec{x}_0$ by an infinite amount. One notices that δ has the geometric significance of the length squared of the tangent vector of the geodesic at x which we may put to one since we deal with timelike geodesics. This further simplifies our formulae to

$$\begin{aligned}
e^{-\sqrt{\delta\Lambda}\gamma} &= \sqrt{\frac{v-1}{v+1}} \\
\vec{r}_0 &= \vec{x} + \frac{\vec{v}}{\sqrt{\Lambda}(v-1)}
\end{aligned}$$

and with these reservations, we obtain that

$$\begin{aligned}
t(s) &= -\frac{1}{\sqrt{\Lambda}}\ln\left(\frac{2e^{-\sqrt{\Lambda}(t+s)}}{v+1-(v-1)e^{-2\sqrt{\Lambda}s}}\right) \\
\vec{x}(s) &= \vec{x} + \frac{\vec{v}}{\sqrt{\Lambda}(v-1)} - \frac{2\vec{v}}{\sqrt{\Lambda}(v-1)\left(v+1-(v-1)e^{-2\sqrt{\Lambda}s}\right)}.
\end{aligned}$$

From the first equation, one can solve v in function of $z = e^{-\sqrt{\Lambda}s}$; the formula is given by

$$v = \frac{2ze^{\sqrt{\Lambda}(t'-t)} - 1 - z^2}{1-z^2}$$

with $z > e^{-\sqrt{\Lambda}(t'-t)}$. Insertion into the second equation fixes z by the polynomial

$$z^2 + 1 - \left(2\cosh(\sqrt{\Lambda}(t'-t)) - \Lambda\left|\vec{x'} - \vec{x}\right|^2 \alpha^2 e^{\sqrt{\Lambda}(t'+t)}\right)z = 0$$

where the evaluation holds for $(t', \vec{x}')$ future timelike related to $(t, \vec{x})$. Notice that we have an asymptotic region of radius $\frac{1}{\sqrt{\Lambda}\alpha e^{\sqrt{\Lambda}t}}$, so unlike Minkowski space time, in our vacuum cosmology, it is impossible for $\vec{x}'$ to become infinite and therefore any observer has a nontrivial horizon. It is easy to solve our equation to

$$s = -\frac{1}{\sqrt{\Lambda}} \ln \left(g(x, x'; \Lambda, \alpha) - \sqrt{g(x, x'; \Lambda, \alpha)^2 - 1} \right)$$

where

$$g(x, x'; \Lambda, \alpha) = \cosh(\sqrt{\Lambda}(t' - t)) - \Lambda \left| \vec{x'} - \vec{x} \right|^2 \frac{\alpha^2 e^{\sqrt{\Lambda}(t'+t)}}{2}.$$

In the limit for $\sqrt{\Lambda}$ to zero, this expression becomes

$$s_0^2 = \lim_{\sqrt{\Lambda} \to 0} \frac{\left((t' - t) \sinh(\sqrt{\Lambda}(t' - t)) - \sqrt{\Lambda} \left| \vec{x'} - \vec{x} \right|^2 \alpha^2 e^{\sqrt{\Lambda}(t'+t)} + O(\lambda) \right)^2}{g(x, x'; \Lambda, \alpha)^2 - 1} = (t' - t)^2 - \alpha^2 \left| \vec{x}' - \vec{x} \right|^2$$

as it should be. This formula can be easily analytically continued to the region

$$-1 < g(x, x'; \Lambda, \alpha) < 1$$

by

$$is' = -\frac{1}{\sqrt{\Lambda}} \ln \left(g(x, x'; \Lambda, \alpha) - i\sqrt{1 - g(x, x'; \Lambda, \alpha)^2} \right)$$

where we have made the branch cut for the complex square root in the upper half plane at for example $\frac{\pi}{2}$. It is then easily computed that

$$-s'(x, x'; \Lambda, \alpha)^2 = -\frac{1}{\Lambda} (\arccos{(g(x, x'; \Lambda, \alpha))})^2$$

and one can again check that the $\sqrt{\Lambda}$ to zero limit is given by

$$-s_0'(x, x'; \alpha)^2 = (t' - t)^2 - \left| \vec{x}' - \vec{x} \right|^2 \alpha^2$$

as it should, so our formula is entirely correct. One can easily see that this result comes by considering the case $\delta < 0$ which corresponds to spacelike geodesics which live a finite amount of time t in the future as well as a finite amount of affine parameter time s in the past and the future. This is again a distinction with Minkowski which is geodesically complete and where spacelike geodesics reach out to infinite values of time in the future. The relevant formulae are deduced by performing the analytic continuation to $\delta < 0$ and putting $\delta = -1$:

$$\begin{aligned} t(s) &= -\frac{1}{\sqrt{\Lambda}} \ln \left(\frac{\sqrt{-\delta}}{\alpha\beta \sin(\sqrt{-\delta\Lambda}(s + \gamma))} \right) \\ \vec{x}(s) &= \vec{r}_0 - \sqrt{\frac{-\delta}{\Lambda}} \frac{\vec{\beta}}{\alpha^2 \beta^2 \tan(\sqrt{-\delta\Lambda}(s + \gamma))}. \end{aligned}$$

As before

$$\begin{aligned}
\vec{\beta} &= \vec{v}e^{2\sqrt{\Lambda}t} \\
e^{i\sqrt{\Lambda}\gamma} &= \frac{v+i}{\sqrt{v^2+1}} \\
\vec{x} &= \vec{r}_0 - \frac{v\vec{v}}{\sqrt{\Lambda}(v^2+1)}.
\end{aligned}$$

This reshapes our solutions as

$$\begin{aligned}
t(s) &= -\frac{1}{\sqrt{\Lambda}}\ln\left(\frac{e^{-\sqrt{\Lambda}t}}{\sin(\sqrt{\Lambda}s)v+\cos(\sqrt{\Lambda}s)}\right) \\
\vec{x}(s) &= \vec{x} + \frac{v\vec{v}}{\sqrt{\Lambda}(v^2+1)} - \frac{\vec{v}(v-\tan(\sqrt{\Lambda}s))}{\sqrt{\Lambda}(v^2+1)(1+v\tan(\sqrt{\Lambda}s))}
\end{aligned}$$

and the reader notices that in the limit $\tan(\sqrt{\Lambda}s) = v$, our assumption $\frac{dt}{ds} \geq 0$ no longer holds. Nevertheless, this solution is past incomplete in the sense that for $s = \frac{1}{\sqrt{\Lambda}}\arctan(-\frac{1}{v})$ it diverges to $t = -\infty$ and $|\vec{x}| \to \infty$. This limit cannot be attained towards the future however and we notice that for $\tan(\sqrt{\Lambda}s) = v$ one has that $\frac{dt}{ds} = 0$ and for later times s, the geodesic evolves again towards lower $t(s)$ values. Our parameter domain reaches only up till $s = \frac{\pi}{2\sqrt{\Lambda}}$ at which point nothing special happens given that the limit of $\vec{x}$ as well as its derivatives are well defined if $\tan(\sqrt{\Lambda}s)$ blows up to infinity. Hence, we need to glue a new solution to the old one, which makes the construction of Synge's function for spacelike geodesics rather complicated. However, we proceed first by determining the world function for the above parametrization, giving the following formulae

$$\begin{aligned}
v &= \frac{e^{\sqrt{\Lambda}(t'-t)} - \cos(\sqrt{\Lambda}s)}{\sin(\sqrt{\Lambda}s)} \\
\Lambda\alpha^2\left|\vec{x}' - \vec{x}\right|^2 e^{2\sqrt{\Lambda}t} &= \frac{(v^2+1)\tan^2(\sqrt{\Lambda}s)}{(1+v\tan(\sqrt{\Lambda}s))^2}
\end{aligned}$$

which leads to

$$s'(x,x';\Lambda,\alpha) = \frac{1}{\sqrt{\Lambda}}\arccos(g(x,x;\Lambda,\alpha))$$

a result which we obtained previously by means of analytic continuation; this formula covers the full spacelike region as the maximal length of a spacelike geodesic equals $\frac{\pi}{\sqrt{\Lambda}}$ which is precisely the range of that function. It is interesting to study the limit for $v \to +\infty$ of our solution; from any starting point in space time one arrives at $t = +\infty$ in a parameter time $s = \frac{\pi}{2\sqrt{\Lambda}}$ at which $\frac{dt}{ds} = 0$ and still the limit of the tangent vectors has unit norm. This means, in particular, that in any direction of space one can trace back these data for smaller t values providing one with a null hypersurface of events in space time demarcating, within the region of events which can be connected by means of a spacelike curve to the initial point, those events which can be reached by a spacelike *geodesic* starting at x. In particular, this horizon is given by

$$|\vec{x}' - \vec{x}| = \frac{1}{\alpha\sqrt{\Lambda}e^{\sqrt{\Lambda}t}} + \frac{1}{\alpha\sqrt{\Lambda}e^{\sqrt{\Lambda}t'}}$$

and it obviously lies fully in the region

$$-1 < g(x, x'; \Lambda, \alpha) < 1.$$

This leads us to the following definition: given a space time point x, the spacelike geodesic horizon $HS(x)$ is the boundary of the region which can be reached by means of a spacelike geodesic. Likewise, we define the future timelike horizon $HT(x)$ at x as the boundary of the region of space time which can be reached by means of timelike geodesics. $HS(x)$ is not necessarily a null hypersurface as it the case for our cosmology and neither does $HT(x)$ need to coincide with the boundary of $J^+(x)$. Note that the outer part of $HS(x)$ coincides, in our case, with the boundary of $J^-(I^+(x))$ which is the standard horizon for timelike signals in a general cosmology. Hence, there is a region of space time which cannot be reached by any geodesic starting at x; this is a novel feature to be taken into account in the quantum theory which we shall do later on. We finish this section by making a comment upon the way the vectorfield e_0 is chosen from local physical considerations. The most obvious criterion is a *quasi*-local one which says that the Riemann curvature squared (or the Ricci curvature squared) of the Riemannian metric on the orthogonal spacelike hypersurface attains an absolute minimum 0. It may be that there exists some *ultra*-local criterium by looking for minima of some function in the space time Riemann tensor components evaluated in a tetrad with timelike vector given by ∂_t. The latter characterization would be preferred in my mind but we leave such fine points for the future.

9.2 The modified propagator on the new vacuum cosmology.

Before we come to the calculation of the two point function, we need to determine the parallel transporter $\Lambda_{\beta}^{\alpha'}(x, y)$ between two points; the latter is defined, as before, by means of transport of a vector along the unique geodesic connecting x with y. Before we come to the explicit computations, let us try to guess the structure of the result based upon symmetry considerations. As is well known $-\sigma_\mu(x, y)$ gives the tangent co-vector at x to the geodesic connecting x with y of length equal to the geodesic length; that is

$$g^{\mu\nu}(x)\sigma_\mu(x, y)\sigma_\nu(x, y) = 2\sigma(x, y)$$

where we have surpressed Λ, α in the notation of Synge's function $\sigma(x, y)$. For future convenience, let us denote by $e_0 = \partial_t$, $e_i = \frac{e^{-\sqrt{\Lambda}t}}{\alpha}\partial_i$ the standard tetrad which is constant under parallel transport on timelike geodesics of constant $\vec{x}$. Hence, the transporter expressed with respect to this tetrad $\Lambda_b^{a'}(x,y)$ is the unit matrix if y has the same space coordinate than x. More in general, one would expect $\Lambda_b^{a'}(x,y)$ to be a Lorentz boost determined by the $e_0, e_a\sigma^a(x,y)$ plane with a magnitude proportional to $\sqrt{\sum_i \sigma_i(x,y)^2}, \sigma^0(x,y)$ where $\sigma^a(x,y) = e^{a\mu}(x)\sigma_\mu(x,y)$ and it has been understood that the a index has been raised with the flat Minkowski metric η^{ab}. Let us now make the explicit computations; the transport equation is given by

$$\begin{aligned} \frac{d}{ds}Z^0(s) + \alpha^2\sqrt{\Lambda}e^{2\sqrt{\Lambda}t}\vec{v}(s).\vec{Z}(s) &= 0 \\ \frac{d}{ds}\vec{Z}(s) + \sqrt{\Lambda}\left(\vec{v}(s)Z^0(s) + \vec{Z}(s)v(s)\right) &= 0 \end{aligned}$$

where $v^\alpha(s)$ is the unit tangent to the geodesic in affine parametrization. From our solutions for timelike and spacelike geodesics, it is easy to see that initial vectors Z perpendicular to e_0 and $\vec{v}$ remain so which confirms our claim that unit vectors perpendicular to e_0 and $e_a\sigma^a(x,y)$ are left invariant for as well spacelike as timelike geodesics[1]. Remains to figure out the boost parameter; here we study the transport of $Z = e_0$. The fact that parallel transport preserves the norm allows us to write

$$Z(s) = (\cosh(\gamma(s)), \sinh(\gamma(s))\frac{\vec{v}(s)}{\sqrt{v(s)^2-1}})$$

for timelike geodesics with $\gamma(0) = x$. Hence, we obtain that the first transport equation reduces to

$$\frac{d\gamma(s)}{ds} = -\sqrt{(v(s)^2-1)\Lambda}$$

and taking the explicit formula for

$$v(s) = \frac{v + 1 + (v-1)e^{-2\sqrt{\Lambda}s}}{v + 1 - (v-1)e^{-2\sqrt{\Lambda}s}}$$

results in

$$\gamma(s) = \left(\ln\left(\frac{1+\sqrt{\frac{v-1}{v+1}}e^{-\sqrt{\Lambda}s}}{1-\sqrt{\frac{v-1}{v+1}}e^{-\sqrt{\Lambda}s}}\right) - \ln\left(\frac{1+\sqrt{\frac{v-1}{v+1}}}{1-\sqrt{\frac{v-1}{v+1}}}\right)\right).$$

Upon substitution by the well known formulae for v in function of t, t', s and s in function of $g(x,x';\Lambda,\alpha)$, we arrive after some algebra at

$$\sqrt{\frac{v-1}{v+1}} = \sqrt{\frac{e^{\sqrt{\Lambda}(t'-t)} - e^{\sqrt{\Lambda}s}}{e^{\sqrt{\Lambda}(t'-t)} - e^{-\sqrt{\Lambda}s}}}$$

[1]Invariant in the sense that the components only undergo a rescaling as to preserve the local norm.

and some rather complicated formula

$$
\begin{aligned}
\gamma(s) &= \ln\left(\frac{1-z^2}{\left(\sqrt{1-ze^{-\sqrt{\Lambda}(t'-t)}}-\sqrt{z^2-ze^{-\sqrt{\Lambda}(t'-t)}}\right)^2}\right) \\
&\quad -\ln\left(\frac{2ze^{\sqrt{\Lambda}(t'-t)}-1-z^2}{2ze^{\sqrt{\Lambda}(t'-t)}-1-z^2-2\sqrt{\left(z^2(e^{2\sqrt{\Lambda}(t'-t)}+1)-z^3e^{\sqrt{\Lambda}(t'-t)}-ze^{\sqrt{\Lambda}(t'-t)}\right)}}\right)
\end{aligned}
$$

where $z = g(x, x'; \Lambda, \alpha) - \sqrt{g(x, x'; \Lambda, \alpha)^2 - 1}$. A similar result holds for spacelike geodesics and the above calculations show already that exact calculations for the two point function will look rather messy. However, regarding the issue of convergence, we can make useful estimates and it is important to notice that

$$-\ln\left(\frac{1+\sqrt{\frac{v-1}{v+1}}}{1-\sqrt{\frac{v-1}{v+1}}}\right) \leq \gamma(s) \leq 0$$

for $s \geq 0$ meaning that in the limit for the affine parameter towards future infinity, the boost parameter converges to a finite negative value. Only in the limit for v towards infinity does $\gamma(s)$ converge to infinity too. Towards the past, $\gamma(s) \to +\infty$ if $t(s) \to -\infty$; for spacelike geodesics, one obtains a different qualitative result which is that in the limit for the affine time towards its finite negative and positive values (with a difference of $\frac{\pi}{\sqrt{\Lambda}}$), $\gamma(s)$ blows up towards minus infinity in the limit towards the positive value and to plus infinity in the limit towards the negative value.

We now come to the determination of the two point function and will denote the relevant formula in terms of first derivatives of Synge's function $\sigma_a(x, x'; \Lambda, \alpha)$ and the boost parameter

$$\gamma(x, x'; \Lambda, \alpha).$$

There is no need to use their explicit expressions to arrive at the desired results and if the reader wants to, he or she can manipulate the final expressions by substituting for the above obtained formulae. The two point function we shall study is given by

$$W_\mu(x, x'; \Lambda, \alpha) = \int \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\theta(k^0)e^{-ik^a\sigma_a(x,x';\Lambda,\alpha)}e^{-\mu(k^0)^2-\mu(\Lambda_a^{0'}(x,x';\Lambda,\alpha)k^a)^2}$$

where x' is causally related to x, since otherwise we would have to include reflection symmetric terms, and $\Lambda_a^{0'}(x, x'; \Lambda, \alpha)$ is given by

$$\Lambda_a^{0'}(x, x'; \Lambda, \alpha)k^a = \cosh(\gamma(x, x'; \Lambda, \alpha))k^0 + \sinh(\gamma(x, x'; \Lambda, \alpha))\frac{\vec{k}.(\vec{x}' - \vec{x})}{|\vec{x}' - \vec{x}|}$$

and x' is supposed to lie within the total geodesic horizon of x (here the total geodesic horizon is defined as the boundary of the set of events which can be reached from x by means of a geodesic). In chapter seven, we studied this integral in Minkowski space time where ∂_t has to be associated to the timelike vectorfield defined by some physical observer making the quantum particle feel an eather due to the him or herself. We shall compute here *explicitely* that precisely the same issues show up and go somewhat deeper into the nature of the Wick transformation. Denoting by $\vec{\sigma}(x,x';\Lambda,\alpha) = (\sigma_i(x,x';\Lambda,\alpha))$, where the i index refers to the spatial part of the vierbein and not to the space components of σ_μ, and correspondingly

$$|\vec{\sigma}(x,x';\Lambda,\alpha)| = \sqrt{\sum_i \sigma_i(x,x';\Lambda,\alpha)^2}$$

we arrive, after some algebra, to

$$\begin{aligned} W_\mu(x,x';\Lambda,\alpha) &= \frac{1}{8\pi^2}\int_0^\infty dk \frac{k}{\sqrt{k^2+m^2}}\int_{-k}^{k} dz e^{-i\sqrt{k^2+m^2}\sigma_0 - \mu(1+\cosh^2(\gamma))(k^2+m^2)} \\ & \quad e^{-\mu\sinh^2(\gamma)\left(z+\left(\frac{\cosh(\gamma)}{\sinh(\gamma)}\sqrt{k^2+m^2}+i\frac{|\vec{\sigma}|}{2\mu\sinh^2(\gamma)}\right)\right)^2} e^{\mu\sinh^2(\gamma)\left(\frac{\cosh(\gamma)}{\sinh(\gamma)}\sqrt{k^2+m^2}+i\frac{|\vec{\sigma}|}{2\mu\sinh^2(\gamma)}\right)^2} \end{aligned}$$

where we have surpressed all dependencies upon x, x', Λ, α in the right hand side. At this point, it is instructive to give some comment about the general structure of the integral. The μ surpression terms we included are sufficient for our purposes just as it is the case for Minkowki. This property is rather independent of the behavior of γ which we have shown to converge to an asymptotic, finite negative value in the limit of the parameter time towards plus infinity for future timelike related events. It may be better to replace the $(V_a k^a)^2$ surpression term by a $h_{ab}k^a k^b$ surpression where h_{ab} is, as before, the Riemannian metric determined by the timelike vectorfield. It is immediately seen that the absolute value of $W_\mu(x,x';\Lambda,\alpha)$ is bounded by a universal constant proportional to $\frac{1}{\mu}$, which is actually sufficient for our proof of finiteness since we have to take into account the Riemannian surpression term due to κ. However, we are interested in more detailed properties of this function and carry on.

Coming back to the calculation of $W_\mu(x,x';\Lambda,\alpha)$, the integral over z is a Gaussian one which cannot be exactly done, but to which we can find a useful upper bound. In particular, we estimate integrals of the type

$$F(k,c) = \int_{a(k)}^{b(k)} dz e^{-a(z+ic)^2}$$

for $c \geq 0$. Taking the differential of $F(k,c)$ with regards to c results in

$$\frac{d}{dc}F(k,c) = i\int_{a(k)}^{b(k)} \frac{d}{dz} e^{-a(z+ic)^2} = i\left(e^{-a(b(k)+ic)^2} - e^{-a(a(k)+ic)^2}\right).$$

Therefore we obtain that

$$|F(k,c)| \leq \int_0^c dz e^{az^2} \left(e^{-a\, b(k)^2} + e^{-a\, a(k)^2} \right) + \frac{\sqrt{\pi}}{\sqrt{a}}$$

and upon using our previous results, the latter expression reduces to

$$|F(k,c)| \leq \frac{1}{acg(\sqrt{a}c)} \left(e^{ac^2} - 1 \right) \left(e^{-a\, b(k)^2} + e^{-a\, a(k)^2} \right) + \frac{\sqrt{\pi}}{\sqrt{a}}.$$

For the purpose of asymptotic analysis, we may clearly ignore the constant on the right hand side, since the resulting expressions converge exponentially fast in the limit for $|\vec{\sigma}|$ towards infinity, and we obtain that

$$\begin{aligned} |W_\mu(x,x';\Lambda,\alpha)| \quad \sim \quad & \frac{1}{4\pi^2 \, |\vec{\sigma}|} \int_0^\infty dk \frac{k}{\sqrt{k^2+m^2}} \\ & e^{-\mu(k^2+m^2)} \left(e^{-\mu \sinh^2(\gamma)\left(k + \frac{\cosh(\gamma)}{\sinh(\gamma)}\sqrt{k^2+m^2}\right)^2} + e^{-\mu \sinh^2(\gamma)\left(k - \frac{\cosh(\gamma)}{\sinh(\gamma)}\sqrt{k^2+m^2}\right)^2} \right). \end{aligned}$$

which shows that $W_\mu(x,x';\Lambda,\alpha)$ converges to zero in the limit for $|\vec{\sigma}|$ to infinity for x' future causally related to x. It is much harder to obtain an estimate in case $|\vec{\sigma}|$ remains finite but σ_0 blows up to plus infinity. The only result I am able to obtain is that of convergence in σ_0 along $|\vec{x}' - \vec{x}| = 0 = \vec{\sigma}$ and $\gamma = 0$ as $\frac{1}{\sigma_0}$.

We now turn our head towards the study of the impact of κ on $W_{\mu,\kappa}(x,x';\Lambda,\alpha)$. Denote by

$$E(x,x';\Lambda,\alpha,\kappa) = e^{-\kappa \int_0^{\tilde{s}} \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}}}$$

the exponentiated energy along the timelike geodesic connecting x with x', then

$$W_{\mu,\kappa}(x,x';\Lambda,\alpha) = E(x,x';\Lambda,\alpha,\kappa) W_\mu(x,x';\Lambda,\alpha)$$

and, in case $|\vec{x}' - \vec{x}| = 0$, then one has

$$E(x,x';\Lambda,\alpha,\kappa) = e^{-\kappa\left|t'-t\right|}.$$

In order for every subintegral of

$$\alpha^3 \int dx' e^{3\sqrt{\Lambda}t'} \left|\Delta_{F,\mu,\kappa}(x,x';\Lambda,\alpha)\right|^n$$

to be finite, it is therefore necessary that $\kappa > 3\sqrt{\Lambda}$, a condition which did not appear in Minkowski space time. Regarding the proof of perturbative finiteness, we will require some other bound to which we will come back to in a short while. Actually, without any further computation, the reader should realize that our cosmology behaves very different from ordinary Minkowski; on one side, one has the existence of all horizons and on the other, one notices that Minkowski can be conformally compactified while the Friedmann cosmology can't. The latter feature causes scattering processes in the future to occur with a higher amplitude which might ultimately not be surpressed anymore by our geodesic energy terms $E(x, x'; \Lambda, \alpha, \kappa)$. This would forbid Type III quantum theories but not Type II or Type I; in Minkowski space time, there is no such distinction between the past and the future and therefore, such behavior is not to be expected. As it will turn out, Type III quantum theories are allowed for as long as κ is sufficiently large. Coming back to our computation, one immediately sees that

$$\int ds \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} = \int_0^{\sqrt{2\sigma(x,y)}} ds \sqrt{2\left(\frac{dt}{ds}\right)^2 - 1}$$

where

$$\frac{dt}{ds} = \frac{v + 1 + (v-1)e^{-2\sqrt{\Lambda}s}}{v + 1 - (v-1)e^{-2\sqrt{\Lambda}s}}$$

an expression wich decreases from v to 1 at $s = \infty$. In Minkowski $\Lambda = 0, \alpha = 1$ and this expression equals $\sqrt{2(\sigma^0(x,y))^2 - 2\sigma(x,y)} = |x - y|$; for a cosmological space time this is very different. In general, we have that,

$$\int_0^{\sqrt{2\sigma}} ds \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} \geq \sqrt{2\sigma} \sim |t' - t|$$

for $|t' - t|$ large and $|\vec{x}' - \vec{x}| < \frac{e^{-\sqrt{\Lambda}t}}{\alpha\sqrt{\Lambda}}$ fixed. Moreover, the inequalities and similarities become equalities in the limit for σ to infinity. Note that σ is infinite within the lightcone and zero on the lightcone in the limit for t' towards ∞, but the pathology on the lightcone needs to be studied further. Actually, one obtains that the energy increases from the symmetrical point $|\vec{x}' - \vec{x}| = 0$ towards the boundary of the lightcone along the "hyperbola" of constant σ which is contained within a domain of compact $\vec{x}'$. We need a finer estimate in order to obtain conclusive results on convergence; some algebra shows that

$$\begin{aligned} \int_0^{\sqrt{2\sigma}} ds \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} \quad &\geq \quad \frac{1}{\sqrt{\Lambda}} \sqrt{\frac{v+1}{v-1}} \\ &\left(\ln\left(\frac{1 + \sqrt{\frac{v-1}{v+1}}}{1 + \sqrt{\frac{v-1}{v+1}} e^{-2\sqrt{2\Lambda\sigma}}} \right) + \ln\left(\frac{1 - \sqrt{\frac{v-1}{v+1}} e^{-2\sqrt{2\Lambda\sigma}}}{1 - \sqrt{\frac{v-1}{v+1}}} \right) \right) \end{aligned}$$

upon substitution of v by

$$v = \frac{2e^{-\sqrt{\Lambda}(\sqrt{2\sigma} - (t'-t))} - 1 - e^{-2\sqrt{2\Lambda\sigma}}}{1 - e^{-2\sqrt{2\Lambda\sigma}}}.$$

In order to study the σ to zero limit, we only need to take into account the second term; this one reduces in leading order to

$$\frac{1}{\sqrt{\Lambda}} \ln\left(\frac{3 + \frac{1}{1 - e^{-\sqrt{\Lambda}(t'-t)}} + \frac{1}{e^{\sqrt{\Lambda}(t'-t)} - 1}}{\frac{1}{1 - e^{-\sqrt{\Lambda}(t'-t)}} + \frac{1}{e^{\sqrt{\Lambda}(t'-t)} - 1} - 1} \right)$$

meaning that for large $|t' - t|$ this expression behaves approximately as $|t' - t| + \frac{\ln(4)}{\sqrt{\Lambda}}$ which is all we need. Actually, due to the nature of the Riemannian metric, we immediately have a lower bound of $|t' - t|$ on the (Lorentzian) energy and an upper bound on the *Riemannian* distance of $|t' - t| + \frac{1}{\sqrt{\Lambda}}$; the constant of $\frac{\ln(4)}{\sqrt{\Lambda}}$ is the only nontrivial thing in the above formula and the reader can easily see that this estimate is very accurate. This means that in the limit for σ equal to zero and $|t' - t|$ towards infinity, the exponentiated energy goes as

$$E(x, x'; \Lambda, \alpha, \kappa) = \frac{1}{(\sigma^0)^{\frac{\kappa}{\sqrt{\Lambda}}}}$$

something which falls quicker off than $\frac{1}{(\sigma^0)^3}$ given our previous bound on κ.

Towards the past, we have that the local energy is an increasing quantity and

$$\infty > \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} \geq \sqrt{2v^2 - 1}$$

which means that the energy is larger than

$$\sqrt{2(\sigma^0)^2 - 2\sigma}.$$

Akin to the future timelike case, this lower bound is actually insufficient as in the limit for $t'(s)$ to minus infinity, one obtains that

$$\sigma^0 = \sqrt{2\sigma} \frac{1 + e^{2\sqrt{2\Lambda\sigma}}}{e^{2\sqrt{2\Lambda\sigma}} - 1}$$

which converges to $\frac{1}{\sqrt{\Lambda}}$ in the limit for σ to zero. Just like in the previous case, one could perform the full integration,

$$\begin{aligned}\int_{-\sqrt{2\sigma}}^{0} ds \sqrt{h_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds}} \quad &\geq \quad \frac{1}{\sqrt{\Lambda}} \sqrt{\frac{v+1}{v-1}} \\ &\left(\ln \left(\frac{1 + \sqrt{\frac{v-1}{v+1}} e^{2\sqrt{2\Lambda\sigma}}}{1 + \sqrt{\frac{v-1}{v+1}}} \right) + \ln \left(\frac{1 - \sqrt{\frac{v-1}{v+1}}}{1 - \sqrt{\frac{v-1}{v+1}} e^{2\sqrt{2\Lambda\sigma}}} \right) \right)\end{aligned}$$

where

$$v = \frac{1 + e^{2\sqrt{2\Lambda\sigma}} - 2e^{\sqrt{\Lambda}(t' - t + \sqrt{2\sigma})}}{e^{2\sqrt{2\Lambda\sigma}} - 1}$$

or simply remark that the energy is always greater or equal to $|t' - t|$, which is all we actually need.

Similar convergence properties apply for spacelike geodesics, as the reader may want to verify for himself which finishes the discussion of this section. The only important conclusion is that the energy is always larger that the t' distance travelled which is sufficient to obtain convergent integrals. There remains something to be said about the Riemannian metric $h_{\alpha\beta}$ associated to our cosmological space time: it is a metric of constant negative sectional curvature $-\Lambda$ and therefore, balls in this metric have a volume which blows up at most exponentially fast in the radius due to a well known theorem in Riemannian geometry. Our Riemannian space has constant sectional curvature but is again not maximally symmetric; this behavior of balls in the Riemannian metric poses however no problem for our Type II quantum theory as the volume of the past lightcone blows up linearly in $-t'$ for t' towards minus infinity in opposition to the volume of the future lightcone which blows up exponentially in $\left|\tilde{t} - t\right|$ and the $t < t' < \tilde{t}$ slice of the lightcone contains the intersection of the future lightcone with the $\left|\tilde{t} - t\right|$ ball which reaches above the $\tilde{t} - t - \frac{1}{\sqrt{\Lambda}}$ slice and therefore has a volume scaling as $e^{3\sqrt{\Lambda}(t'-t)}$ which indeed shows exponential scaling of the balls for late times t'.

All considerations in this chapter reveal that the hyperbolic behavior of the Wick rotation $h_{\mu\nu}$ of $g_{\mu\nu}$ has an effect on the quantum theory regarding the behavior at late times t. The latter impact is important and regards many fine details in the next chapter, but it is not insurmountable as we shall see and a Type III quantum theory can be defined on the cosmological vacuum. The reader understands by now that the entire analysis regarding finiteness of Feynman diagrams is going to rely upon the following property: a Riemannian geometry is called exponentially finite if and only if for any x, we have that

$$\int_{\mathcal{M}} P(d(x,y))e^{-\kappa d(x,y)}\sqrt{h(y)}dy < R(P,\kappa)$$

for any $\kappa > 0$, polynomial P and some $R(P,\kappa) > 0$. Here $R(P,\kappa)$ is supposed to go to zero in the limit for κ to plus infinity. Euclidean space time, the Wick rotation of Minkowski, is exponentially finite but the Wick rotated Friedmann cosmology is *not* so when considering the entire asymptotic future. It is however exponentially finite towards the geodesic region of every point x restricted to the sub-space time $t \leq \tilde{t}$ and we have worked our way towards this. In other words, the exponential blow up in the radius r for Riemannian balls $B(x,r)$ poses no problem when considering the intersection with the region contained within the (Lorentzian) geodesic horizon of x restricted to $t \leq \tilde{t}$ given that for large r, this intersection blows up *linearly* in r as opposed to the short scale r^4 behavior. It may be clear that we can nevertheless accomodate for the entire cosmological vacuum by means of the following notion: a Riemannian geometry is called exponentially finite on a scale $\zeta > 0$ if and only if for any x we have that

$$\int_{\mathcal{M}} P(d(x,y))e^{-\kappa d(x,y)}\sqrt{h(y)}dy < R(P,\kappa)$$

for any $\kappa > \zeta > 0$, polynomial P and some $R(P,\kappa) > 0$. This remark concludes this chapter.

Chapter 10

Perturbative finiteness.

In this chapter, we gather all our insights obtained so far and prove that the interacting theory is well defined at the perturbative level, meaning that every Feynman diagram is finite, and we aspire to obtain useful bounds. We proceed step by step and start by investigating the regularized propagators for spin-$\frac{1}{2}, 1, 2$ particles and obtain the required bounds on the propagator as well as on the spin-derivatives theirof with respect to the preferred $SO(3)$-class of vierbeins. From hereon, we eliminate all non-trivial structure of the interaction vertices so that we are left with ordinary integrals over space time of the function

$$e^{-\kappa d(x,y)}$$

which allows one to obtain several bounds on the respective Feynman diagrams. Our bounds on the propagators reveal that we have to work in space times such that the physical Wick rotation provides for an exponentially finite Riemannian geometry. It is with respect to that class that all our results pertain.

However, before coming to all that, it is somewhat amusing to quiet the mind of the impatient physicist who might have misunderstood quite some details of semi-classical gravitational physics and related to that, the so-called cosmological constant problem. Indeed, the suggested fascination of some physicists with this problem, amongst which 't Hooft, has always baffled me since the CC-problem is really no stranger than the infinite renormalizations occuring in ordinary quantum field theory. This is something 't Hooft can live with, probably because he got a Nobel prize for that piece of mathematical "art", but on the other hand, he "feels" that there is something deep behind the CC-issue probably requiring a deterministic quantum mechanics. So, as I told you, there are still those who aspire to become electron psychologists, an ambition which is correlated to the juvinile delusion that one can become a man "who knows everything". I will make it very clear now that in our setup, there is no cosmological constant problem and the reader may appreciate this at several levels: indeed, there is even no theory of semi-classical gravitation.

10.1 No CC-problem.

One might at this point reflect if one can still couple geometry semiclasically to our novel definition of a quantum theory and whether it is meaningful to do so. The main point of our discussion so far turned around the two point function and the Feynman propagator: any reference regarding quantum fields has been omitted; therefore, the point of view of an energy momentum tensor is not natural anymore. More in particular, the creation and annihilation processes at events x and y respectively came with a local energy momentum dependent "viscosity" so that the total process is not of Hamiltonian nature anymore and therefore no conserved currents, apart from the trivial one, can be constructed - something which is badly needed if one might want to look for a source for gravitation in the Einsteinian sense. This last fact implies that our framework does not contain a natural, nonzero, energy momentum tensor anymore; indeed, the only definition would be given by the following quadratic expression

$$\langle 0|T_{\mu\nu}(x)|0\rangle = \lim_{y\to x}\left(\partial_\mu\partial_{\nu'}W(x,y) - \frac{1}{2}g_{\mu\nu'}(x,y)\left(g^{\alpha\beta'}(x,y)\partial_\alpha\partial_{\beta'}W(x,y) - m^2W(x,y)\right)\right)$$

and we will now spawn some comments hereupon. In our *full* regularization scheme involving μ, κ, L we obtain that

$$\langle 0|T_{\mu\nu}(x)|0\rangle = 0$$

being an equality which is of course covariantly conserved. Therefore, in the approximation of no interactions, the state with no electrons present does not contribute as a source for the gravitational field which is obviously the only sensible answer. Again, in case interactions are included or nontrivial states are considered no such expression can be found in our theory. We will see now what happens in the μ, κ regularization scheme: in this case, the expression does not exist because the limit differs when y approaches x from the spacelike or timelike side. The fundamental reason herefore is to be found in the "reflection symmetry" in the surpression terms for spacelike geodesics, something which only depends upon an axis and not a magnitude, nor a specific orientation. We recall that this symmetry was needed to obtain bose statistics which crucially determined the definition of the Feynman propagator. Now, it may very well be that bose statistics is something which does not survive in a curved space time, but then the Feynman propagator would depend upon a frame of reference as there is no canonical way to define it. This is an avenue which we shall not take here; the reader, moreover, notices that the limit taken for y in the future lightcone of x gives an expression which is not covariantly conserved at all. This can be easily seen by noticing that for $y \in I^{\pm}(x)$ sufficiently close to x one has that

$$W_\mu(x,y) = \int \frac{d^4k}{(2\pi)^3}\delta(k^2 - m^2)\theta(k^0)e^{-ik^a\sigma_a(x,y)}e^{-\mu\left(K_{ab}(x)k^ak^b + K_{a'b'}(y)k^{a'}_{\star\,-\sigma^c(x,y)}k^{b'}_{\star\,-\sigma^c(x,y)}\right)}$$

where $\sigma(x,y)$ denotes as usual Synge's world function and the index a refers to the operation $e^\mu_a(x)\partial_\mu$ applied to it. The quadratic form $K_{ab}k^ak^b$ satisfies the property that it blows up quadratically in any Lorentz frame towards infinity if k^0 goes to infinity. In this limit $y \to x$, $W_\mu(x,x)$ becomes a smooth function of $K_{ab}(x)$ only since $\sigma_a(x,x) = 0$. The latter, however, does not satisfy a conservation law since generically $K_{ab;\nu}(x) \neq 0$ and the same reasoning applies to the whole energy momentum tensor where second covariant derivatives of $K_{ab}(x)$ come ito play and the expression becomes much more complicated. More abstract and from first principles, there is a-priori no good reason why the coincidence limit of derivatives applied to an amplitude for particle propagation should have something to do with a *vacuum* expectation value of some energy momentum tensor. The way geometry is influenced by quantum particles must therefore be encoded in a new theory which requires a super metric, a universal, and therefore background independent, metric on the space of all Lorentzian geometries (and matter configurations thereupon). This author has written ideas regarding this super-metric up in his Phd thesis.

10.2 Bounds on regularized spin-zero Feynman propagators.

This section will be brief and technical, but the underlying physical and mathematical ideas should be clear. We will regularize the (Feynman) propagator in such a way that all norms of covariant derivates of the latter, where the norm is defined with respect to the $SO(3)$-class of vierbeins, are bounded as $Ce^{-\kappa' d(x,y)}$ where C is some constant. This "universal" property, which is possible due to the nature of our regularization, is sufficient to arrive at a universal proof for finiteness of Feynman diagrams for any interacting theory with *any* kind of interaction vertices, at least in space times such that the Wick rotation is exponentially finite on some high scale. To summarize our results so far, we obtained for spin-0 particles that

$$\begin{aligned} W_{\mu,\kappa,L}(x,y) &= \int \frac{d^4k}{(2\pi)^3}\delta(k^2-m^2)\theta(k^0) \sum_{w:\exp_x(w)=y} e^{-\frac{2L^2}{w^a w_a}} e^{ik^a w_a} e^{-\kappa\int_0^1 \sqrt{h(w(s),w(s))}ds} \\ & \qquad e^{-\mu(V^\alpha k_\alpha)^2 - \mu(V_{\beta'}k^{\beta'}_{\star w})^2 + R(w)} \text{ symmetric if } w \text{ is spacelike.} \end{aligned}$$

on the condition that there exist open neighborhoods $\mathcal{V}_x$ and $\mathcal{W}_y$, as well as $\mathcal{O}_{x,w} \subset T^\star\mathcal{M}$ such that: (a) $\mathcal{O}_{x,w} \cap \mathcal{O}_{x,w'} = \emptyset$ (b) for any $x' \in \mathcal{V}_x, y' \in \mathcal{W}_y$ there is exactly one $w' \in \mathcal{O}_{x,w}$ such that $\exp_{x'}(w') = y'$ and

$$\phi_{\mu,\kappa,L}(x', k^{a'}, y') = \sum_{w':\exp_{x'}(w')=y'} \tilde{\phi}_{\mu,\kappa,L}(x', k^{a'}, w').$$

This regularization procedure does clearly not depend upon any local Lorentz frame and the integral is Lorentz invariant as it should. Before we come to the higher spin operators, whose analysis is only slightly more complicated than the one for spin-0 particles, let me indicate the bounds we are going to construct and are needed for the subsequent analysis. First, let us show under which generic conditions a bound of the type

$$|W_{\mu,\kappa,L}(x,y)| < C(\mu,\kappa,g,V,\epsilon)e^{-(\kappa-\epsilon)d(x,y)}$$

where C is a constant depending upon the geometry and μ, $0 < \epsilon \ll \kappa$ and d the Riemannian distance defined by h, can be constructed prior to showing that the same can be done for higher derivatives given that we smoothened the lightcone.

Sometimes, it happens that an infinite number of geodesics between two points exists on a space time with a trivial first homotopy group such as is the case for a closed Friedmann universe which has topology $S^3 \times \mathbb{R}$. When Wick rotating this space time, there exists a minimal length on the closed spacelike geodesics, which are also geodesics in the Wick rotated metric; albeit every such geodesic can be deformed to a point. Therefore, in this particular example, a closed geodesic that winds around n-times has at least (exactly) n-times this minimal length. This is the main feature we are interested in; suppose now that the closed geodesics are due to a nontrivial first homotopy group and that arbitrary winding numbers occur. Under rather generic conditions, we may associate to each homotopy generator a minimal length squared $M(h) > 0$ (Gromov) such that the energy of a curve with winding number $n > n_0 > 0$ between[1] x and y is greater than $d(x,y) + n\frac{M(h)}{d(x,y)+1}$ which is another expression of the fact that higher winding numbers come with a multiple of a fixed length. These considerations lead one to

$$|W_{\mu,\kappa,L}(x,y)| \leq e^{-\kappa d(x,y)} \sum_{w:\exp_x(w)=y} C_{\mu,\kappa}(x,w)e^{-\frac{2L^2}{w_a w^a}} e^{-\kappa n(w)\frac{M(h)}{d(x,y)+1}}$$

where $0 < C_{\mu,\kappa}(x,w) < C_{\mu,\kappa}$. Here, further estimates regarding

$$\sum_n e^{-\kappa \frac{nM(h)}{d(x,y)+1}} = \frac{1}{1 - e^{-\kappa\frac{M(h)}{d(x,y)+1}}} \leq C(\mu,\kappa,h)(d(x,y)+1)$$

can be made and the division through $d(x,y)+1$ stems from infinitely large homotopy classes and can be ignored when all nontrivial topology resides in a compact region of space time. These conditions are not always true on non-compact space times in case singularities are present, giving rise to topology change and $M(h) = 0$.

Therefore, under rather generic conditions, we obtain that

$$|W_{\mu,\kappa,L}(x,y)| < D(\mu,\kappa,h)(1+d(x,y))e^{-\kappa d(x,y)}$$

[1] The n_0 serves here to avoid the pathological cases where, for example, the length of two h-geodesics equals the minimal distance d and, moreover, they have a relative winding number of one.

which leads to a bound of the type

$$|W_{\mu,\kappa,L}(x,y)| < C(\mu,\kappa,h,\epsilon)e^{-(\kappa-\epsilon)d(x,y)}$$

for any $0 < \epsilon \ll \kappa$. Bounding the derivatives is a far more difficult task to perform given that in the above, the specific details of Synge's function or the energy functional didn't matter. Proving the assertions following below is a task in global analysis which has never been made before as far as I know. Therefore, the reader should take them as assumptions which are most likely true for our class of cosmological vacua; the formal proof of which constitutes a gap in our knowledge. More in particular, we shall assume that, given some $N \in \mathbb{N}$ there exists some $0 < \epsilon < \kappa$ such that

$$\sqrt{W_{\mu,\kappa,L;\beta_1,\ldots,\beta_i}(x,y)h^{\alpha_1\beta_1}\ldots h^{\alpha_i\beta_i}W_{\mu,\kappa,L;\alpha_1,\ldots,\alpha_i}(x,y)} < C_i(\mu,\kappa,L,g,h,\epsilon)e^{-(\kappa-\epsilon)d(x,y)}$$

for $i : 0\ldots N$ and α_k, β_k any index-pair referring to x or y respectively. The same assumption will hold regarding the Feynman propagator: I am unaware under what circumstances one can strengthen this assumption for an arbitrary number of derivatives and research of such fine points is left for the future.

10.3 Bounds on regularized spin-$\frac{1}{2}$ propagators.

It was shown in chapter six that the correct frictionless two point function for particles and anti-particles of spin-$\frac{1}{2}$ are given by

$$W_p(x,y)_i^{j'} = \int_{T^\star\mathcal{M}_x} \frac{d^4k}{(2\pi)^3}\delta(k^2-m^2)\theta(k^0) \sum_{w:\exp_x(w)=y} (\Lambda^{\frac{1}{2}}(x,w))_r^{j'}(k_a(\gamma^a)_i^r+m\delta_i^r)e^{ik^aw_a}$$

and

$$W_a(x,y)_{j'}^{i} = \int_{T^\star\mathcal{M}_x} \frac{d^4k}{(2\pi)^3}\delta(k^2-m^2)\theta(k^0) \sum_{w:\exp_x(w)=y} (k_a(\gamma^a)_r^i-m\delta_r^i)((\Lambda^{\frac{1}{2}}(x,w))^{-1})_{j'}^{r}e^{ik^aw_a}$$

where $\Lambda^{\frac{1}{2}}(x,w)$ is the spin transformation associated to parallel transport of a spinor along the geodesic between x and y determined by w. It has been shown that

$$W_p(x,y)_i^{j'} + W_a(y,x)_i^{j'} = 0$$

for $x \sim y$ using reflection symmetry so that spin-$\frac{1}{2}$ particles exhibit Fermi-statistics. We now intend to regularize this propagator in the same way as it occured for the spin-0 particle while still maintaining the Fermi property: again, this is a *condition* posed on the regularization scheme and by no means a proof. This relates to my previous comment that particle statistics relates to Minkowski space time and that there is no a-priori reason why particle statistics should hold in a curved space time with local geometric excitations. The viewpoint of the generalized Schrodinger equation suggests that

$$\widetilde{W}_{p,a}^{\mu,\kappa,L}(x,y) = \sum_{w:\exp_x(w)=y\text{ and w is causal.}} W_{p,a}^{\mu,\kappa,L}(x,w)+$$

$$\frac{1}{2}\sum_{w:\exp_x(w)=y \text{ and w is spacelike.}} \left(W^{\mu,\kappa,L}_{p,a}(x,w) - W^{\mu,\kappa,L}_{a,p}(y,-w_{\star w})\right)$$

a definition which differs in a slight but not unimportant way from the original one in my publications. There, we did not bother about smoothening the light-cone since no derivatives of Fermi-propagators are ever taken, which resulted in the alternative definition

$$\widetilde{W}^{\mu,\kappa,L}_{p,a}(x,y) = \frac{1}{2}\left(W^{\mu,\kappa,L}_{p,a}(x,y) - W^{\mu,\kappa,L}_{a,p}(y,x)\right)$$

if $x \sim y$ and $W^{\mu,\kappa,L}_{p,a}(x,y)$ otherwise. The difference regards the treatment in both prescriptions of the spacelike geodesics connecting x with y which remain spacelike once y crosses the boundary of $J^{\pm}(x)$. This difference has been taken away in the novel definition which allows for a correct smoothening procedure near the lightcone. Both definitions do not depend upon any local Lorentz frame and the propagator has the correct transformation properties under combined Lorentz and spin transformations. Again, the reader may infer that on a general class of backgrounds

$$\widetilde{W}^{\mu,\kappa,L}_{p}(x,y) = e^{-\kappa d(x,y)} \sum_{w:\exp_x(w)=y} \Lambda^{\frac{1}{2}}(x,w)\left(C^{\mu,\kappa,L}_{p;b}(x,w)\gamma^b + C^{\mu,\kappa,L}_{p}(x,w)1\right)$$

and

$$\widetilde{W}^{\mu,\kappa,L}_{a}(x,y) = e^{-\kappa d(x,y)} \sum_{w:\exp_x(w)=y} \left(C^{\mu,\kappa,L}_{a;b}(x,w)\gamma^b + C^{\mu,\kappa,L}_{a}(x,w)1\right)(\Lambda^{\frac{1}{2}}(x,w))^{-1}$$

where we have used that $\Lambda^{-\frac{1}{2}}(y,-w_{\star w}) = \Lambda^{\frac{1}{2}}(x,w)$. Just like in the previous case do we obtain from elementary considerations that

$$\sum_{w:\exp_x(w)=y} \left|C^{\mu,\kappa,L}_{\alpha;b}(x,w)\right| < C(\mu,\kappa,L,g,h)(d(x,y)+1)$$

and likewise for $\sum_{w:\exp_x(w)=y}\left|C^{\mu,\kappa,L}_{\alpha}(x,w)\right|$ where $\alpha \in \{a,p\}$. Here, the above inequalities are taken with respect to the preferred $SO(3)$-class of cosmological vierbeins which is evident given that $C^{\mu,\kappa,L}_{a;b}(x,w)$ behaves as a Lorentz vector and likewise do we need a norm estimate, with respect to the same vierbein, of the propagator. The relevant matrix norm is given by

$$||A|| = \left(\mathrm{Tr}(A^{\dagger}A)\right)^{\frac{1}{2}}$$

and an elementary computation yields

$$||W^{\mu,\kappa,L}_{p}(x,y)|| = e^{-\kappa d(x,y)}||\sum_{w:\exp_x(w)=y}(\Lambda^{\frac{1}{2}}(x,w))\left(C^{\mu,\kappa,L}_{p;b}(x,w)\gamma^b + C^{\mu,\kappa,L}_{p}(x,w)1\right)|| \leq$$

$$2e^{-\kappa d(x,y)}\sum_{w:\exp_x(w)=y}\sqrt{\mathrm{Tr}\left(\Lambda^{\frac{1}{2}}(x,w)^{\dagger}\Lambda^{\frac{1}{2}}(x,w)\right)}\sqrt{\sum_b\left|C^{\mu,\kappa,L}_{p;b}(x,w)\right|^2 + \left|C^{\mu,\kappa,L}_{p}(x,w)\right|^2}$$

which can be further bounded to

$$2e^{-\kappa d(x,y)}\left(\sup_{w:\exp_x(w)=y}\sqrt{\mathrm{Tr}\left(\Lambda^{\frac{1}{2}}(x,w)^{\dagger}\Lambda^{\frac{1}{2}}(x,w)\right)}\right)\left(\sum_{w:\exp_x(w)=y}\left(\sum_b\left|C^{\mu,\kappa,L}_{p;b}(x,w)\right|+\left|C^{\mu,\kappa,L}_{p}(x,w)\right|\right)\right).$$

Finally, this is majorated by

$$D(\mu,\kappa,g,V)e^{-\kappa d(x,y)}\left(\sup_{w:\exp_x(w)=y}\sqrt{\mathrm{Tr}\left(\Lambda^{\frac{1}{2}}(x,w)^{\dagger}\Lambda^{\frac{1}{2}}(x,w)\right)}\right)(1+d(x,y))$$

which puts us into the position to make the following global definition. We say that the tuple (g,V) defines a spin-transport which is $0<\delta$ exponentially finite if and only if

$$\sup_{w:\exp_x(w)=y}\sqrt{\mathrm{Tr}\left(\Lambda^{\frac{1}{2}}(x,w)^{\dagger}\Lambda^{\frac{1}{2}}(x,w)\right)}\leq F(g,V)e^{\delta d(x,y)}.$$

Obviously, we assume $\delta<\kappa$ so that

$$||W^{\mu,\kappa,L}_p(x,y)||<E(\mu,\kappa,g,V,\epsilon)e^{-(\kappa-\delta-\epsilon)d(x,y)}$$

for any $0<\epsilon\ll\kappa-\delta$ which finishes our discussion of the necessary bounds on the spin-$\frac{1}{2}$ propagator.

As mentioned previously, it is not costumary to develop theories in which derivatives of the Fermi-propagator are taken but if the reader wishes to, he or she may assume a similar bound to hold on the derivatives of the propagator. It would be worthwhile to study all these issues in far greater detail and try to say something about it from the point of view of global analysis. However, I am unaware of any such result and this book is just a temporary, organic reflection of what I know at this instant in time about this topic. As mentioned in the introduction, the reader is more than welcome to work on the remaining open issues which need further attention.

10.4 Spin-one and two propagators.

Qualitatively, all important details about regularized propagators have been revealed, but we shall nevertheless present the case of spin-$1,2$ given that those constitute the only propagators from which derivatives have to be taken in interacting theories such as non-abelian gauge theory and gravity. For a massless spin-one particle associated to a compact symmetry group, the propagator is given by

$$W^{\mu,\kappa,L}_{\gamma\nu';\alpha,\beta'}(x,y)=-g_{\alpha\beta'}\sum_{w:\exp_x(w)=y}g_{\gamma\nu'}(x,w)W_{\mu,\kappa,L}(x,w)$$

where

$$g_{\gamma\nu'}(x,w) = (\Lambda^{-1}(x,w))^{\mu}_{\nu'} g_{\gamma\mu}(x)$$

is the parallel transport of the bi-tensor along the geodesic determined by w. Here, $W_{\mu,\kappa,L}(x,w)$ constitutes the usual spin-0 expression and the reader remembers that in interactions this propagator is contracted with the vierbein $e_a(x)$. Therefore, with respect to our $SO(3)$-class of cosmological vierbeins, we demand that

$$||\Lambda(x,w)^{a'}_{b}|| \leq F(g,V)e^{\delta d(x,y)}$$

a condition of δ exponential finiteness. In the same way as before, we obtain that

$$||W^{\mu,\kappa,L;a'}_{b;\alpha\beta'}(x,y)|| \leq |g_{\alpha\beta'}|\, E(\mu,\kappa,L,g,V,\epsilon)e^{-(\kappa-\delta-\epsilon)d(x,y)}$$

for all $0 < \epsilon \ll \kappa - \delta$. We shall assume similar bounds to hold on the first N spin derivatives of this expression so that all tricky aspects of non-abelian gauge theory are covered for.

We haven't said too much about spin two particles yet and we shall make up this deficit to some extend here; from symmetry considerations, one can derive that the two point function is given by

$$W^{\mu,\kappa,L}(x,y)_{\alpha\beta,\alpha'\beta'} =$$

$$\sum_{w:\exp_x(w)=y} \left(g_{\alpha\alpha'}(x,w)g_{\beta\beta'}(x,w) + g_{\beta\alpha'}(x,w)g_{\alpha\beta'}(x,w) - \frac{1}{2}g_{\alpha\beta}(x)g_{\alpha'\beta'}(y) \right) W_{\mu,\kappa,L}(x,w).$$

The factor $\frac{1}{2}$ has been chosen such that

$$W^{\mu,\kappa,L}(x,y)_{\alpha\beta,\alpha'\beta'}g^{\alpha\beta}(x) = W^{\mu,\kappa,L}(x,y)_{\alpha\beta,\alpha'\beta'}g^{\alpha'\beta'}(y) = 0$$

thereby eliminating all trace degrees of freedom. As before, we evaluate this propagator with respect to our $SO(3)$-class of vierbeins leading to the tensor

$$W^{\mu,\kappa,L}(x,y)_{ab,a'b'} = \sum_{w:\exp_x(w)=y} \left(\Lambda(x,w)_{a'a}\Lambda(x,w)_{b'b} + \Lambda(x,w)_{a'b}\Lambda(x,w)_{b'a} - \frac{1}{2}\eta_{ab}\eta_{a'b'} \right) W^{\mu,\kappa,L}(x,w)$$

where a norm estimate of the "coefficients"

$$C(x,w)^{ba'}_{ab'} = \Lambda(x,w)^{a'}_{a}\Lambda(x,w)^{b}_{b'} + \Lambda(x,w)^{a'b}\Lambda(x,w)_{b'a} - \frac{1}{2}\delta^{b}_{a}\delta^{a'}_{b'}$$

can be made due to the δ-exponentially finite character of the Lorentz transporters. In general, we therefore have that

$$||W^{\mu,\kappa,L}(x,y)^{ba'}_{ab'}|| \leq E(\kappa,\mu,L,g,V,\epsilon)e^{-(\kappa-\epsilon-2\delta)d(x,y)}$$

where the trace-norm has been taken with respect to the usual $SO(3)$-class of vierbeins. Now, we shall assume that a similar bound holds for the first N

spin-derivatives of the two point function, a condition which should again be further examined.

This finishes our treatment of propagators and quite evidently, we assume all the above to be valid for the Feynman propagator even if the presentation has been made for the two point function. This is quite legitimate as no essential points in the above analysis change by doing so; we are all set and ready now to prove perturbative finiteness under the geometrical constraints mentioned so far in this and previous chapters. I deem these constraints to be quite mild and again, they should be further investigated in the future.

10.5 Bounds on Feynman diagrams.

In this section, we will proceed in a few steps by treating first the so-called ϕ^4 theory and give two different types of bounds on the Feynman diagrams. The first one is specific to the theory while the second one is of a more universal nature: we shall illustrate that point of view by reducing the calculations for non-abelian gauge theory to the ones of ϕ^4 theory. Next, we move to the theory of massless spin two particles or gravitons, which can also be reduced in this way, but contains a few details relating to nonperturbative aspects which are not present in the latter theories. In particular, the friction parameter μ shall be related to the Planck length squared l_p^2; everything which will be said below pertains to Lorentzian geometries g having an exponentially finite Wick tranform h at some scale $\zeta > 0$.

Let us start by mentioning an obvious equation for general Feynman diagrams, interpreted as graphs, which is that

$$V - I = C - L$$

where V is the total number of *internal* vertices of a Feynman diagram, I its number of internal lines, hereby excluding the legs towards the external points, and L is the number of loops. Finally, C is the number of components of the graph; for ϕ^4 theory and connected diagrams C is bounded by

$$C \leq \frac{n+m}{2}$$

where n, m are the number of IN and OUT vertices respectively. With these conventions, we have that the absolute value of every Feynman diagram is bounded by

$$c(m,\mu)^{I + \frac{n+m+n'+m'}{2}} \int dz_1 \sqrt{h(z_1)} \ldots \int dz_V \sqrt{h(z_V)} \prod_{\text{all lines } (\alpha_i, \alpha_j)} e^{-\kappa d(\alpha_i, \alpha_j)}$$

where we have used that the spin-0 Feynman propagator is bounded by $c(m,\mu)e^{-\kappa d(x,y)}$ and $\alpha_i \in \{z_k, x_i, y_j\}$. Moreover, for ϕ^4 theory, one has that

$$I + \frac{n' + m'}{2} = 2V$$

where $0 \leq n' \leq n$ and $0 \leq m' \leq m$ so that the prefactor may be exactly written as

$$c(m,\mu)^{2V + \frac{n+m}{2}}$$

and

$$2V - \frac{n' + m'}{2} \geq L = V + C - \frac{n' + m'}{2} \geq 0$$

so that $V \geq \frac{n'+m'}{4}$. Here n', m' denote the number of IN or OUT vertices which are connected to an internal vertex. Before we proceed, let us mention some easy to see fact about the Friedmann cosmology; if z is within the geodesic horizon of x and y, then it is in the geodesic horizon of the midpoint of x and y in the Riemannian metric[2]. This observation is most convenient in the following estimates which constitute a straightforward generalization of our previous inequalities. We start by deducing a universal and simple bound which does not depend at all on the details of the interaction vertices as well as on the distances between the exterior vertices. It is simply given by

$$c(m,\mu)^{2V + \frac{n+m}{2}} R(1,\kappa)^V$$

which is most easily proved by induction on the number of internal vertices V. Here,

$$\int_{\mathcal{M}} d^4y \, \sqrt{h(y)} e^{-\kappa d(x,y)} < R(1,\kappa)$$

as defined at the end of chapter nine. If $V = 0$, then the bound is easily seen to hold since $e^{-\kappa d(\alpha,\beta)} \leq 1$ for every leg joining two external vertices. Suppose now that the bound is true for $V \geq 0$, we will prove it for $V + 1$. Take any internal vertex connected by one edge to an exterior vertex α and remove it; the effect is that we obtain a diagram with four extra external vertices (we copied four times the internal vertex) but with one internal vertex less. Remove the edge to α from the diagram, then the remaining part is bounded by

$$R(1,\kappa)^{V-1}.$$

Now there remains to identify the four vertices again and perform the remaining integration over this vertex; the latter gives an extra factor of $R(1,\kappa)$ because we still have the leg to α which proves the result. This shows that the diagram blows up in a suitable way, but there remains of course the "entropy" factor associated to all Feynman diagrams with V internal vertices and n IN and m OUT vertices. The latter remains to be investigated in the next chapter but it

[2]This follows most easily from the convexity of the horizon of z in the Riemannian metric d which the reader may prove as an exercise.

is very well possible that unitarity may have to be given up to make the series analytic.

This is by far the easiest proof that the Feynman diagrams are finite; in case of ϕ^4 theory, it is possible to make another, useful, estimate in case the geometry is spherical; in either, we assume that our Riemannian manifold h satisfies a volume bound for a ball of radius r around x by

$$\mathrm{Vol}_4(B(x,r)) \leq Kr^4$$

for some metric dependent constant K. This includes a Type II theory for the cosmological vacuum as mentioned previously. Consider n points z_i and take the integral

$$\int_{\mathcal{M}} dz \sqrt{h(z)}\, e^{-\kappa \sum_{i=1}^{n} d(z_i,z)}$$

then, as previous, this may be bouded by

$$e^{-\frac{\kappa}{n-1}\sum_{i<j} d(z_i,z_j)}$$

$$\left(\int_{y;\,\exists z_i,z_j : d(y,\frac{z_i+z_j}{2}) < \frac{3}{2} d(z_i,z_j)} \sqrt{h(y)} dy \,+\, \int_{y;\,\forall i,j\; d(y,\frac{z_i+z_j}{2}) \geq \frac{3}{2} d(z_i,z_j)} dy\, e^{-\frac{\kappa}{2(n-1)}\left(\sum_{i<j} d(y,\frac{z_i+z_j}{2})\right)} \sqrt{h(y)} \right).$$

Note here the factor of 2 in the denominator of the exponential in second integral; this originates from the fact that in a general Riemannian space

$$d(x,y) + d(y,z) \geq d(x,z) + \frac{1}{2} d(y, \frac{x+z}{2})$$

for $d(y,\frac{x+z}{2}) \geq \frac{3}{2} d(x,z)$ whereas in Euclidean space this factor $\frac{1}{2}$ is not present. The latter formula can again be estimated by

$$e^{-\frac{\kappa}{n-1}\sum_{i<j} d(z_i,z_j)} \left(\left(\frac{3}{2}\right)^4 K \sum_{i<j} d(z_i,z_j)^4 + R\left(1, \frac{\kappa}{2(n-1)}\right) \right)$$

and the only thing the reader should notice is the division of κ through $n-1$ which lowers convergence for diagrams with multiple internal vertices. We will not apply the above estimate consistently but look for a finer estimate which will provide one with better convergence properties. Actually, we will be set with a Kirchoff diagram where the flow is given by some rational proportion of $\kappa d(x_i,z_j)$ or $\kappa d(y_j,z_k)$; at any instant of the computation, these proportions add up to one. The optimal way of spreading around is by ensuring that the you do not subdivide into smaller portions; in that way, the surpression factor at the vertex remains constant κ. Homogeneous fractalizing is the worst that can happen since it lowers κ substantially after a few vertices have been run through. Loops make no difference whatsoever, in case we have a loop and there are three external vertices, two with current κ and one with current 2κ then we obtain that κ does not get renormalized, nor at the vertex nor at the legs. Also, in case we have a loop with only two external points each with current κ, there is no lowering of κ neither at the vertex nor at the legs.

Let us reason why homogeneous spreading is a bad idea; in case any of the currents associated to a leg consists out of several pieces, then a lowering of κ will occur, but such lowering will always be less than is the case for a vertex with four external currents associated to four distinct graph points. We will now determine the maximal contribution of homogeneous fractalizing: start at any vertex z_i, then the most severe contribution regarding the integral comes when no loop is present and likewise, this situation divides κ through the largest number three. Pick now any neighboring vertex, then again, the largest division occurs again when there are three other external legs, dividing the $\frac{1}{3}$ leg into 3 times $\frac{1}{9}$ and the remaining $\frac{2}{3}$ per other leg by two which gives $\frac{1}{3}$ and yields the surpression factor of $\frac{\kappa}{6}$ on the second vertex. In the third step, the worst that can happen is that a leg of the first and second vertex meet since that would cause maximal diversification. The leg from the first vertex contains two factors $\frac{1}{3}$ and 3 factors $\frac{1}{9}$ and the same for the leg coming from the second vertex. Therefore, diversification would lead to 4 times $\frac{1}{6}$ and 6 times $\frac{1}{18}$ on the other two legs, giving a surpression of $\frac{\kappa}{12}$ at the third vertex. Clearly, this reasoning is catastrophic and we now turn our head towards no fractalizing.

This case is easy and one can partition the set $S = \{x_i, y_j\}$ into pairs $(\alpha_{2i-1}, \alpha_{2i})$; with these reservations, the quantitative result reads

$$c(m,\mu)^{2V+\frac{n+m}{2}} P(d(\alpha_{2i-1},\alpha_{2i})\,; i = 1 \ldots \frac{n'+m'}{2}) e^{-\kappa \sum_{i=1}^{\frac{n+m}{2}} d(\alpha_{2i-1},\alpha_{2i})}$$

where P is a polynomial of degree $4V$ and the highest order coefficient is bounded by

$$\left(\frac{3}{2}\right)^{4V} K^V \left(2^4\right)^{\frac{V(V-1)}{2}} .$$

It is the behavior of this last coefficient which makes our bound on the series nonanalytic. The above formula is always true for any diagram as the reader may wish to show by induction on the number of internal vertices, by integrating out a vertex without altering the connectivity properties[3], and does not hinge upon special features of the diagram such as the property that there exists a partition of the edges into paths, connecting the exterior points, and loops such that no internal vertex belongs to two loops. It is always possible to cover a graph by means of curves connecting the exterior points and loops

[3]Such a vertex always exists as the following reasoning shows: start at an exterior vertex and go in the diagram by means of the edge e. On the first vertex v one meets, there is another edge f which can be connected to a different exterior vertex without coming back to v given that every vertex is connected to at least two different exterior vertices. If the other two edges of v are identified and therefore form a loop, then connect e with f and integrate out v. Otherwise connect e and f with one of the remaining edges each and integrate out v, which preserves the connectivity properties of the diagram.

but sometimes it is the case that two loops always intersect[4]. The reader might wonder wether the above estimate is not too crude given that we do not rely upon the details of $W_\mu(x,x')$ at all. Also, we replaced the Lorentzian geodesic energy by the inferior Riemannian distance, wich is an approximation as well. My answer is a resounding no: these approximations will not significantly influence the result for the following reasons. Regarding W_μ, only very slight falloff behavior towards infinity can be shown which effectively can be minorized by means of a slight renormalization of κ (increasing its value a bit). Concerning the replacement of the energy term by the Riemannian distance; not much is to be expected here since they coincide in Minkowski given that the geodesics of both metrics are the same. Therefore, in a general analysis, these details should not matter.

The reader notices that both bounds have their advantages but that the first one is universal in nature and did not depend at all upon the four-valency of the interaction vertex. We shall now turn our head towards the perturbative renormalizability of non-abelian gauge theory; the proof of which reduces fully to the one above. The proof is almost self-evident given that every Feynman diagram consists out nothing but a product of spin-one propagators and at most second derivatives, one in each end vertex, thereof as well as Fermi and ghost propagators. The intertwiners $f_{\alpha\beta\gamma}, g_{\alpha\beta}$ and $(\gamma^a)^i_j, \eta_{ab}$ are all uniformly bounded so that the total Feynman diagram reduces to V-integrals of exponential factors

$$e^{-(\kappa-\delta-\epsilon)d(x,y)}$$

associated to all, up to second order, derivatives of *any* propagator. A fully analogous reasoning as before then shows that the contribution of any Feynman diagram is bounded by

$$C(D)E(\mu,\kappa,L,g,V,\epsilon)^{I+\frac{n+m+n'+m'}{2}}R(1,\kappa-\delta-\epsilon)^V$$

where $C(D)$ is a factor associated to the specific diagram D and function of the relevant intertwiners; all further symbols have the same meaning as before. I want to stress again that this bound holds in our special $SO(3)$-class of reference frames and that local boosts at the end vertices can make this number as large as one wants to.

We now finish this section by further fleshing out the graviton theory, at least at the perturbative level; comments regarding non-perturbative aspects will follow. At the perturbative level, we will need supplementary bounds on the Riemann tensor of g such as

$$R_{abcd}(x)\delta^{aa'}\delta^{bb'}\delta^{cc'}\delta^{dd'}R_{a'b'c'd'}(x) < C$$

where the Lorentz indices are taken with respect our special $SO(3)$-class of tetrads. This implies that all interaction intertwiners $Z(x)$ are uniformly bounded

[4]The reader may easily find an example of such diagram.

in these Lorentz frames and therefore, the contribution of any Feynman diagram is estimated by

$$\prod_{i=1}^{V} C(Z_i)\,(C(\kappa,\mu,L,g,V,\epsilon))^E \int dz_1 \sqrt{h(z_1)} \ldots \int dz_V \sqrt{h(z_V)} \prod_{\text{edges }(\alpha,\beta)} e^{-(\kappa-\epsilon-2\delta)d(\alpha,\beta)}$$

where $C(Z_i)$ is a constant depending upon the intertwiners Z_i, V is the number of internal vertices, E the number of edges (internal and external) and α, β are the coordinates of an internal or external vertex respectively. The bound on the propagators is valid up till the fourth covariant derivatives of the graviton propagator, with maximally two covariant derivatives per vertex each, and the same for the ghost propagator

$$\Delta^{\mu,\kappa,L}_{F;\nu\nu'}(x,y) = - \sum_{w:\exp_x(w)=y \text{ and w is future causal or spacelike at x}} \theta(x)\overline{\theta(y)} g_{\nu\nu'}(x,w) W_{\mu,\kappa,L}(x,w) +$$

$$\sum_{w:\exp_x(w)=y \text{ and w is past causal at x}} \theta(x)\overline{\theta(y)} g_{\nu'\nu}(y,-w_{\star w}) W_{\mu,\kappa,L}(y,-w_{\star w}).$$

As we will show in the next chapter, there are some peculiar nonperturbative aspects of the graviton theory which are not present in any other interaction theory considered so far. We will show that the friction parameter μ has a bound depending upon the Planck length squared putting therefore a lower bound on the "nonlocal range" of a creation or annihilation process.

10.6 Comments.

The reader might probably wonder if a replica of the second bound on ϕ^4 theory is possible for the other theories; the answer is *no* as he may verify for himself in the case of quantum electrodynamics. Therefore, what we have done represents a kind of universal optimum and the reader should cherish the cheer simplicity of the proof enabled by the Riemannian nature of the problem. The simplicity as well as universality of the obtained results are in sharp contrast to the poverty displayed by means of the renormalization procedure: this mathematical monster has only caused the illusion that there was something deep hiding behind it and has diverted attention from the simple fact that the principle of general covariance would lead one to a class of *natural* solutions. This is the real lesson of this book that when people like Kallen, Weinberg, 't Hooft and many others are shouting that mathematical rigor has to be abandoned permanently, that one *must* logically think that these grandmasters of illegitimate manipulations are missing an essential part of physics.

Since I am somewhat in a political mood at this moment, let me stress that there is a point where a generation has to go on official retirement; the healthy age at which this occurs is when a professor becomes 65. By then, he might wish to follow Wolfgang Pauli with his fascination for the number 137 which indicates he has lost all contact with the reality of the activity of science. He probably also has a hidden agenda to ressurect a deterministic quantum mechanics, a fight he has since long lost with his collegues but which he hopes to win by appealing to rather silly politicians, mediocre minds and undereducated Royalty. Also, he is sceptical about paranormal events and maintains totally ludricrous opinions about what would count for him as a proof of paranormal activity: perhaps he is paid by the pharmaceutical industry or left wing politicians to feed those who experience paranormal phenomena with all kinds of drugs and anti-psychotic "medication". For him, it is time to withdraw before he goes into history as the *worst* theoretical physicist ever. Long live Heisenberg and Einstein, one a practical "follower" of the Nazi regime and another one hounted by the Aryans; no politician ever has the right to decide about nature since She will always be victorious. Also, scientists come into different political colours, but that shouldn't matter given that science is the only language on earth which has the potential to unite the minds of all people while music can do the same for all hearts. If cultural differences should once not matter anymore in interpersonal contacts, it is because we are speaking the language of Nature to some extend. I doubt wether lousy democrats will ever be able to follow that beat instead of adhering to some artificial, surpressing ideology made for mediocrity. Let me also stress that modern democracy is only 70 years old and if she continues like that, it might be unlikely that she makes it to a century.

Chapter 11

Entropy and analycity.

From the previous chapter, we learned that for any Feynman diagram, under the assumptions made, its value can be estimated by

$$C^V D^E R(1, \kappa')^V$$

where C is a constant arising from the intertwiners and can be uniformly bounded if the theory contains a finite number of them, whereas D is a constant coming from the propagators and $R(1, \kappa')$ has been defined before. This reveals that, in order to get a grasp on the entire perturbation series, we have to investigate the number of connected Feynman diagrams of the type (n, m, V, E) where every symbol has the same meaning as in the previous chapter. More precisely, we shall investigate the sum

$$A(n, m, V, E) = \sum_{D \text{ of type } (n,m,V,E)} \frac{1}{s(D)}$$

and this task is the easiest to perform for ϕ^4 theory while by far the most difficult for the graviton theory which contains an infinite number of interaction vertices; therefore, in this chapter, we shall proceed by making the exercise for ϕ^4 theory, next for non-abelian gauge theory and finally for the graviton theory. Insights regarding "unitarity" are obtained and further discussed.

11.1 Non-perturbative aspects of ϕ^4 theory.

The reason why I have postponed this issue in my papers towards this book project is because it is rather obvious to perform, at least for ϕ^4 and non-abelian gauge theory. The conclusions we shall reach are, on the other hand, interesting and this fact motivates the existence of this chapter. Let me stress from the outset that no approach to relativistic quantum theory has even reached the stage where one is in a position to adress these questions; our bound on the contribution of one Feynman diagram to the series is however sufficient for us to adress the issue. We shall show now that an upper bound for $A(n,m,V,E)$ is given by

$$A(n,m,V,E=2V+\frac{n+m}{2}) \le \sum_{0\le n'\le n;0\le m'\le m:n-n'=m-m',n'+m'>0} \frac{(4V)!(n-n')!}{(2V-\frac{n'+m'}{2})!(4!)^V 2^{2V-\frac{n'+m'}{2}}}$$

where n', m' have the same meaning as before. Given that any exterior vertex under consideration is connected to an interior vertex by means of an edge E, there are $4V(4V-1)\ldots(4V-n'-m'+1)$ possible choices whereas the remaining $4V-n'-m'$ lines emanating from the V internal vertices have to be identified internally. This leads to a factor

$$\frac{(4V-n'-m')!}{2^I I!}$$

where I is the number of internal lines and we know from previous considerations that $2I+n'+m'=4V$. Clearly the $I!$ stands for the number of permutations of the internal lines whereas the factor 2^I is associated to the swapping of orientation of them. Finally, the $(4!)^V$ in the denominator stems from the permutation freedom of the edges at the vertices and the $(n-n')!$ denotes the number of propagators between the remaining $n-n'=m-m'$ IN and OUT vertices. This upper bound is pretty thight and clearly provides one with the right kind of asymptotics in terms of V; also, I believe it would be hard to obtain a better one given that symmetry properties of individual diagrams would become important.

We shall now estimate its asymptotic behavior for large V and hence large $E=2V+\frac{n+m}{2}$ keeping n,m fixed. Clearly $n'+m'$ can be ignored when it comes together with V so that

$$A(n,m,V) \le \frac{(4V)!}{(2V-\frac{n+m}{2})!(4!)^V 2^{2V}} 2^{\frac{n-m}{2}} \left(\sum_{0\le n'\le n;0\le m'\le m:n-n'=m-m',n'+m'>0} (n-n')! 2^{n'} \right).$$

Therefore, our scattering amplitudes are bounded by

$$|\langle \mathrm{OUT}\, m|\mathrm{IN}\, n\rangle| \le 2^{\frac{n-m}{2}} D^{\frac{n+m}{2}} \left(\sum_{0\le n'\le n;0\le m'\le m:n-n'=m-m',n'+m'>0} (n-n')! 2^{n'} \right)$$

$$\sum_{V=0}^{\infty} \frac{|\lambda|^V (4V)!}{(2V-\frac{n+m}{2})!(4!)^V 2^{2V}} (CD^2 R(1,\kappa'))^V$$

and the right hand side is easily seen to diverge for any λ. This brings me back to comments I have previously made in my papers as well as the introduction which boil down to the fact that "unitarity" or the structure of the coefficients

$$\frac{(-i\lambda)^V}{s(D)}$$

will have to be changed for diagrams with a large number of internal vertices and we have just shown that this needs to be the case. We had of course anticipated already in chapter four, on general covariance, that unitarity was incompatible with it, but now we are forced to investigate deeper implications of this fact.

Unfortunately, I have at the moment no obvious substitute for the principle of "unitarity" which had no natural place in our theory anyway. It is a remnant from the old quantum theory on flat space time which leads to all kinds of inconsistencies mentioned previously in this book. The reader must again understand that this is not a weakness in my viewpoint but a liberty which is enforced upon the theory by means of our broader perspective on microscopic physics. The latter turned out to be necessary to tame the divergencies in the Feynman diagrams and to make the theory well defined; the principle of general covariance had similar implications for gravitational physics and so does it have for quantum theory. Therefore, I am *not* going to propose any specific coefficients which might make ϕ^4 theory well defined nonperturbatively but which could fail miserably for the graviton theory. Only experiment should guide us herein; the freedom associated with those coefficients should not be mistaken with a choice of an infinite number of "coupling constants". Indeed, in practise, only the first few terms of the perturbation series, regarding diagrams with a low number of internal vertices, are important and the rest can be safely ignored which tells you something about the *effect* of the remaining coefficients. To my feeling, this is as thight as the jacket can reasonably be and going over to higher values of the coupling constant opens up an infinite new world which remains unseen in ours.

11.2 Non-perturbative aspects of non-abelian gauge theory.

A similar qualitative result as the one just obtained for ϕ^4 is expected to hold in non-abelian gauge theory albeit the counting is somewhat more difficult since one disposes of four types of interaction vertices: a tri and four valent gauge boson vertex, a trivalent ghost-gauge boson vertex as well as a trivalent particle-gauge boson vertex. Since all details of the interaction vertices are washed out in the constant C and likewise so for all details of the propagators regarding the constant D, we are left with diagrams having tri and four valent vertices as well as a consistent labelling with p, a, b where p stands for particle, a for antiparticle and b for gauge boson, on the edges adjacent to the exterior vertices. This labelling should be extendible, in a non-unique way, to the interior edges when supplemented with a ghost and anti-ghost g, ag label. Therefore, we are interested in estimating amplitudes written down abstractly as

$$\langle \alpha_i, i = 1 \ldots n | \beta_j, j = 1 \ldots m \rangle$$

where $\alpha_i, \beta_j \in \{p, a, b\}$ and the ordering in the states is from one to n and one to m respectively. It is clear that we cannot provide for the exact number of labellings since that depends from graph to graph in the sum and we shall therefore provide for a reasonable upper bound for a graph with V_4 four-valent vertices and V_3 three-valent ones, ignoring the number of edges as well as n, m. An obvious upper bound is given by 7^{V_3} given that there is only one type of four-valent vertex with identical particle lines and every tri-valent vertex, together with isolation of the v, Ψ-line, fixes all other lines. There are in general three types of trivalent vertices and three lines per vertex to place the v, Ψ, hence 7^{V_3}; of course, this constitutes an overestimation of the state of affairs given that a tri-valent vertex connected to a four-valent one only leaves for two possibilities but all such details are graph dependent.

The estimates we are interested in here concern

$$D^{\frac{n+m}{2}} \sum_{0 \leq n' \leq n, 0 \leq m' \leq m, n'+m' > 0; V_3, V_4 \geq 0; 2I + n' + m' = 3V_3 + 4V_4} A(n', n, m', m, V_3, V_4) 7^{V_3} |\tilde{g}|^{2V_4 + V_3}$$

$$C^{V_3 + V_4} D^{\frac{3}{2}V_3 + 2V_4} R(1, \kappa')^{V_3 + V_4}$$

where $A(n', n, m', m, V_3, V_4)$ equals the number of diagrams with V_3 and V_4 trivalent, respectively four valent, vertices and all other symbols have the same meaning as before. $\tilde{g}$ is the coupling constant of the theory in the standard representation, see chapter eight. The reader understands that this is a fairly substantial overestimation of the state of affairs given that we do not take the nature of the exterior vertices into account in the determination of $A(n', n, m', m, V_3, V_4)$ implying that a particle can be connected to a four-valent vertex. However, one would expect the "real" number to be of the same magnitude which means we probably capture the right asymptotics in terms of V_3, V_4 and this is our only point of concern. $A(n', n', m', m', V_3, V_4)$ can again be estimated by

$$A(n', n', m', m', V_3, V_4) \leq \frac{(3V_3 + 4V_4)!}{2^I I! (3!)^{V_3} (4!)^{V_4}}$$

where $2I + n' + m' = 3V_3 + 4V_4$; hence,

$$A(n', n, m', m, V_3, V_4) \leq \frac{(3V_3 + 4V_4)!(n - n')!}{2^{\frac{3}{2}V_3 + 2V_4 - \frac{n'+m'}{2}} (\frac{3}{2}V_3 + 2V_4 - \frac{n'+m'}{2})! (3!)^{V_3} (4!)^{V_4}}.$$

This reduces our original sum to

$$D^{\frac{n+m}{2}} \sum_{V_3, V_4 \geq 0} 7^{V_3} |\tilde{g}|^{2V_4 + V_3} C^{V_3 + V_4} D^{\frac{3}{2}V_3 + 2V_4} R(1, \kappa')^{V_3 + V_4} \frac{(3V_3 + 4V_4)!}{2^{\frac{3}{2}V_3 + 2V_4} (3!)^{V_3} (4!)^{V_4}}$$

$$\sum_{0 \leq n' \leq n, 0 \leq m' \leq m, n'+m' > 0, n - n' = m - m'} \frac{(n - n')! 2^{\frac{n'+m'}{2}}}{(\frac{3}{2}V_3 + 2V_4 - \frac{n'+m'}{2})!}$$

which diverges again due to the superexponential factor

$$\frac{(3V_3 + 4V_4)!}{(\frac{3}{2}V_3 + 2V_4 - \frac{n+m}{2})!}.$$

Therefore, we reach again the conclusion that unitarity cannot hold for a quantal gauge theory to be well defined; in particular, diagrams with a high number of internal vertices need to be superexponentially surpressed in these parameters.

11.3 Gravitons.

Until now, we have received the lesson that diagrams with a large number of internal vertices should be super-exponentially surpressed; in a graviton theory we anticipate another lesson which is that diagrams with large vertices should be super-exponentially surpressed too. More in particular, one meets interaction vertices with coefficient l_p^{2n} having $2n$ legs such that one obviously needs a factor $a(n)$ such that $a(n)(2n)! \to 1$ in the limit for n to infinity. This would lead to a bound of the kind

$$1 > l_p^2 \frac{C(g, V, \kappa, \epsilon)}{\mu} ||W(x, y)_{ab'}^{ba'}|| > 0$$

given that a Feynman diagram contibutes the n'th power of that, which implies that

$$\mu > l_p^2 \alpha(g, V, \kappa, \epsilon)$$

is the kind of bound on the friction term μ one should anticipate in a graviton theory on a generic background. This is all I have to say about this for now, these results require deep reflection as they destroy the traditional structure of quantum field theory.

11.4 Conclusions.

We already knew that in traditional quantum field theory, the value of a Feynman diagram was not uniquely defined and moreover, that any regularization scheme is rather ad-hoc and lacks physical motivation. We rectified that by looking for modified propagators falling in the class delineated by our physical principles; in this chapter we moreover figured out that the traditional expansion series does not converge either requiring equally drastic modifications to the theory. The level of precision obtained in this book is unparalled in the literature and should constitute enough motivation for the reader to further investigate these matters. These notes conclude the main body of this book, the remaining chapter being merely an exposition about some thoughts of mine of how to formulate a physical principle giving rise to a free quantum theory for the background metric field. The particular proposal I will suggest properly reinstates "time" and strongly criticizes and departs from the timeless physics of Einstein. In particular, we will work again towards a generalized Fourier transformation and use this to define a free theory; interactions between different universes shall not be discussed.

Chapter 12

Quantum gravity type one.

The search for a theory of quantum gravity is one of new principles of nature and involves the question if and how the superposition principle should be applied to space time itself. Crucial in our story so far was the presence of a classical space time metric and therefore, a quantum theory of the space time metric appears to call for a super metric: a metric on the space of all Lorentzian geometries. Those, who keep on insisting upon a Feynman integration theory are facing the question of the canonical character of the "measure" where the latter has to be understood in some limiting, rather than a fundamental, sense since the space of all space times is not locally compact in any known Hausdorff topology. This is not the only worry one has regarding such discrete constructions: one has also to show that the limiting kinematical configurations are arbitrarily close to any classical space time in a suitable sense implying that the action principle at hand converges too. There is a very important distinction here between gravity and all other action principles in field theory, which is that the latter all depend upon first derivatives only while the former depends upon second derivatives of the metric field. There exists a discretization procedure invented by Regge, which can account for the second derivatives in a distributional sense but it requires flexibility in the degrees of freedom of the discrete structure (a simplicial manifold) so that, locally, on the $n-2$ simplices, where n is the dimension of the simplicial manifold, the deficiency angles go to zero sufficiently fast. The "curvature" of the approximating simplicial manifolds then converges to the Ricci scalar in a weak distributional sense. I am unaware of any suitable substitute for the Ricci tensor and Riemann curvature in this kinematical framework. I am also unaware of any approach to quantum gravity which manages to offer a suitable answer to these elementary matters of principle: the measure in the causal dynamical triangulations approach heavily depends upon the kinematical restrictions which, moreover, do *not* approximate any classical space time. Indeed, not only is it clear that Regge's scheme does not apply, the "local" curvature is a diverging quantity in the distributional sense when the continuum limit is taken.

What I have described above can be called "quantum gravity type one" where there is no classical metric background on which computations are performed. One can of course maintain that the universe consists also out of classical degrees of freedom providing one with a dynamical classical background on which it is possible to regard the quantization of the gravitational force as the quantum theory of the graviton. This can be called "quantum gravity type two"; such a theory has long been believed to be impossible due to the non-renormalizability of the gravitational force on a Minkowski background. It his here that our novel nonunitary quantum theory offers a way out since the theory is finite, a result which does not depend at all on the structure of the Feynman diagrams as has been shown here. In particular, loops played no special role at all in our analysis and were treated on pair with other internal legs which shows that quantum gravity type two is a perfectly safe theory in our framework.

12.1 Quantum gravity type one.

Personally, I have never made a choice between both types, both reflect different world views, which in my opinion were equally valid, see the philosophical account in chapter two on that matter. The fact that type two did not seem to work out technically has always been regarded by me with the necessary amount of scepticism since in my opinion, QFT did not work for QED nor the standard model either. Only sloppy and overprotective field theorists could take something like that seriously, but I was rigorous and not even protective regarding my deepest beliefs. So I have always felt that on the level of relativistic particle theory, we were lacking a few crucial insights. What I knew already for a long time was that ultimately type one was going to be the most difficult to realize; I will come back to that issue in a forthcoming publication given that some other things have to be straightened out first in our approach. These notes are about obstacles one will meet regarding the formation of a type one theory, but a real theory, not just something we can all pull out of our hats within five minutes but which lacks canonical beauty and predictive power.

In our approach so far, there are two remaining open questions (all others have been answered thorougly): (a) what is an appropriate substitute for unitarity (b) why should local gauge invariance be a principle of nature? In particular, why should the interaction structure be limited in some peculiar way? I have at this point no good answer to those two, but maybe future investigations will elucidate these matters. The confusion around these topics is, in my opinion, the work of an entire generation of post war physicists who did abandon mathematical rigour culminating in the renormalization generation of 't Hooft and Veltman. It is not gratuitious that these gentlemen are Dutchman as indeed it requires some form of talent to sell Heineken, one of the world's worst beers and secondary to any Belgian beer, to the public. The same holds for their renormalization results, they have a flair of mathematical ingenuity, but deep down it is all arbitrary nonsense. As I have shown, they have been missing quite some important *physics* which evaporates the distinction between non-renormalizable and renormalizable theories. Indeed, it has been shown that quantum gravity type two has *nothing* to do with supergravity and supersymmetry in particular. Good, let us return to type one which is further ahead of us.

Here, one immediately faces a couple of problems regarding the fact that standard formulations of quantum mechanics are *not* covariant. This is seldomly highlighted, but the problem is really everywhere: in the path integral approach, it is in the non-covariance of the measure, in the Heisenberg approach, it resides in the non-covariance of the total Hamiltonian and therefore the vacuum state, and finally in the Schrodinger approach, it is blatanly visible because the probability density does not transform as a density under coordinate transformations of space. For examples and more in depth comments regarding those issues, see chapter four. In field theory for example, one will obtain that distinct lattice regularizations, in either different choices of "measure", will give rise to different continuum limits and we wish physics to be devoid of such ambiguity. In that respect is our quantum theory generally covariant: it does not depend upon geometrical structures or coordinate choices which have to be imported. There is no choice of vacuum state, no Hamiltonian, no measure, everything has been poored in a manifest space time language. This, of course, is a great starting point for some ideas regarding a quantum gravity type one theory to mature. So, up till now, *every* approach to quantum gravity suffers from one of these drawbacks: in the discrete theories based upon the Feynman path integral, such as causal sets and causal dynamical triangulations, one remains with the choice of the measure associated to the particular regularization scheme. Some researchers accept this as a fact they have to live with, most of them are not even aware of the issue.

So, how can we extend our novel line of thought to space time itself? For example, how to define the momenta of the theory which have to serve for a gravitational uncertainty principle and what are the constraints upon the momenta replacing the on-shell mass condition for relativistic particles? Clearly, in a continuum theory of the universe some infinite dimensional integration would have to be performed which again will lose its appeal through the non-canonical character of the limit of measures (the limiting measure does not exist). Hence, if one were to have to define a quantum theory for the gravitational field, one should resort to the choice of a preferred discrete structure as being really there and not just being some approximation to the continuum situation. This is the main reason why I have always thought a quantum gravity type one theory to be discrete in some sense; a feature which is not mandatory at all in our setup for the type two graviton theory.

In a discrete universe, one obviously does abandon local Lorentz covariance in a well defined sense, albeit this does not need to have disastrous implications upon the physics defined on it. It is an important kinematical question to ask oneself how close two (discrete) universes are and I have adressed this question in my PhD work where I have defined and investigated to some extend a Gromov distance on “Lorentz geometries”. Albeit I have been very humble about the applications of this work in the past, given the recent importance of the metric in defining quantum theory, it occured to me that this Gromov distance could be of direct physical significance too when defining geodesics and therefore the Fourier transform on a space of space times. This could even be calculated exactly for finite universes albeit it would become a very complex task to do so when considering large universes. Apart from the technical complications associated to such scheme, there is an immediate philosophical issue regarding the Riemannian nature of this Gromov super metric. Why should it not be a Lorentzian one albeit the natural criterion for closeness immediately leads one to a Riemannian instead of Lorentzian structure. In the latter case could one entertain concepts such as the “cause of causality” where, as explained in chapter two, it would be better to speak about the evolution of properties rather than causality. Strictly speaking, there is no need for time in the evolution of geometries since time already lives inside the universe, so a Riemannian distance will do just fine.

I can ensure the reader that the Gromov distance is canonical and given the finiteness of the space of kinematical structures, we should obtain a fairly unique definition of the Fourier transform which allows one to define the quantum theory. In a sense, this “weakens” my objection against the rather arbitrary character of the quantum measure by means of choice of a kinematical regularization scheme, given that we have chosen here some finite structures ourselves in the construction of the theory. Of course, the choice of measure has still many more degrees of freedom than merely picking out a kinematical structure as often the uniform measure does not provide one with suitable convergence properties. Moreover, as we did notice in this book, unitarity is dead which supports the idea that the quantal measure will not be obvious or canonical either. Therefore, our scheme appears to have the least amount of freedom and to be the most canonical possible. These considerations immediately imply that a quantum universe can have only a finite extend and that therefore, the full universe needs to have classical components. So, our type one quantum theory also leads one to consider a classical-quantum universe, see chapter two for further thoughts about these issues. I could at this point be more technical, but I do not wish to be so given that I intend to investigate these matters in the future; the interested reader can always find my PhD thesis on arXiv or contact me if he or she wishes to collaborate on the issue.

Let $\mathcal{M}$ be a compact $n+1$-dimensional manifold with initial and final boundaries Σ_1 and Σ_2 which we shall assume to be spacelike regarding any further metric specification. Regarding non-compact manifolds, we compact consider sub-cobordisms with a boundary part which is timelike and normal initial data are kept fixed as well during variation so that fluctuating initial conditions

(h, π) provide for a unique solution to the Einstein equations. Consider a metric $g_{\alpha'\beta'}(x, y)$ where the primed indices refer to the y-coordinate and unprimed ones to the x-coordinate. Actually, g is a bitensor on the product manifold $\mathcal{M} \times \mathcal{M}$ with the ordinary product differentiable structure at least what concerns the second factor; mathematically, it belongs to $\mathcal{M} \times T_2\mathcal{M}_{\mathbf{sym}}$. The data on Σ_i are specified by a vierbein E_a and a vacuum solution $g_{\alpha'\beta'}(x, y)$ for any $x \in \Sigma_1$ and $y \in \mathcal{M}$ with initial data $h_{\alpha'\beta'}(x, y) = h_{\alpha'\beta'}(y) = \sum_i E_{i\alpha'}(y)E_{i\beta'}(y)$ where $E_i \in T^1\Sigma$ could be constructed. One could further personify the metric and induce a particular x-dependency; this is certainly not at odds with free will but we shall not consider this option in this paper. Now, every $h \in T_2\Sigma$ defines a Fourier transform in $L^2(\Sigma, h, x)$ where x is a reference point in Σ:

$$\pi_{h,x}(y) = \int_{\mathbb{R}^n} d^n k \widehat{\pi_{x,h}}(k)\psi_{\Sigma,h}(x, y, k)$$

where $\psi_{\sigma,h}(x, y, k)$ has been defined in my previous book on covariant quantum mechanics and constitute nothing but the basic Fourier modes with respect to a reference point x. We shall henceforth assume this Fourier transform to define a diffeomorphism modulo a certain ambiguity on the function space $L^2(\mathbb{R}^n, d^n k, h, x)$ which is modelled by linear operators $\Psi_{x,h} : L^2(\mathbb{R}^n, d^n k, x, h) \to L^2(\mathbb{R}^n, d^n k, x, h)$ which are not necessary unitary with respect to the standard inner product $\langle\pi|\psi\rangle_{x,h} = \int d^n k \widehat{\pi_{x,h}}(k)\widehat{\psi_{x,h}}(k)$ where some representants $\widehat{\psi_{x,h}}, \widehat{\pi_{x,h}}$ have been chosen. Notice that these expressions are not necessarily the same as $\int d^n y \pi(y)\psi(y)$ and could be moreover heavily x dependent. Choose any minimal subspace $\mathcal{Z}_{x,h}$ of $L^2(\mathbb{R}^n, d^n k, h, x)$ containing one representant only and which glue nicely together to a Hilbert bundle on Σ modelled on $T^1\Sigma$. Likewise, it is possible to consider the space of initial data $\pi_{\alpha'\beta'}(y)$ on $T^2\Sigma$ and define a Fourier transform defined by

$$\widehat{\pi_{ab;x,h}}(k)$$

where a, b are Euclidean indices in x and the latter is a function in $\mathcal{Z}_{x,h}$. For more details, see a previous publication on the Fourier transform. Point is that the Hamiltonian constraint $H(h, \pi) = 0$ together with the diffeomorphism constraints must be rewritten in $\otimes_{n(n+1)}\mathcal{Z}_{x,h}$ which defines an infinite dimensional Riemannian submanifold $\mathcal{C}_{h,x}(\pi)$ in the induced Riemannian metric as an embedding space of an infinite dimensional Hilbert space. A canonical limiting measure may now be defined on this infinite dimensional Riemannian manifold by putting k space in finite balls of radius L and taking periodic boundary conditions in the radius as well as spherical modes with integer factor m bounded by $|m| \leq N$, so that the measure spaces $\mathcal{Z}_{x,h}(L, N)$ are finite dimensional. The natural thermodynamic limit is $N \to \infty$ followed by $L \to \infty$ in every $x \in \Sigma$. A natural inner product space and measure are therefore defined by means of these truncations and a scalar product between different (h, π) data in $\otimes_{n(n+1)}\mathcal{Z}_{x,h}$ is given by the linear extension in the π factor of

$$\langle(h_1, \pi_1)|(h_2, \pi_2)\rangle = \int_\Sigma d^n x (h_1(x)h_2(x))^{\frac{1}{4}} \int d^n k \widehat{\pi^1_{x,h_1}}(k)\widehat{\pi^2_{x,h_2}}(k)$$

and likewise so for the Riemannian metrics themselves. This scalar product has no suitable additive properties with respect to the spatial metric. The respective scalar product has a differentiable action of Diff(Σ) by means of unitary operators and therefore the identity component of the spatial diffeomorphism group is generated by unitary operators. This brings us to a first point of comparison with more popular quantization methods such as Loop Quantum Gravity. There, no adequate scalar product has been found and generators of the diffeomorphism constraints simply do not exist mathematically whereas they do here as well defined operators.

This brings us back to the idea that a spacetime in the state $|h\rangle$ represents the birth of an entire universe and that the appropriate integration over different metrics is done by means of this "background" metric. This intertwining between dynamics and kinematics was present already in my covariant quantum theory published herefore and it canonically defines spacetime particles and the appropriate propagators associated to the birth of a spatial universe in contrast to the standard quantum theory on Minkoswki where the maximal symmetry makes this association canonical. This subtly breaks the naive linearity of Minkowskian quantum mechanics and makes the moduli problem totally redundant given that a reference metric is part of the appropriate parlee to frame physical questions. Specifically, denote by the metric h on the boundary Σ the *birth* metric of a spatial universe; then a wave function Φ in h' satisfying $Z_i(h') = 0$ can be fully developed with respect to h. That is,

$$h'_{\alpha'\beta'} = \Gamma^a_{\alpha'}(x,y)\Gamma^b_{\beta'}(x,y) \int d^n k \widehat{h'}_{ab;x,h}(k)\psi(x,y,k,x,h)$$

assuming that x, y are connected by exactly one h geodesic and $\Gamma^a_{\alpha'}(x,y)$ is the propagator along that geodesic. A more complicated but equally adequate formula exists when multiple h geodesics connect x with y. The diffeomorphism constraints on h' are most easily expressed in terms of integral equations of the Fourier components $\widehat{h'}_{ab;x,h} \in \mathcal{Z}_{x,h}$. Hence, $\Phi(h')$ must be decomposed with respect to the induced metric on the h' submanifold $\mathcal{D}(h,h',x)$ in $\otimes_{\frac{n(n+1)}{2}} \mathcal{Z}_{x,h}$ defined by the momenta satisfying the constraints $Z_i(h,\pi) = Z_i(h',\pi) = 0$. The usual L, N filtration gives then a sequence of measures $d\mu_{L,N}$ and Fourier transforms defined by the canonical flat metric on tangent space such that

$$\Phi_{h,x}(h') = \lim_{L,N\to\infty} \int_{T_{0,x}\mathcal{D}(h,h',x)(L,N)} d\pi_{ij;x,h} \widehat{\Phi_{x,h;L,N}}(\pi_{ij;x,h,h'})\psi(h,h',x,\pi_{ij;x,h,h'})$$

where $\pi_{ij;x,h,h'}$ are the cotangent bundle coordinates in a orthogonal basis of $T_{0,x}\mathcal{D}(h,x)(L,N)$ and ψ denotes the usual Fourier functions. The Fourier decompositions differ when varying the point x but crucially depend upon h and one canonically defines a canonical diffeomorphism invariant scalar product

$$\langle \Psi_h | \Phi_h \rangle = \lim_{L,N\to\infty} \int_{\Sigma_1} d^n x \sqrt{h(x)} \int_{T_{h,x}\mathcal{D}(h,x)(L,N)(h)} d\pi_{ij;x,h} \widehat{\Psi_{x,h;L,N}}(\pi_{ij;x,h}) \widehat{\Phi_{x,h;L,N}}(\pi_{ij;x,h})$$

for wavefunctions regarding a universe born at (Σ_1, h). Given a compatible pair (h, π), one solves for the vacuum Einstein equations of motion

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 0$$

with any acceptable lapse and shift vector so that all variables remain well defined; for example, the Gaussian gauge could break down before Σ_2 is reached. Now, we wish to study the evolution of $\Psi_h(h')$ in "time". Given a point y in $(\mathcal{M}, g(h, \pi, N_\mu))$ to the future of Σ_1, denote by $P(y, g)$ the subset of pairs $(x, v) \in \Sigma_1 \times T\mathcal{M}_{|\Sigma_1}$ such that $\exp_x(v) = y$. The spacetime Gaussian $\psi(h, \pi, N_\mu; h', y)$ is given as

$$\psi(h, \pi, N_\mu; h', y) = \int_{P(y, g(h, \pi, N_\mu))} d^n x \sqrt{\tilde{h}(x)} \psi(h, \pi, N_\mu; h', x, v)$$

where $\tilde{h}$ is the metric induced on $P(y, g)$ by means of the product metric $h \otimes \delta$ on $\Sigma_1 \times T_1 \Sigma_1$. The functions $\psi(h, \pi, N_\mu; h', x, v)$ are determined by means of the standard Schroedinger equation

$$\frac{d}{ds} \psi(h, \pi, N_\mu; h', x, v; s) = i \sqrt{\pi_{\mu\nu;s}(\gamma_{x,v}(s)) g'^{\mu\kappa}(\gamma_{x,v}(s)) \dot{\gamma}^{\nu}_{x,v}(s) \pi_{\kappa\lambda;s}(\gamma_{x,v}(s)) \dot{\gamma}^{\lambda}_{x,v}(s)} \psi(h, \pi, N_\mu; h', x, v; s)$$

with conditions $\gamma_{x,v}(0) = x, \gamma_{x,v}(1) = y$ and the tensors π, h' on Σ_1 have canonical extensions towards $\mathcal{M}$. The tensors $\pi_{\mu\nu;x}(\gamma_{x,v}(s))$ are covariantly constant along the g-geodesic and their value at the endpoint may vary from starting point to starting point x; also, the initial condition $\psi(h, \pi, N_\mu; h', x, v; 0) = 1$ holds as usual. The Hamiltonian constraint is as such preserved along any geodesic and is entirely foliation independent. Keeping N^μ fixed, we therefore can define a propagator

$$D(h, N_\mu, h'; x, y) = \lim_{L,N \to \infty} \int_{\Sigma_1} d^n z \sqrt{h(z)} \int_{\mathcal{C}_{z,h,h'}(L,N)} d\pi_{ij;z,h,h',L,N} \delta(H(h, \pi)) \delta(H(h', \pi))$$
$$\psi(h, \pi, N_\mu; h', x, y)$$

which is nothing but the adequate substitute for the graviton propagator on the background (h, π) in a first quantized theory. Here, $\mathcal{C}_{z,h,h'}(L, N)$ is the Hilbert manifold of momenta π expressed with respect to the birth metric h and the conditions $Z_i(\pi, h) = Z_i(\pi, h') = 0$ are satisfied. Alas, as we shall soon see, this object is not of much use. A second quantized theory would demand a four index bitensor propagator with two indices in every reference point each.

The evolution therefore is

$$\Phi_{h,\Sigma_1}(\overline{h}, N_\mu, \Sigma_2) = \lim_{L,N \to \infty} \int_{\Sigma_2} d^n y \sqrt{\overline{h}(y)} \int_{\Sigma_1} d^n x \sqrt{h(x)}$$
$$\int_{\mathcal{C}_{h,x}(L,N)} d\pi_{ij;x,h;L,N} \int_{K(h,\pi,N_\mu,\Sigma_1,x;\overline{h},y,\Sigma_2)} d\mu_{h,\pi,N_\mu,\Sigma_1,x;\overline{h},y\Sigma_2}(h') \Phi_{h,\Sigma_1}(h', \Sigma_1)$$

$$\delta(Z_i(\pi, h'))\delta(H(h, \pi))\delta(H(h', \pi))\psi(h, \pi, N_\mu; h', x, y)$$

where $K(h, \pi, N_\mu, \Sigma_1, x; \overline{h}, y, \Sigma_2)$ is the Hilbert manifold of Riemannian metrics h' on Σ_1 canonically metricized and equipped with the limiting volume form by means of $T\Sigma_{1,x}$ such that the projection of

$$\int_{P(y, g(h, \pi, N_\mu))} d^n z \sqrt{\tilde{h}(z)} g'_{\mu\nu;z}(y)$$

on Σ_2 equals $\overline{h}_{\alpha\beta}(y)$ and $H(h', \pi) = 0, Z_i(\pi, h') = 0$. All these formulae are divergent and need further regularization akin to the procedures encountered in the second chapter of this book.

Chapter 13

Quantization of the free string.

The Virasoro problem in string theory arises most clearly in the covariant quantization where one has hermitian generators L_n with $n \in \mathbb{Z}$ which have to be regarded as constraints; that is physical states have to satisfy $L_n|\Psi\rangle = 0$ for $n \neq 0$ and $L_0|\Psi\rangle = a|\Psi\rangle$ with $a \neq 0$. The Virasoro algebra without central anomalies $c(n)$,

$$[L_n, L_m] = i(n-m)L_{n+m} + c(n-m)1$$

makes this impossible given that

$$0 = [L_n, L_{-n}]\,|\Psi\rangle = 2inL_0|\Psi\rangle = 2ina|\Psi\rangle$$

which contradicts $a \neq 0$. The "fix" of the problem is to keep the constraints $L_n|\psi\rangle = 0$ for $n > 0$ while dropping the others. This leads to physical operators changing particle species, spin and angular momentum causing all known conservation laws of particle physics to fail (but not largely in practice). The downside is that the geometrical description of the theory is totally lost at the quantum level even in a Minkowski background and that everything becomes therefore gauge dependent. This is not expected given that quantum theory works perfectly fine for flat geometries and we shall trace back the problem to the non-geometric character of quantum theory itself. In that context, the worldsheet formulation evaporates and only reparametrisations of the type $t'(t)$ and $s'(s)$ can be made such that the Virasoro problem dissapears giving rise to two mutually commuting symmetry algebra's as the *full* symmetry algebra.

13.1 Strings from the viewpoint of covariant quantum theory.

Given a closed string worldsheet $\gamma(t, s)$, we define two vectorfields $\mathbf{V} = \partial_t\gamma(t, s)$ and $\mathbf{Z} = \partial_s\gamma(t, s)$ where $t \in [0, T]$ and $s \in [0, L]$ with periodic boundary conditions; obviously $[\mathbf{V}, \mathbf{Z}] = 0$.

The law one is looking for clearly is of the kind

$$\nabla_{\mathbf{V}}\mathbf{V} = \mathbf{F}(\mathbf{V}, \mathbf{Z}, \nabla_{\mathbf{Z}}\mathbf{Z}, \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{R}(\mathbf{Z}, \mathbf{V})\mathbf{Z}, \mathbf{h})$$

where $\mathbf{A} = \nabla_{\mathbf{Z}}\mathbf{Z}$ is a kind of acceleration, $\mathbf{h}$ the, possibly degenerate, metric on the string and all Riemann curvature terms involve the intrinsic geometry of the string. The problem so far is that the velocity field $\mathbf{V}$ is randomly chosen and that therefore it is desirable to impose constraints on $\nabla_{\mathbf{V}}\mathbf{Z}$. We have basically two types: (a) one involving the extrinsic geometry and the latter only the intrinsic geometry. In other words, we have (I)

$$\nabla_{\mathbf{V}}\mathbf{Z} = \mathbf{G}(\mathbf{V}, \mathbf{Z}, \nabla_{\mathbf{Z}}\mathbf{Z}, \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{R}(\mathbf{Z}, \mathbf{V})\mathbf{Z}, \mathbf{h})$$

or (II)

$$\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z}, \mathbf{V}) = P(\mathbf{V}, \mathbf{Z}, \nabla_{\mathbf{Z}}\mathbf{Z}, \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{R}(\mathbf{Z}, \mathbf{V})\mathbf{Z}, \mathbf{h})$$

whereas a condition of the kind

$$\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z}, \mathbf{Z}) = Q(\mathbf{V}, \mathbf{Z}, \nabla_{\mathbf{Z}}\mathbf{Z}, \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{R}(\mathbf{Z}, \mathbf{V})\mathbf{Z}, \mathbf{h})$$

is meaningless given that consistency would bring it down to an algebraic condition on $\mathbf{g}(\mathbf{Z}, \mathbf{Z})$. Such theories are usually empty and therefore not interesting at all.

One has to demand now that the dynamics preserves the constraint; that is

$$\nabla_{\mathbf{V}}\nabla_{\mathbf{V}}\mathbf{Z} = \nabla_{\mathbf{V}}\mathbf{G} = \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V} + \nabla_{\mathbf{Z}}\mathbf{F}$$

a consistency condition. Note that

$$\nabla_{\mathbf{Z}}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}) = (\nabla_{\mathbf{Z}}\mathbf{R})(\mathbf{V}, \mathbf{Z})\mathbf{V} - \mathbf{R}(\mathbf{G}, \mathbf{Z})\mathbf{V} + \mathbf{R}(\mathbf{V}, \mathbf{A})\mathbf{V} - \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{G}$$

which can be reduced to, by means of the second Bianchi identity to

$$\nabla_{\mathbf{Z}}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}) = (\nabla_{\mathbf{V}}\mathbf{R})(\mathbf{V}, \mathbf{Z})\mathbf{Z} + \mathbf{R}(\mathbf{Z}, \mathbf{G})\mathbf{V} + \mathbf{R}(\mathbf{V}, \mathbf{A})\mathbf{V} - \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{G}.$$

On the other hand, a similar compuation gives that

$$\nabla_{\mathbf{V}}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{Z}) = (\nabla_{\mathbf{V}}\mathbf{R})(\mathbf{V}, \mathbf{Z})\mathbf{Z} + \mathbf{R}(\mathbf{F}, \mathbf{Z})\mathbf{Z} + \mathbf{R}(\mathbf{V}, \mathbf{G})\mathbf{Z} + \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{G}$$

where no second Bianchi identity has been used and the other terms do not allow for comparison between $\mathbf{F}$ and $\mathbf{G}$ by means of the latter identity. Contractions with the spacetime metric do allow for further use of the first Bianchi identity and gives rise to a larger margin to construct stringy laws. Hence, in light of the conservation law for the constraint,

$$\mathbf{F}(\mathbf{V}, \mathbf{Z}, \nabla_{\mathbf{Z}}\mathbf{Z}, \mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{h})$$

and

$$\mathbf{G}(\mathbf{V},\mathbf{Z},\nabla_{\mathbf{Z}}\mathbf{Z},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{h})$$

with

$$\frac{\delta\mathbf{F}}{\delta\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}} = \frac{\delta\mathbf{G}}{\delta\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z}}.$$

We also have that

$$\frac{\delta\mathbf{G}}{\delta\mathbf{Z}}\triangle\mathbf{G} + \frac{\delta\mathbf{G}}{\delta\mathbf{V}}\triangle\mathbf{F} + \frac{\delta\mathbf{G}}{\delta\mathbf{A}}\triangle\nabla_{\mathbf{V}}\mathbf{A}$$

$$= \frac{\delta\mathbf{F}}{\delta\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}}\triangle(-\mathbf{R}(\mathbf{G},\mathbf{Z})\mathbf{V}+\mathbf{R}(\mathbf{V},\mathbf{A})\mathbf{V}-\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G}-\mathbf{R}(\mathbf{F},\mathbf{Z})\mathbf{Z}-\mathbf{R}(\mathbf{V},\mathbf{G})\mathbf{Z}-\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G})$$

$$+\frac{\delta\mathbf{F}}{\delta\mathbf{Z}}\triangle\mathbf{A} - \frac{\delta\mathbf{F}}{\delta\mathbf{V}}\triangle\mathbf{G} + \frac{\delta\mathbf{F}}{\delta\mathbf{A}}\triangle\nabla_{\mathbf{Z}}\mathbf{A} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}.$$

From a generalist point of view, this would suggest

$$-\mathbf{R}(\mathbf{G},\mathbf{Z})\mathbf{V}+\mathbf{R}(\mathbf{V},\mathbf{A})\mathbf{V}-\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G}-\mathbf{R}(\mathbf{F},\mathbf{Z})\mathbf{Z}-\mathbf{R}(\mathbf{V},\mathbf{G})\mathbf{Z}-\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G} = 0$$

as well as

$$\frac{\delta\mathbf{G}}{\delta\mathbf{Z}}\triangle\mathbf{G} + \frac{\delta\mathbf{G}}{\delta\mathbf{V}}\triangle\mathbf{F} + \frac{\delta\mathbf{G}}{\delta\mathbf{A}}\triangle\nabla_{\mathbf{V}}\mathbf{A} = \frac{\delta\mathbf{F}}{\delta\mathbf{Z}}\triangle\mathbf{A} - \frac{\delta\mathbf{F}}{\delta\mathbf{V}}\triangle\mathbf{G} + \frac{\delta\mathbf{F}}{\delta\mathbf{A}}\triangle\nabla_{\mathbf{Z}}\mathbf{A} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}.$$

It is immediately seen that, in general and independent of this ansatz,

$$\frac{\delta\mathbf{F}}{\delta\mathbf{A}} = 0$$

given that higer spatial derivatives do not occur elsewhere in the formula and therefore

$$\mathbf{F}(\mathbf{V},\mathbf{Z},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V})$$

given that we have already neglected $\mathbf{h}$. On the other hand

$$\nabla_{\mathbf{V}}\mathbf{A} = \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z} + \nabla_{\mathbf{Z}}\mathbf{G}$$

which implies that

$$\frac{\delta\mathbf{G}}{\delta\mathbf{A}} = 0$$

due to consistency given that no algebraic relations are allowed for between higher spatial derivatives. Hence,

$$\mathbf{G}(\mathbf{V},\mathbf{Z},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z})$$

and we conclude from the remaining master equation that only intrinsic contractions of the Riemann tensor with $\mathbf{V},\mathbf{Z}$ are allowed for to eliminate the nasty

$$\mathbf{R}(\mathbf{V},\mathbf{A})\mathbf{V}$$

term. This however happens in two different ways $\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{A})\mathbf{V},\mathbf{V}) = 0$ identically whereas contractions of the kind

$$\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z})$$

require a balancing between

$$\frac{\delta \mathbf{F}}{\delta \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}} \triangle \mathbf{R}(\mathbf{V},\mathbf{A})\mathbf{V}$$

and

$$\frac{\delta \mathbf{F}}{\delta \mathbf{Z}} \triangle \mathbf{A}$$

in the sense that they have to be equal to one and another due to the first Bianchi identity. As a conclusion, we further specify that

$$\mathbf{F} = X(\mathbf{V},\mathbf{Z},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}))\mathbf{V} + Y(\mathbf{V},\mathbf{Z},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}))\mathbf{Z}$$

which automatically satifies this requirement by means of symmetries of the Riemann tensor. This further limits

$$\mathbf{G} = R(\mathbf{V},\mathbf{Z},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{Z}))\mathbf{V} + S(\mathbf{V},\mathbf{Z},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{V}))\mathbf{Z}$$

with

$$\frac{\delta X}{\delta \mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z})} = -\frac{\delta R}{\delta \mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{Z})}$$

and

$$\frac{\delta Y}{\delta \mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z})} = -\frac{\delta S}{\delta \mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{Z})}.$$

This shows that ∇ cannot be the Christoffel connection of a Riemannian metric and $\mathbf{R}$ its associated Riemann tensor. Although the Riemann tensor of any connection satisfies the second Bianchi identies, the first Bianchi identies and the associated symmetries of the Riemann tensor follow from the metric and torsionless character. Therefore, the connection needs torsion for the subsequent analysis to hold.

Given that one would expect only curvature to occur in the acceleration law of the string and moreover that the acceleration is of the geodesic type so that the sring t coordinate is nothing but a rescaling of the geodesic time, reparamentrization invariance has to be given up in the light of the fact that no $g(\mathbf{Z},\mathbf{Z})$ or $g(\mathbf{Z},\mathbf{V})$ terms may occur due to an inappropriate appearance of $\mathbf{A}$ in the

$$\frac{\delta \mathbf{F}}{\delta \mathbf{Z}} \triangle \mathbf{A}$$

term. Therefore,

$$\nabla_{\mathbf{V}}\mathbf{V} = \frac{c}{L^3}\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z})\mathbf{V}$$

where c is the speed of light and L has units of meters. This is the correct way of looking at it given that the curves are ordinary geodesics again but then reparametrized in a way as to balance the tidal forces; t can be reparametrized but generally speaking only *one* worldline of a point of the circle can have unit time parametrization. This is a salient feature given that strings will not induce superluminal effects in this way by means of its nonlocal character. In particular, we have that if x is a point past to the string and $\mathbf{V}$ is a future pointing timelike vectorfield, then the entire string will remain within $I^+(x)$. Finally,

$$\nabla_{\mathbf{V}}\mathbf{Z} = K(\mathbf{g}(\mathbf{V},\mathbf{Z}),\mathbf{g}(\mathbf{V},\mathbf{V}),\mathbf{g}(\mathbf{Z},\mathbf{Z}))\mathbf{V} + \\ -\frac{c}{L^3}\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{Z},\mathbf{Z}))\mathbf{V} + L(\mathbf{g}(\mathbf{V},\mathbf{Z}),\mathbf{g}(\mathbf{V},\mathbf{V}),\mathbf{g}(\mathbf{Z},\mathbf{Z}))\mathbf{Z}.$$

The consistency equation has now been reduced to

$$\frac{\delta\mathbf{G}}{\delta\mathbf{Z}}\triangle\mathbf{G} + \frac{\delta\mathbf{G}}{\delta\mathbf{V}}\triangle\mathbf{F} =$$

$$-\frac{\delta\mathbf{F}}{\delta\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}}(\mathbf{R}(\mathbf{G},\mathbf{Z})\mathbf{V} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G} + \mathbf{R}(\mathbf{F},\mathbf{Z})\mathbf{Z} + \mathbf{R}(\mathbf{V},\mathbf{G})\mathbf{Z} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{G})$$

$$-\frac{\delta\mathbf{F}}{\delta\mathbf{V}}\triangle\mathbf{G} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}.$$

As expected one page ago, this equation can only have solution in case $\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}$ equals its projection on the string worldsheet determined by the $\mathbf{V},\mathbf{Z}$ plane which is in general impossible except for Einstein spaces. Therefore, it might be possible to develop a type I string theory for Einstein spaces with torsion but given such restriction it is utterly clear that type II is the only physical case.

Here, we might try to arrive at a theory with equation of motion

$$\nabla_{\mathbf{V}}\mathbf{V} = \mathbf{F}$$

and constraint equations

$$\begin{aligned} \mathbf{g}(\mathbf{V},\nabla_{\mathbf{V}}\mathbf{Z}) &= \alpha\mathbf{g}(\mathbf{V},\mathbf{G}) \\ \mathbf{g}(\mathbf{E}_i,\nabla_{\mathbf{V}}\mathbf{Z}) &= \alpha\mathbf{g}(\mathbf{E}_i,\mathbf{G}) \\ \frac{1}{2}\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z},\nabla_{\mathbf{V}}\mathbf{Z}) &= \alpha\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z},\mathbf{G}) \end{aligned}$$

where E_i is a $n-2$ bein orhogonal to $\mathbf{V},\mathbf{Z}$. In vector language, this gives

$$\nabla_{\mathbf{V}}\mathbf{Z} - \alpha\mathbf{G} = \mathbf{W}$$

with $\mathbf{W}$ perpendicular to the $n-1$ plane defined by $\mathbf{V},\mathbf{E}_i$. Moreover,

$$\mathbf{g}(\mathbf{W} - \alpha\mathbf{G},\nabla_{\mathbf{V}}\mathbf{Z}) = 0.$$

Hence,

$$\mathbf{g}(\mathbf{W},\mathbf{W}) = \alpha^2 \mathbf{g}(\mathbf{G},\mathbf{G}).$$

The structure of these equations is as such that they are preserved during time evolution. Time evolution of the first gives

$$\mathbf{g}(\mathbf{F},\nabla_{\mathbf{V}}\mathbf{Z}) + \mathbf{g}(\mathbf{V},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V}) + \mathbf{g}(\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{F}) =$$

$$\alpha\mathbf{g}(\mathbf{F},\mathbf{G}) + \alpha\mathbf{g}(\mathbf{V},\nabla_{\mathbf{V}}\mathbf{G})$$

which generically leads to

$$\mathbf{g}(\mathbf{F},\nabla_{\mathbf{V}}\mathbf{Z} - \alpha\mathbf{G}) = 0,\ \mathbf{g}(\mathbf{V},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V} + \nabla_{\mathbf{Z}}\mathbf{F} - \alpha\nabla_{\mathbf{V}}\mathbf{G}) = 0.$$

The other equations are

$$\mathbf{g}(\mathbf{E}_i,\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V} + \nabla_{\mathbf{Z}}\mathbf{F} - \alpha\nabla_{\mathbf{V}}\mathbf{G}) = 0$$

and

$$\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z},\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V} + \nabla_{\mathbf{Z}}\mathbf{F} - \alpha\nabla_{\mathbf{V}}\mathbf{G}) = 0$$

supplemented with

$$\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{F} + \mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{G}) = 0$$

which obviously gives the same problems as before. It appears some more delicate analysis is necessary: clearly, one would like

$$\nabla_{\mathbf{V}}\mathbf{V} = \mathbf{F}(\mathbf{V},\mathbf{Z},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}))$$

given that $\mathbf{Z}$ is chosen according to arc length and evolution should only depend upon the intrinsic geometry and only as far on the directions perpendicular to the infinitesimal string surface as the acceleration goes. That is

$$\mathbf{g}(\mathbf{Z},\mathbf{A}) = 0$$

and one would like to preserve this property under evolution, in either keep it as a constraint. Time evolution gives

$$\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{Z},\mathbf{A}) + \mathbf{g}(\mathbf{Z},\nabla_{\mathbf{Z}}\nabla_{\mathbf{V}}\mathbf{Z}) = 0 = \nabla_{\mathbf{Z}}(\mathbf{g}(\mathbf{Z},\nabla_{\mathbf{V}}\mathbf{Z})).$$

Therefore, we should add as constraint

$$\mathbf{g}(\mathbf{Z},\nabla_{\mathbf{V}}\mathbf{Z}) = \nabla_{\mathbf{Z}}\mathbf{g}(\mathbf{Z},\mathbf{V}) - \mathbf{g}(\mathbf{A},\mathbf{V}) = 0$$

which follows from

$$\mathbf{g}(\mathbf{Z},\mathbf{V}) = \mathbf{g}(\mathbf{A},\mathbf{V}) = 0.$$

In ordinary string theory in flat Minkowski $\mathbf{F} = \mathbf{A}$ for a Lorentzian flat worldsheet metric and $\mathbf{F} = -\mathbf{A}$ for a Riemannian worldsheet metric and the former two conditions give by means of the equation of motion

$$\mathbf{g}(\mathbf{Z},\mathbf{V}) = \nabla_{\mathbf{V}}\mathbf{g}(\mathbf{V},\mathbf{V}) = 0.$$

The first of those is the usual Virasoro constraint

$$\partial_t \gamma . \partial_s \gamma = 0$$

whereas the second equals

$$\partial_t (\partial_t \gamma . \partial_t \gamma) = 0$$

which is the time derivative of one of the other constraints. Our original constraint was

$$\partial_s (\partial_s \gamma . \partial_s \gamma) = 0$$

which is the space derivative of the last Virasoro constraint. It is now possible to *impose* the constraints

$$\mathbf{g}(\mathbf{Z}, \mathbf{V}) = 0 = \mathbf{g}(\mathbf{V}, \mathbf{V}) = \mathbf{g}(\mathbf{A}, \mathbf{V})$$

where we have eliminated one integration function depending upon s only. One could leave a positive integration constant

$$\mathbf{g}(\mathbf{V}, \mathbf{V}) = \gamma$$

so that strings would move on timelike curves excluding therefore massless particles in their description. Similarly, we could demand that

$$\mathbf{g}(\mathbf{Z}, \mathbf{Z}) = 0$$

where we have eliminated a space integration constant β arising from

$$\nabla_{\mathbf{Z}} \mathbf{g}(\mathbf{Z}, \mathbf{Z}) = 2\mathbf{g}(\mathbf{A}, \mathbf{Z}) = 0.$$

We show now that the remaining three constraints close under time evolution

$$\nabla_{\mathbf{V}} \mathbf{g}(\mathbf{V}, \mathbf{Z}) = \mathbf{g}(\mathbf{F}, \mathbf{Z}) + \mathbf{g}(\mathbf{V}, \nabla_{\mathbf{V}} \mathbf{Z}) = \mathbf{g}(\mathbf{F}, \mathbf{Z})$$

where we have used the torsionless character of the Levi Civita connection and the commuting of the coordinate fields. This does not impose any constraints on the $\mathbf{F}$ field given the constraints. Finally

$$\nabla_{\mathbf{V}} \mathbf{g}(\mathbf{V}, \mathbf{V}) = 2\mathbf{g}(\mathbf{V}, \mathbf{F}) = 0$$

for similar reasons and

$$\nabla_{\mathbf{V}} \nabla_{\mathbf{V}} \mathbf{g}(\mathbf{V}, \mathbf{V}) = 2\nabla_{\mathbf{V}} \mathbf{g}(\mathbf{F}, \mathbf{V}) = 2\mathbf{g}(\nabla_{\mathbf{V}} \mathbf{F}, \mathbf{V}) + 2\mathbf{g}(\mathbf{F}, \mathbf{F}) = 0$$

where the last equality only holds in case

$$\mathbf{g}(\kappa(\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{Z}), \mathbf{A})\nabla_{\mathbf{V}} \mathbf{Z}, \mathbf{V}) + \mathbf{g}(\delta(\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{Z}), \mathbf{A})\nabla_{\mathbf{V}} \mathbf{A}, \mathbf{V}) + \delta^2(\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{Z}), \mathbf{A})\mathbf{g}(\mathbf{A}, \mathbf{A}) = 0.$$

In particular, the equation reduces by means of the torsionless character to

$$\delta(\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{Z}), \mathbf{A}) \left[\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{Z}, \mathbf{V}) + \mathbf{g}(\nabla_{\mathbf{Z}} \nabla_{\mathbf{Z}} \mathbf{V}, \mathbf{V})\right] + \delta^2(\mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{Z})\mathbf{V}, \mathbf{Z}), \mathbf{A})\mathbf{g}(\mathbf{A}, \mathbf{A}) = 0$$

where

$$\mathbf{g}(\nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{V}) = -\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{V}) = 2\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{V})\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{K}) + 2\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{Z})\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{L})$$

$$-\sum_{i=1}^{n-4}\eta_{ii}(\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{E}_i))^2$$

where η_{ij} is an $n-4$ dimensional Lorentzian vielbein in the remaining orthogonal space assuming a spacetime metric with three time directions and linearly independent $\mathbf{V},\mathbf{Z}$. Obviously, the helicity components only remain in the sense that

$$\mathbf{g}(\nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{V}) = -\sum_{i=1}^{n-4}\eta_{ii}(\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{E}_i))^2.$$

There is something extremely important about $\mathbf{g}(\nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{V})$ which we expect to happen and that is that it equals

$$\pm\mathbf{g}(\mathbf{A},\mathbf{A}).$$

In general, there is a deep connection with the equation of motion so that it is wise to suggest that

$$\delta(\mathbf{A},\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{V}),\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}))$$

which we are free to since it does not intervene with our previous analysis. That is, we have found a consistent theory with constraints

$$\mathbf{g}(\mathbf{Z},\mathbf{Z}) = \mathbf{g}(\mathbf{V},\mathbf{Z}) = \mathbf{g}(\mathbf{V},\mathbf{V}) = \mathbf{g}(\mathbf{F},\mathbf{V}) = 0$$

supplied with the condition

$$\mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{V})) = \frac{\mathbf{g}(\mathbf{R}(\mathbf{V},\mathbf{Z})\mathbf{V},\mathbf{Z}) + \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{V})}{\mathbf{g}(\mathbf{A},\mathbf{A})}.$$

Classically, these equations cannot be solved for in a spacetime with a Lorentzian signature unless $\mathbf{V}\sim\mathbf{Z}$ which is rather boring and actually occurs in "string theory". It therefore appears clear that standard string theory would require at least two independent time directions which would endager the whole edifice of causality and make no sense at all unless those time directions are compactified of some sort and far beyond our scale of observation. Hence, a fibre structure is needed for the spacetime manifold with a four dimensional Lorentzian base manifold and Lorentzian fibre. In standard quantum theory of the string with a Lorentzian world sheet and spacetime metric, one solves for right and left moving strings which should be kept strictly separate to impose the constraints. Alas, such line of reasoning is inconsistent with the Heisenberg commutation relations given that $\mathbf{V}$ and $\mathbf{Z}$ should fluctuate independently; hence, the Virasoro problem. In our setup, there are two possibilities, either one keeps $\beta < 0$ so that $\mathbf{g}(\mathbf{Z},\mathbf{Z}) = \beta$ and therefore $\mathbf{Z}$ is always spacelike involving $\mathbf{F}(\mathbf{V},\mathbf{A})$ *or* one goes over to the higher time formalism such that the projection of **Z** on the base manifold is spacelike and varying in case the fibre is one dimensional and the standard Virasoro picture with a more general force field may hold.

We now discuss these things in the next section. The real reason why in a general spacetime the left and right moving modes cannot be defined is due to the dependency of the metric on the string world sheet coordinates and hence world sheet coordinates must be endowed with a spacetime geometric meaning which is precisely what happens here. In the standard Minkowski quantization, this feature does not arise and therefore the standard procedure gives the wrong results due to the wrong signature of spacetime. Finally, *if* one would insist upon *one* string theory to describe the entire particle spectrum, then it is utterly clear that the ultrahyperbolic fibre picture with a Riemannian flat world sheet metric is obliged for. It is this picture we shall further develop in subsequent chapter.

13.2 Quantization of the string.

In ordinary particle theory, we look for the little group of the momentum vector which for massive particles equals $SO(3)$ and for massless particles E_2, the Euclidean group in two dimensions at least if the spacetime dimension equals four. To have a similar thing in string theory, we need to go to $7 = 2 + 5$ dimensions where the little group is $SO(3) \times \mathbb{R}^3 = E_3$ taking into account that the helicity has to be perpendicular to **V** as well as **Z**. This provides one with a richer particle spectrum and suggests that massive particles can travel at the speed of light in case **Z** resides exclusively in the fibre. The velocity field, being timelike, then has the speed of light in the base four dimensional spacetime which is a contradiction to standard particle theory. In plain words, the string is entirely in the fibre and behaves as a point particle from the point of view of the observer; this is a violation to Einstein's mass formula and in conflict to special relativity. It is clear that the string veloity needs to have a timelike component in the fibre manifold for a massive particle to arise; clearly, mass quantization can only occur when the bulk momenta are quantized which necessitates closed timelike curves in the fibre. Therefore, mass and in particular the mass gap, are dynamical quantities closely related to the microscopic structure of the timelike fibre which are in turn determined by the string length which suggests that the fibre is incapable of blowing up given a fixed string length scale. On the other hand, given that the fibre winding number cannot increase and the string length is fixed, formation of singularities of the bulk is excluded unless the string expands drastically in the base manifold in which case it beoomes extremely heavy as an ordinary base spacetime particle (the mass increases much more if there are extra spatial dimensions in the fibre added). Therefore, local mass eigenstates, in the sense explained below, appear to be stable unless the dynamics forces the strings to blow up in which case the mass runs astray.

In order to go over to the quantum theory, let us first define the suitable strings

$\zeta(t,s)$ where $t \in \mathbb{R}^+$ and $\zeta(0,s) \sim S^1$. Let $\mathbf{R}$ be the projection of $\mathbf{Z}$ on the plane perpendicular to $\mathbf{T} = \partial_t \zeta(t,s)$. Such hyperplane is always an $n-1$ dimensional Lorentzian, ultrahyperbolic or of a null-Lorentzian geometry. We propose now the following dragging law

$$\nabla_{\mathbf{T}}\mathbf{T} = \mathbf{F}(\mathbf{T},\mathbf{R},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{T}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{T}))$$

with

$$\mathbf{g}(\mathbf{V},\mathbf{T}) = \mathbf{g}(\mathbf{V},\mathbf{A}) = \mathbf{g}(\mathbf{Z},\mathbf{Z}) = \mathbf{g}(\mathbf{T},\mathbf{Z}) = \mathbf{g}(\mathbf{V},\mathbf{V}) = \mathbf{g}(\mathbf{T},\mathbf{A}) = 0$$

and

$$\nabla_{\mathbf{T}}\mathbf{V} = \mathbf{F}(\mathbf{V},\mathbf{R},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V})).$$

Also, it is necessary to make $\mathbf{g}(\mathbf{T},\mathbf{T}) = \lambda$ into a constant with respect to $\mathbf{T}$ and $\mathbf{Z}$. Before we check that these constraints are preserved under evolution $\nabla_{\mathbf{T}}$, we must remark that it requires at least a $3+5$ picture of spacetime or a four dimensional Lorentzian base space with a three dimensional negative Lorentzian fibre. This may sound unappealing but it has some potential to explain the origin of time in our universe as a kind of symmetry breaking due to small extra time dimensions.

First of all, it is easy to convince oneself that *all* those constraints are necessary; in case one chooses the constraint

$$\mathbf{g}(\mathbf{V},\mathbf{T}) = 0$$

to start with, then time derivation gives

$$\mathbf{g}(\mathbf{F}(\mathbf{V},\mathbf{R},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V})),\mathbf{T}) + \mathbf{g}(\mathbf{V},\mathbf{F}(\mathbf{T},\mathbf{Z},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{T}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{T}))) = 0$$

which suggests

$$\mathbf{g}(\mathbf{V},\mathbf{Z}) = \mathbf{g}(\mathbf{V},\mathbf{A}) = 0$$

and

$$\mathbf{g}(\mathbf{T},\mathbf{A}) = 0$$

as well. To be precise, the string we have so far is a worldsheet $\zeta(t,s)$ where $\mathbf{T} = \partial_t\zeta(t,s)$ and $\mathbf{Z} = \partial_s\zeta(t,s)$; $\mathbf{V}$ is treated as a vectorfield along $\zeta(t,s)$ and serves to infinitesimally thicken the worldsheet $\zeta(t,s,r)$ where the thickening is a possibly nonholonomic in the sense that $\mathbf{V}$ does not need to commute with $\mathbf{T},\mathbf{Z}$ for $r \in (-\epsilon,\epsilon)$ and all previous equations only hold for $r = 0$. We are now in position to compute the time evolution of $\mathbf{g}(\mathbf{V},\mathbf{Z}) = 0$ which results in

$$\mathbf{g}(\mathbf{F}(\mathbf{V},\mathbf{R},\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V})),\mathbf{Z}) + \mathbf{g}(\mathbf{V},\nabla_{\mathbf{T}}\mathbf{Z}) = 0$$

which suggests $\mathbf{F}$ not to depend upon $\mathbf{Z}$, $\mathbf{g}(\mathbf{A},\mathbf{Z}) = 0$ and $\mathbf{g}(\mathbf{V},\nabla_{\mathbf{T}}\mathbf{Z}) = 0$. The preservation in time of the former condition gives

$$\mathbf{g}(\nabla_{\mathbf{Z}}\nabla_{\mathbf{T}}\mathbf{Z},\mathbf{Z}) + \mathbf{g}(\mathbf{A},\nabla_{\mathbf{T}}\mathbf{Z}) = 0$$

which results again in

$$\nabla_{\mathbf{Z}}\mathbf{g}(\nabla_{\mathbf{T}}\mathbf{Z},\mathbf{Z}) = 0.$$

This constraint is as usual replaced by considering that

$$\mathbf{g}(\nabla_{\mathbf{T}}\mathbf{Z},\mathbf{Z}) = 0$$

giving rise to

$$\mathbf{g}(\mathbf{T},\mathbf{Z}) = \rho$$

which is equivalent to

$$\nabla_{\mathbf{T}}\mathbf{g}(\mathbf{Z},\mathbf{Z}) = 0.$$

Time evolution of the former constraint gives rise to $\rho = 0$ or a further functional restriction to $\mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V}))$ and $\mathbf{g}(\mathbf{T},\mathbf{T}) = \lambda$. This constraint is only preserved only in case $\mathbf{g}(\mathbf{T},\mathbf{T}) = 0$ or in case the former restriction on the force field holds which is the minimal case in a sense. The former condition is clearly nonsensical so that *quantum theory* imposes a restriction on $\mathbf{F}$ which was not available classically; that is,

$$\mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V})).$$

We shortly remark now that adding $\mathbf{g}(\mathbf{V},\nabla_{\mathbf{T}}\mathbf{Z}) = 0$ as a constraint is a hard thing to do as it is generally not preserved in time t. The crucial remark however is that we do not really need it if we restrict our force field as previously and we shall simply add the constraints $\mathbf{g}(\mathbf{T},\mathbf{Z}) = 0 = \mathbf{g}(\mathbf{Z},\mathbf{Z})$ and $\mathbf{g}(\mathbf{T},\mathbf{T}) = \lambda$ to our portofolio instead of $\mathbf{g}(\mathbf{V},\mathbf{Z}) = 0$. To verify that time evolution of $\mathbf{g}(\mathbf{V},\nabla_{\mathbf{T}}\mathbf{Z}) = 0$ gives trouble, we remark that

$$\mathbf{g}(\mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V})),\nabla_{\mathbf{T}}\mathbf{Z})+$$

$$\mathbf{g}(\mathbf{V},\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{T}) + \mathbf{g}(\mathbf{V},\nabla_{\mathbf{Z}}\mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{T}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{T})))$$

which reduces further to

$$\delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{V}),\mathbf{A})\mathbf{g}(\mathbf{A},\nabla_{\mathbf{T}}\mathbf{Z}) + \mathbf{g}(\mathbf{V},\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{T}) -$$

$$\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V},\mathbf{A})\delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{T}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{T}),\mathbf{A}).$$

It is clear that this constraint on $\nabla_{\mathbf{Z}}\mathbf{V}$ is hard to preserve in time and therefore the original constraint $\mathbf{g}(\mathbf{V},\mathbf{Z}) = 0$ is inadequate.

Remains to further investigate the consistency of the two remaining constraints; one sees that the preservation of $\nabla_{\mathbf{T}}\mathbf{g}(\mathbf{Z},\mathbf{Z}) = 0$ in the sense that

$$\nabla_{\mathbf{T}}\nabla_{\mathbf{T}}\mathbf{g}(\mathbf{Z},\mathbf{Z}) = 0$$

is equivalent to

$$\nabla_{\mathbf{T}}\mathbf{g}(\mathbf{T},\mathbf{A}) = 0 = \delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{T}),\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T},\nabla_{\mathbf{Z}}\mathbf{T}),\mathbf{A})\mathbf{g}(\mathbf{A},\mathbf{A})+\mathbf{g}(\mathbf{T},\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z})+\mathbf{g}(\mathbf{T},\nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{T}).$$

Here,

$$\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{T}) = -\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{T})$$

leading to

$$\delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V})) = \frac{\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{T},\mathbf{Z}) + \mathbf{g}(\nabla_{\mathbf{T}}\mathbf{Z}, \nabla_{\mathbf{T}}\mathbf{Z})}{\mathbf{g}(\mathbf{A},\mathbf{A})}.$$

Finally,

$$\nabla_{\mathbf{T}}\mathbf{g}(\mathbf{V},\mathbf{A}) = 0$$

is fully equivalent for the defining equation of $\delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V}), \mathbf{A})$, namely

$$\delta(\mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z},\mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V}), \mathbf{A})\mathbf{g}(\mathbf{A},\mathbf{A}) + \mathbf{g}(\mathbf{V}, \nabla_{\mathbf{T}}\mathbf{A}) = 0.$$

One moreover has that

$$\mathbf{g}(\mathbf{V}, \nabla_{\mathbf{T}}\mathbf{A}) = \mathbf{g}(\mathbf{V}, \mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z}) + \mathbf{g}(\nabla_{\mathbf{Z}}\nabla_{\mathbf{Z}}\mathbf{T}, \mathbf{V}) = -\mathbf{g}(\mathbf{Z}, \mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{V}) - \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V}).$$

This gives $\mathbf{F}$ back its old functional description which shows the adequacy of our approach.

It is clear that the constraint $\mathbf{g}(\mathbf{V},\mathbf{T}) = 0$ and especially its consequence $\mathbf{g}(\mathbf{V},\mathbf{Z}) = 0$ are not be needed at all, but that it is possible to formulate an alternative theory with constraints

$$\mathbf{g}(\mathbf{Z},\mathbf{Z}) = \mathbf{g}(\mathbf{V},\mathbf{V}) = \mathbf{g}(\mathbf{T},\mathbf{Z}) = \mathbf{g}(\mathbf{T},\mathbf{A}) = \mathbf{g}(\mathbf{V},\mathbf{A}) = 0$$

as well as $\mathbf{g}(\mathbf{T},\mathbf{T}) = \lambda$. It is easy to see that this theory closes and that it regards another characterization of string momentum space where the classical limit $\mathbf{T} \sim \mathbf{V}$ is more subtle. Note that this picture can be accomplished in a 2+5 setting and that the little group of a string generally is given by $SO(3) \times \mathbb{R}^3$. We now come to the characterization of quantal free string states.

13.3 Fourier transform for strings in covariant quantum theory.

We shall be brief here and leave further development for subsequent work. It is clear we have to consider $\phi(S, \mathbf{V}_S, S')$, where S is a null string parametrized with arclength and $\mathbf{V}_S$ is a null vectorfield defined on the string. Rescalings of $\mathbf{T}$ and $\mathbf{V}_S$ with a different constant leaves the physics invariant as it should as occurs for null particles but not massive ones. Clearly, the previous evolution equations for some $\mathbf{T}$ satisfying the appropriate constraints must bring S to S' and define a transporter one vector, one co-vector bi-field $\Lambda(S, S')$ giving rise to $\Lambda(S, S')(\mathbf{V}_S)$. Clearly, the correct equation is, for example, given by

$$\frac{d}{dt}\phi(S, \mathbf{V}_S, S'; t) = i\frac{\kappa}{L}\left(\int_0^L ds\, \mathbf{g}(\mathbf{V}(t,s,0), \mathbf{T}(t,s,0))\right)\phi(S, \mathbf{V}_S, S'; t)$$

with $\phi(S, \mathbf{V}_S, S'; 0) = 1$ and $\phi(S, \mathbf{V}_S, S'; 1) = \phi(S, \mathbf{V}_S, S')$. Here, κ is a dimensionless constant and L is the string length which eliminates all affine rescalings $\mathbf{Z} \to \gamma \mathbf{Z}$ from the theory. Trying to eliminate it would be a hopeless task since the natural expression is given by

$$\frac{\mathbf{g}(\mathbf{V}(t,s,0), \mathbf{T}(t,s,0))\mathbf{g}(\mathbf{V}(t,s,0), \mathbf{Z}(t,s,0))}{\mathbf{g}(\mathbf{V}(t,s,0), \mathbf{V}(t,s,0))}$$

which involves a division by zero and therefore makes little sense. In the second quantum mechanical theory, the momentum factor is time independent and given by

$$\int_0^L ds\, \mathbf{g}(\mathbf{V}(0,s,0), \mathbf{T}(0,s,0))$$

a salient feature completely analogous to the quantum theory of a point particle. Regarding the first quantum theory, the energy factor vanishes which necessitates further terms in the integral. There is however a term which does not require a string length and is given by

$$\int_0^L ds\, \mathbf{g}(\nabla_{\mathbf{Z}} \mathbf{V}, \mathbf{T})$$

at least if the $\mathbf{V}$ field is treated in a dimensionless way. This integral can be computed to be

$$-\int_0^L ds\, \mathbf{g}(\mathbf{V}, \nabla_{\mathbf{T}} \mathbf{Z}) = -\partial_t \int_0^L ds\, \mathbf{g}(\mathbf{V}, \mathbf{Z}).$$

This integral, however, is time dependent as the integrand cannot be written as a total s derivative; indeed,

$$\partial_t \int_0^L ds\, \mathbf{g}(\mathbf{V}, \nabla_{\mathbf{T}} \mathbf{Z}) = \int_0^L ds\, \mathbf{g}(\mathbf{F}(\mathbf{A}, \mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z}, \mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}} \mathbf{T}, \nabla_{\mathbf{Z}} \mathbf{V})), \nabla_{\mathbf{T}} \mathbf{Z})$$

$$+ \int_0^L ds\, \mathbf{g}(\mathbf{V}, \mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{T}) + \int_0^L ds\, \mathbf{g}(\mathbf{V}, \nabla_{\mathbf{Z}} \mathbf{F}(\mathbf{A}, \mathbf{g}(\mathbf{R}(\mathbf{T},\mathbf{Z})\mathbf{Z}, \mathbf{T}), \mathbf{g}(\nabla_{\mathbf{Z}} \mathbf{T}, \nabla_{\mathbf{Z}} \mathbf{T}))).$$

This involves as usual $\nabla_{\mathbf{Z}} \mathbf{A}$ which cannot be eliminated by any means and produces terms which are not total $\mathbf{Z}$ derivatives. A salient feature of this term is that the Fourier wave only depends upon the values of the fields on the endpoints but somehow, the length of the path is not included. That is,

$$\phi'(S, \mathbf{V}_S, S') = e^{i\left(\int_0^L ds\, \mathbf{g}(\mathbf{V}(\mathbf{1},\mathbf{s},\mathbf{0}), \mathbf{Z}(1,s,0)) - \int_0^L ds\, \mathbf{g}(\mathbf{V}(\mathbf{0},\mathbf{s},\mathbf{0}), \mathbf{Z}(0,s,0))\right)}.$$

This term includes nonclassical stringy effects given that, classically, the $\mathbf{V}$ and $\mathbf{Z}$ field must be perpendicular to one and another. The quantity measured here involves "stringy" variations of the $\mathbf{T}$ field and are therefore not very important from the particle perspective albeit they might indicate novel effects beyond the standard model. To rescue the first string theory, we need another contribution given that nothing else works. Considering

$$\int_0^L ds \, \frac{\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V}, \nabla_{\mathbf{Z}}\mathbf{T})}{\mathbf{g}(\mathbf{V}, \mathbf{Z})}$$

we now verify that the classical limit of the integrand makes sense. In particular,

$$\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V}, \nabla_{\mathbf{Z}}\mathbf{T}) = -\mathbf{g}(\nabla_{\mathbf{T}}\nabla_{\mathbf{Z}}\mathbf{V}, \mathbf{Z}) + \nabla_{\mathbf{T}}\nabla_{\mathbf{Z}}\mathbf{g}(\mathbf{V}, \mathbf{Z})$$

which can be further computed to be

$$\mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{V}, \nabla_{\mathbf{Z}}\mathbf{T}) = -\mathbf{g}(\mathbf{R}(\mathbf{T}, \mathbf{Z})\mathbf{V}, \mathbf{Z}) - \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{F}(\mathbf{A}, \mathbf{g}(\mathbf{R}(\mathbf{T}, \mathbf{Z})\mathbf{Z}, \mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V}))), \mathbf{Z}) + \nabla_{\mathbf{T}}\nabla_{\mathbf{Z}}\mathbf{g}(\mathbf{V}, \mathbf{Z})$$

and further

$$-\mathbf{g}(\mathbf{R}(\mathbf{T}, \mathbf{Z})\mathbf{V}, \mathbf{Z}) + \mathbf{g}(\mathbf{F}(\mathbf{A}, \mathbf{g}(\mathbf{R}(\mathbf{T}, \mathbf{Z})\mathbf{Z}, \mathbf{V}), \mathbf{g}(\nabla_{\mathbf{Z}}\mathbf{T}, \nabla_{\mathbf{Z}}\mathbf{V})), \mathbf{A}) + \nabla_{\mathbf{T}}\nabla_{\mathbf{Z}}\mathbf{g}(\mathbf{V}, \mathbf{Z})$$

leading to $\nabla_{\mathbf{T}}\nabla_{\mathbf{Z}}\mathbf{g}(\mathbf{V}, \mathbf{Z}) = 0$. Further calculation reveals,

$$-\mathbf{g}(\nabla_{\mathbf{T}}\nabla_{\mathbf{V}}\mathbf{Z}, \mathbf{Z}) = -\mathbf{g}(\nabla_{\mathbf{V}}\nabla_{\mathbf{T}}\mathbf{Z}, \mathbf{Z})$$

which leads again to

$$\nabla_{\mathbf{V}}\mathbf{g}(\nabla_{\mathbf{T}}\mathbf{Z}, \mathbf{Z}) = 0$$

assuming that $\mathbf{V}$ is integrable at $r = 0$. It may be utterly clear that this integral is not preserved in time indicating that the extensiveness of the string imposes a drag on the phase of the Fourier wave going beyond parallel propagation with a standard Lorentzian distance factor $\mathbf{g}(\mathbf{T}, \mathbf{T})$. It appears utterly clear to me that the second theory is preferred here and gives the most physical results.

13.4 The free string propagator.

The particular feature about the string propagator is that it involves an infinite dimensional integration over momentum space $\mathbf{V}_S$ and we limit in the subsequent analysis ourselves to a product manifold $\mathcal{M} \times \mathcal{N}$ where $\mathcal{M}$ is a $3 + 1$ dimensional Lorentzian base manifold endowed with a $2 + 1$ dimensional Lorentzian fibre which are each for sake of simplicity of conceptuality taken flat so that a string which is originally completely hidden in the fibre, meaning that the $\mathbf{Z}, \mathbf{A}$ field are within $\mathcal{N}$ and the $\mathbf{T}$ field has been chosen in $\mathcal{M}$ (albeit an addition of the term $\gamma(s)\mathbf{Z}$ is possible), remains so. The $\mathbf{T}$ field is as such that after parameter time one the string S specified above moves into a string S' with nontrivial projection into $\mathcal{M}$ due to s variations of the $\mathbf{T}$ field; that is, the projection of $\partial_s\mathbf{T}$ on the base is different from zero. Hence,

$$D(S, S') = \int_{\mathbf{V}_S} d\mu(\mathbf{V}_S)\delta(\mathbf{g}(\mathbf{V}_S, \mathbf{V}_S))\theta_{\mathcal{M}}(\mathbf{V}_S)\delta(\mathbf{g}(\mathbf{V}_S, \mathbf{A}))\phi(S, \mathbf{V}_S, S')$$

where we have chosen a time direction on $\mathcal{M}$ and $\theta_{\mathcal{M}}(\mathbf{V}_S)$ concerns positivity of the projection of $\mathbf{V}_S$ on that time field. The problem here regards the usual definition of the path integral as a limiting measure and dealing with the ultra violet modes. Obviously, the approach goes by means of the Fourier transform and take the limit of the modes to infinity after performing the subsequent real coefficient integrals over the null modes perpendicular to the respective modes of the $\mathbf{A}$ field. Further regularuzation of this propagator is needed and postponed for subsequent work.

Chapter 14

Conclusions.

As mentioned in the introduction, this book is an extensive report of work in progress and albeit we have obtained some novel insights in the failure of quantum field theory as well as the succes of our approach, many issues remain to be investigated. Science is a never ending story indeed and only future progress from the theoretical as well as experimental side will further clear the sky. A topic which we did not treat here regards "the" classical limit of the theory and these issues are postponed to future research. I have written this book in a self contained way and hope to have explained the importance of this topic to the interested reader. A novel book, expanding further on spirituality and its applications is currently being written.