Tools of quantum field theory over curved backgrounds

Lectures given
on the 8th-9th of January 2010
Theoretisch-Physikalisches Institut
Friedrich-Schiller Universität Jena
and
on the 29th-30th of January 2010
Institut für Theoretische Physik
Universität Leipzig

Claudio Dappiaggi

February 1, 2010

Abstract

These lectures are thought as a short introduction to the main tools of quantum field theory over curved backgrounds. The topics are thus introduced assuming that a potential reader has no previous specific knowledge on any of the used concepts except for quantum field theory over Minkowski spacetime. The main goal will be to show that, whenever the spacetime is not flat, the most effective approach to quantization is the algebraic one; to this avail, we will also consider the more common techniques and we shall show their limits of applicability. Particularly our desire is to put a reader in the condition of reading most if not all the papers dealing with concrete applications of the algebraic approach to quantum field theory over curved backgrounds without forcing him to browse through dozens of other publications in order to understand the techniques used.

Contents

Guideline to the notes

The aim of these lecture notes is to recollect and to expand the material presented during a series of lectures given at the University of Jena and at the University of Leipzig at the beginning of 2010. The overall goal is to provide a reliable and quick survey of the main mathematical and physical tools which lie at the heart of quantum field theory over curved backgrounds. The approach we shall follow is that of the algebraic formulation which stems from the seminal paper of Haag and Kastler [31] and which emphasizes the algebra structure of the observables of a physical theory and their representation on a suitably chosen Hilbert space.

My personal experience suggests that, for several reasons, this way of dealing with quantum field theory is not always very popular, being perceived as mathematically involved, difficult to learn and, worse of all, hard to connect to the real physical quantities, one wishes to measure through experiments. To a certain extent I agree that there exists a tendency to focus more on the certainly beautiful mathematical structures and the physical picture is often hidden.

Therefore I wish to set an ambitious goal: in these notes I shall try to guide the reader in the discovery of quantum field theory over curved backgrounds starting from the usual ideas which are introduced in the standard undergraduate courses of quantum field theory. I shall try to generalize these concepts without resorting to hard mathematical concepts for as long as possible and I shall also try to motivate the need for introducing every single new concept, be it of geometrical or of algebraic nature.

The final goal will be first to convince a potential reader that the algebraic formulation is necessary to overcome certain obstructions of the more common approaches and hence there is nothing to fear from it. Subsequently I aim to prove that, regardless of what it is commonly thought, the used concepts as well as the obtained results fully reproduce the standard ones and, from time to time, these are even clarified. To this avail, these notes are divided in three main parts; in chapter 1, the basic tools of differential geometry are recalled. Their introduction does not follow the standard pattern leading to Einstein's equations, and actually the goal is to show that they are an essential ingredient to define a full-fledged classical and quantum field theory over a curved background. The end point of the whole discussion will be the definition of a globally hyperbolic spacetime as the natural playground for a well-defined quantum field theory. Chapter 2 deals instead with the analysis of a classical field theory and particularly the structure of the space of solutions of the equation(s) of motion is carefully studied. A certain effort is put both in introducing the standard language of the algebraic formulation of quantum field theory and in showing that the standard concepts, one is used to, can be fully recovered, particularly with reference to the notion of classical observables. The last chapter is instead fully devoted to the description of a quantization scheme. First we shall cope with the so-called ultrastatic spacetimes where the standard picture of quantization as in Minkowski spacetime can be transported with minor efforts. The aim is to show that this operation can succeed only in a limited number of cases and, thus, a more general approach is needed. This is the algebraic formulation which is subsequently developed. A particular emphasis is put in the description of the Weyl algebra of observables and in the clarification of the notion of algebraic state. As a last topic, the Hadamard condition for algebraic states is sketched both in its microlocal and in its local form, hence allowing the readers to understand the standard notion of a physically sensible ground state in a curved background.

Chapter 1

A walk through the geometry of curved spacetimes

If one thinks about *Quantum Field Theory* (QFT), she/he is looking at one of the most successful theories developed in the 20th century whose great credit is to provide a coherent and concrete framework thanks to which one can discuss the interaction between the matter constituents and, at the same time, perform explicit computations of quantities like, for example, cross sections which have been probed via dozens of experiments. It is certainly not here the place to recall that the measurements of the various predictions, made out of theories such as quantum electrodynamics, reached an astonishing degree of precision and agreement with the theory. At the same time QFT enjoys the characteristic of admitting an extremely precise mathematical formulation and, in between its main aspects, it is important to recall that it embodies the very concepts of special relativity thus somehow encompassing in its own foundations the notion of spacetime, even though only the flat one. Nonetheless this particular aspect is also one of the main limitations of the theory since it clearly discards the role of the second major success of the former century, namely general relativity. It is well established even from an experimental point of view that the spacetime itself is not necessarily described by the Minkowski one, as suggested by special relativity, since the matter content of the Universe modifies its shape and only in certain limits one can recover as an approximation the standard picture of Minkowski spacetime.

This leads to the formulation of Einstein's theory of general relativity which teaches us that the background itself is not the usual $\mathbb{R}^4$ where distances are measured through the flat metric η, but actually one needs to allow for a more general scenario where one has (M, g).

Hence, from the point of view of QFT, we face the following scenario: we know that general relativity is around and, therefore, that the underlying spacetime is curved, no matter what. This means that we cannot expect that the ordinary formulation of QFT on $(\mathbb{R}^4, \eta)$ is suited to describe field phenomena at all scales and in all the scenarios present in the observed Universe. At the same time, the standard assertion that also the gravitational field must be quantized can certainly be agreed with, but it is an experimental fact that there are distance scales (one can think of single galaxies, clusters of galaxies or even cosmological scales) where general relativity plays an important role yielding significant modifications to the Minkowskian picture. Yet one cannot expect that, in these scenarios, quantum gravitational effects, whatever quantum gravity is, play any significant role. It is indeed in this window that QFT over curved backgrounds takes its shape and must be formulated and used to discuss physical phenomena. We shall thus peep in this realm.

Nonetheless, as promised before, we must have a clear physical and mathematical reason to introduce any single concept we shall use. To this avail I cannot help but recalling my first (or maybe it was the second) lecture in quantum field theory during my undergraduate studies when I was told that one of the main problems of quantum mechanics in the thirties was to implement Poincaré invariance, the heart of special relativity. To this avail, the first solution was developed by Klein and Gordon, who introduced the notion of a relativistic real scalar field, the simplest of all possible models. It consists of a map $\Phi : \mathbb{R}^4 \to \mathbb{R}$

whose dynamics is ruled by the Euler-Lagrange equations minimizing the following action:

$$S[\Phi] = \int\limits_{\mathbb{R}^4} d^4x \left(\eta^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi + \frac{m^2}{2}\Phi^2 \right), \tag{1.1}$$

where m is the mass of the field and, hence, it has to be taken positive. A simple variational calculation yields that

$$\begin{cases} \Box_\eta\Phi - m^2\Phi = 0 \\ \Phi(0,\vec{x}) = f, \quad \partial_t\Phi(0,\vec{x}) = g \end{cases}, \tag{1.2}$$

where $\Box_\eta$ is the d'Alembert wave operator constructed out of η, here taken in the diagonal form such that $\eta_{00} = -1$. The functions $f, g \in C_0^\infty(\mathbb{R}^3)$ are the compactly supported initial data to be assigned on an arbitrary constant time spatial surface, here taken at $t = 0$. Standard results in the theory of partial differential equations allow to prove, that, the field $\Phi \in C^\infty(\mathbb{R}^4)$.

From now on we shall take the above simple example as our guidance, namely we shall try to introduce the basic concepts which are at the heart of differential geometry as well as of general relativity and we shall show how all of them are needed to construct a counterpart of (1.1) and of (1.2) whenever the underlying spacetime is not Minkowski. This way of proceeding is for the author of these notes a sort of experiment aimed at showing that we are not forcing a theory (QFT) to be compatible to the outcome of a second one (GR). On the contrary, it turns out that basically all the main tools of general relativity arise naturally also from QFT when we try to enlarge its domain of applicability as much as possible. As in most of the cases, I fear that it is almost impossible to properly learn the tools of differential geometry, we shall need, by just reading this notes. Hence an interested reader might wish refer to more complete texts and the author personally recommends [11] which gives a pedagogical introduction to differential geometry without emphasizing the mathematical aspects, but rather the physical ones. For a more precise analysis, the author favourite book is nowadays [44], though we should certainly suggest [57] from the GR side and [39, 40] for an exclusively mathematical perspective.

1.1 Manifolds and differentiable structure: what is a smooth field?

One of the net advantages of using a global object like $\mathbb{R}^n$ with $n \in \mathbb{N}$ is that the concept of a continuous, of a smooth or of a singular function is absolutely clear and easy to use. Hence, if one inspects (1.2), there is no problem in dealing with smooth initial data or with $\Phi \in C^\infty(\mathbb{R}^4)$. What happens now if we switch on gravity and therefore we have to give up on $\mathbb{R}^4$?

While the notion of map is rather flexible, ultimately relying on sets, if one thinks of symbols like C^∞, C^0 and so on and so forth, he/she is really looking at the standard topology of $\mathbb{R}^4$, a concept which is rather manageable to use. Therefore, even though, one might be forced to give up to $\mathbb{R}^4$ globally, it would be desirable to preserve its local structure since this is all we need to define the concept of continuity, differentiability and smoothness, as well as most of the geometrical tools we shall need. Therefore one is motivated to come up with the following definition

Definition 1.1.1. *M is called a d-dimensional smooth real* **manifold** *if it is a set with a collection $\{\mathcal{O}_\alpha\}$ of subsets such that α is an index running over the (possibly infinite) natural numbers and*

- **Covering Property***: Each point $p \in M$ lies in at least one $\mathcal{O}_\alpha$,*
- **Locally** $\mathbb{R}^d$*: For each α, there exists a $1:1$ map $\psi_\alpha : \mathcal{O}_\alpha \to U_\alpha$ where $U_\alpha \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$,*
- **Local smoothness***: If $\mathcal{O}_\alpha \cap \mathcal{O}_\beta \neq \emptyset$, then $\psi_\alpha \circ \psi_\beta^{-1} : \psi_\beta(U_\alpha \cap U_\beta) \to \psi_\alpha(U_\alpha \cap U_\beta)$ is a smooth function with respect to the standard topology of $\mathbb{R}^d$. These maps are also called transition functions.*

Each ψ_α is called *chart* (mathematical terminology) or *coordinate system* (physical terminology) and we use them both throughout the text. The whole collection is instead known as *atlas*. It is important to stress that the above is a definition of manifold with minimal requirements. In this way we have a clear differentiable structure which locally mimics that of $\mathbb{R}^d$, but one can easily ask for more properties at a topological level. In particular we wish to draw the attention on the following example:

Example: The line with two origins. Let us consider now a one dimensional scenario where our candidate M is constructed as follows. Let us take two distinct copies of $\mathbb{R}$, say $\mathbb{R}_1$ ad $\mathbb{R}_2$ and let us consider the following equivalence relation: $x \sim x'$ with $x \in \mathbb{R}_1$ and $x' \in \mathbb{R}_2$ if and only if $x = x' \neq 0$. Therefore we can consider M as the union of $\mathbb{R}_1$ and $\mathbb{R}_2$ together with the said identification. The outcome is locally homeomorphic to $\mathbb{R}$ since we can construct a collection $\mathcal{O}_\alpha$ which are open sets of the real line, whenever not including the origin together with $(\mathcal{O}(0_1), \psi_1)$ and $(\mathcal{O}(O_2), \psi_2)$. These are open sets in M whose image in $\mathbb{R}$ is an open interval containing the origin. The choice can be made so that $\psi_1(\mathcal{O}(0_1)) = \psi_2(\mathcal{O}(0_2))$.

It is clear that the above scenario is hardly physical though it can be easily generalized to any dimension, for example considering the Cartesian product $M \times \mathbb{R}^n$. In order to avoid a potential pathology like the presence of a spacetime with two origins we shall require that the following property is always fulfilled:

Definition 1.1.2. *A manifold M is called* **Hausdorff (or T_2)** *if for any $p, p' \in M$, there exist $\mathcal{O} \ni p$ and $\mathcal{O}' \ni p'$ such that $\mathcal{O}, \mathcal{O}' \subset M$ and $\mathcal{O} \cap \mathcal{O}' = \emptyset$.*

We have given above an example of a pathological manifold, but it is certainly worthwhile to give also an example of a well-defined manifold where definition 1.1.1 is shown to be explicitly fulfilled. In this case there is no need to be original: we shall simply discuss the structure of a sphere.

Example: Let us consider the **2-sphere** $\mathbb{S}^2$ as embedded in $\mathbb{R}^3$, *i.e.*,

$$\mathbb{S}^2 = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 \mid \sum_{i=1}^{3} x_i^2 = 3 \right\}.$$

This is a manifold since, if we take the point $p \in \mathbb{S}^2$ of coordinate $(0, 0, 1)$ which we call north-pole, we can define the open set $\mathcal{O}_1 \doteq \mathbb{S}^2 \setminus \{p\}$ together with $\psi_1 : \mathcal{O}_1 \to U_1 \subseteq \mathbb{R}^2$ such that

$$\psi_1(x_1, x_2, x_3) = (y_1, y_2) = \left(\frac{2x_1}{1 - x_3}, \frac{2x_2}{1 - x_3} \right).$$

Since we are covering with $\mathcal{O}_1$ the whole 2-sphere except a point, we can just consider $p' \in \mathbb{S}^2$ of coordinate $(0, 0, -1)$. Then a second open set is $\mathcal{O}_2 \doteq \mathbb{S}^2 \setminus \{p'\}$ together with $\psi_2 : \mathcal{O}_1 \to U_2 \subseteq \mathbb{R}^2$ such that

$$\psi_2(x_1, x_2, x_3) = (z_1, z_2) = \left(\frac{2x_1}{1 + x_3}, \frac{2x_2}{1 + x_3} \right).$$

Per construction the first two hypotheses of definition 1.1.1 are fulfilled, hence we just need to check the third one. A direct computation shows that $\mathcal{O}_1 \cap \mathcal{O}_2 = \mathbb{S}^2 \setminus \{p \cup p'\}$ and that the transition functions are given by the following relations

$$\begin{cases} z_1 = \frac{4y_1}{y_1^2 + y_2^2} \\ z_2 = \frac{4y_2}{y_1^2 + y_2^2} \end{cases},$$

which are clearly smooth functions over $\mathbb{R}^2$. Hence $\mathbb{S}^2$ is a manifold and it is also Hausdorff since $\mathbb{R}^n$ is such, $\mathbb{R}^3$ in particular.

Consequence 1.1.0. We have learned the notion of an Hausdorff manifold (for short, henceforth, just manifold) and it has been built to look like locally as $\mathbb{R}^n$. Therefore, given any two manifolds M and N, we call **smooth functions** from M to N, or $C^\infty(M; N)$ the set of functions $f : M \to N$ such that, for any chart

$\{\psi_\alpha\}$ of M and $\{\psi'_\beta\}$ of N, then, calling $\circ$ the composition of functions, $\psi'_\beta \circ f \circ \psi_\alpha^{-1} : U_\alpha \subseteq \mathbb{R}^n \to U'_\beta \subseteq \mathbb{R}^d$ is smooth with respect to the standard topology of $\mathbb{R}^n$. If $N \equiv \mathbb{R}$ we shall omit it and simply write $C^\infty(M)$. In terms of (1.2) we have understood how to make sense of a smooth map, hence of smooth fields and initial conditions, whenever we are not dealing with $\mathbb{R}^n$, but, rather, with a generic manifold M.

In between the set of all smooth functions between two manifolds, of particular mathematical and physical relevance are the following:

Definition 1.1.3. *We call* **smooth diffeomorphism** *a bijective map* $f \in C^\infty(M; N)$ *such that* $f^{-1} \in C^\infty(N; M)$.

1.1.1 Vectors and Tangent space: what is the derivative of a field?

If one gives a closer look at (1.1) and (1.2), she/he can realize that the notion of smoothness of a field or of an initial datum is a prerequisite in order to understand whether the object we cope with is differentiable and hence if we can actually perform a derivative, a basic operation whenever we want to discuss the dynamics of a physical system. Nonetheless this does not suffice since the action of the Klein-Gordon field in Minkowski spacetime also involves the concept of direction and of directional derivative, a tool which is of immediate definition whenever we cope with functions on $\mathbb{R}^n$. Let us now see how can we make sense of a similar concept on a generic manifold.

Even in this case, the general strategy will be to exploit that the local geometry of a manifold M coincides with that of $\mathbb{R}^n$, where n is the dimension of M. As a matter of fact if we consider $\mathbb{R}^n$ endowed with the standard Cartesian coordinate system, there is a direct correspondence between n-dimensional vectors $\vec{v} = (v_1, ..., v_n)$ and directional derivatives acting on smooth functions $C^\infty(\mathbb{R}^n)$, namely

$$v \mapsto D_v \doteq \sum_{i=1}^{n} v_i \frac{\partial}{\partial x_i}.$$

On an arbitrary background we can almost slavishly follow the same procedure, *i.e.*,

Definition 1.1.4. *For a generic manifold M, a vector v at $p \in M$ is a map*[1] $v : C^\infty(M) \to \mathbb{R}$ *which fulfils the following two properties*

- **Linearity:** $v(af + bg) = av(f) + bv(g)$ *for all* $f, g \in C^\infty(M)$ *and for all* $a, b \in \mathbb{R}$,
- **Leibniz rule:** $v(fg) = f(p)v(g) + v(f)g(p)$ *for all* $f, g \in C^\infty(M)$.

The set of all vectors at a point $p \in M$ is called the **tangent space** T_pM *of M at p.*

An important consequence of this definition of tangent space is the following (for more details and for a similar approach, refer to [5]).

Proposition 1.1.1. *Let M and N be manifolds and $F \in C^\infty(M; N)$. Then F induces a natural homomorphism $F_* : T_pM \to T_{F(p)}N$, called* **push-forward**, *such that*

$$(F_*v)(f) \doteq v(f \circ F). \quad \forall f \in C^\infty(N), \text{ and } \forall v \in T_pM.$$

Proof. We need to verify that F_*v is indeed a vector at $F(p)$ according to definition 1.1.4. Let us start with linearity, the simplest condition. Let $f, g \in C^\infty(N)$ and let $a, b \in \mathbb{R}$, then

$$(F_*v)(af + bg) = v[(af + bg) \circ F] = v[af \circ F + bg \circ F] = av(f \circ F) + bv(g \circ F) = a(F_*f)(v) + b(F_*g)(v),$$

[1] To be precise, one should not use the set of smooth functions over M, but actually the set of smooth functions whose domain of definition includes an open set containing p. This set must be endowed with an equivalence relation where $f \sim g$ if they agree on any neighbourhood of p. The resulting objects are sometimes indicated as **germs** of C^∞ functions and the whole set as $C^\infty(p)$.

where, in the various equality we exploited only the definition of F_* and the linearity property of v. Let us now show the Leibniz rule; let us consider $f, g \in C^\infty(N)$, then

$$(F_*v)(fg) = v(f \circ Fg \circ F) = (f \circ F)(p)v(g \circ F) + v(f \circ F)(g \circ F)(p) = f(F(p))(F_*v)(g) + (F_*v)(f)g(F(p)),$$

where, besides the definition of F_*, we used that the product fg is meant pointwise, that is $(fg)(x) = f(x)g(x)$ for all $x \in N$. $\square$

Consequence 1.1.0. The definition 1.1.4 and the above proposition yield that T_pM is a vector space and to any open (coordinate) neighbourhood $\mathcal{O} \subseteq M$, there corresponds a natural basis e_μ with $\mu = 1, ..., n$ for every $p \in \mathcal{O}$. Hence $\dim T_pM = \dim M = n$. To be precise, if the local map $\psi : \mathcal{O} \to U \subseteq \mathbb{R}^n$ is assigned, for every $v \in T_pM$ and for every $f \in C^\infty(M)$, we can write

$$v(f) = \sum_{\mu=1}^{n} v^\mu e_\mu(f) = \sum_{\mu=1}^{n} v^\mu(\psi_*^{-1}\partial_\mu)(f) = \sum_{\mu=1}^{n} v^\mu \left.\frac{\partial(f \circ \psi^{-1})}{\partial x_\mu}\right|_{\psi(p)},$$

where e_μ is defined as the pull-back via ψ of the vector $\frac{\partial}{\partial x_\mu}$ at $\psi(p)$, once a coordinate system x^μ has been assigned to $\mathbb{R}^n$. In particular we have understood how to define **the notion of (directional) derivative at a point** $p \in M$ of a smooth function and, thus, of a real scalar field. We can thus make sense of $\partial_\mu\Phi$ at any but fixed point point in a curved manifold.

Yet this cannot suffice since, on the one hand, we want to have a global and not a pointwise notion of derivative and, on the other hand, the given definition applies to scalar fields, while, in general, we have to keep in mind that we might have to deal with additional structures such as, for example, with vector or spinor fields (Dirac, Majorana, Proca and Maxwell fields in the language of Minkowski spacetime). In this case the standard directional derivative will be superseded by the covariant one, which we will later introduce.

Before concluding the section, let us discuss a few further related concepts which we might later mention.

Definition 1.1.5. *We call* **cotangent space** *of M at p, T_p^*M, the dual space of T_pM.*

Since the tangent space of M at p is a finite dimensional vector space, we can introduce the dual pairing $(,)$ between T_pM and T_p^*M to conclude that also the latter is a vector space and that $\dim T_p^*M = \dim T_pM = \dim M$. Hence, whenever we assign a base X_μ of T_pM with $\mu = 1, ..., \dim M$ then we can automatically define a base $X^{*\mu}$ of T_p^*M out of $(,)$ as $(X^{*\mu}, X_\nu) = \delta^\mu_\nu$.

As a last step, we have to "glue" together the information of all the (co)-tangent spaces and, to this avail, we call

- **Tangent space** $TM \doteq \bigcup_{p\in M} T_pM$,
- **Cotangent space** $T^*M \doteq \bigcup_{p\in M} T_p^*M$.

Accordingly we can now continuously assign a vector at each point $p \in M$ and the result is called *vector field* X, namely $X : M \to TM$. Thanks to the discussion of the previous section, we know that we can also consistently require that X is smooth and the set of all these vector fields is either indicated in the literature as $\Gamma(M, TM)$ or as $\mathfrak{X}(M)$. In the first case one usually stresses the nature of smooth vector fields as sections of TM, thought as a bundle over the manifold, while the second symbol stems from the more traditional literature in differential geometry and general relativity. Notice that, when dealing with a generic vector field $v \in \mathfrak{X}(M)$, we can still refer to its components v^μ and this entails that we are actually constructing them pointwisely and that, for each fixed μ, $v^\mu \in C^\infty(M)$.

1.1.2 The metric structure: what plays the role of $\eta^{\mu\nu}$?

One of the main advantages of using Minkowski spacetime in relativistic theories is that it comes endowed with a metric, namely an inner product over $T_p\mathbb{R}^4 \sim \mathbb{R}^4$, which enables us to measure distances and angles, to define volume forms and to associate covectors, elements of $T_p^*\mathbb{R}^4$, to vectors at p. This operation is symbolically known as "the raise of indexes", that is for every $v \in T_pM$ whose components are v^μ, we associate the components of an element in $T_p^*\mathbb{R}^4$ as $v_\mu = \eta^{\mu\nu}v_\nu$.

We seek now to extend this concept to a generic manifold M. To this avail we need to go one step beyond the notion of vectors and covectors.

Definition 1.1.6. *A* **tensor** *of type* (k,l) *is a multilinear map*

$$\tau : \underbrace{T_p^*M \times ... \times T_p^*M}_{k} \times \underbrace{T_pM \times ... \times T_pM}_{l} \to \mathbb{R}.$$

The set of all the tensors of fixed type is a vector space $T(k,l)$ *whose generic element* τ *can be expanded as*

$$\tau = \tau^{\mu_1...\mu_k}_{\nu_1...\nu_l} X_{\mu_1}...X_{\mu_k} X^{*\nu_1}...X^{*\nu_l},$$

where X_μ *is a base of* T_pM.
Alternatively τ *can be interpreted as an element of* $\underbrace{T_pM \times ... \times T_pM}_{k} \times \underbrace{T_p^*M \times T_p^*M}_{l}$. *The coefficients* $T^{\mu_1...\mu_k}_{\nu_1...\nu_l}$ *are usually called tensor components.*

On the set of tensors, one can define natural operations:

- **Contraction** $C : T(k,l) \to T(k-1, l-1)$, such that for every $\tau \in T(k,l)$ with $k \geq 1$ and $l \geq 1$, the tensor components of $C[T]$ read
$$C[T] = T^{\mu_1...\sigma...\mu_k}_{\nu_1...\sigma...\nu_l},$$
where the upper and lower σ are located at the fixed i-th and j-th entry.
- **Outer Tensor Product** $\boxtimes : T(k,l) \times T(k',l') \to T(k+k', l+l')$ such that for every $\tau \in T(k,l)$ and $\tau' \in T(k',l')$, the tensor $\tau \boxtimes \tau'$ has components
$$(\tau \boxtimes \tau')^{\mu_1...\mu_{k+k'}}_{\nu_1...\nu_{l+l'}} \doteq \tau^{\mu_1...\mu_k}_{\nu_1...\nu_l} \tau'^{\mu_{k+1}...\mu_{k+k'}}_{\nu_{l+1}...\nu_{l+l'}}.$$

It is often customary to work with coordinates and, hence, to switch from the abstract basis e_μ to $\partial_\mu \equiv \frac{\partial}{\partial x_\mu}$ which induces a specific e_μ along the lines of consequence 1.1.2. If we choose a different coordinate system, say $\psi'_\mu \equiv x'_\mu$ in place of $\psi_\mu \equiv x_\mu$, then the tensor components transform as follows:

- *tensors of type* $(1,0)$: for any $v \in T_pM$, then $v^\mu \to v^{\mu'} = \frac{\partial x'^{\mu'}}{\partial x^\mu} v^\mu$
- *tensors of type* $(0,1)$: for any $\omega \in T_p^*M$, then $\omega_\mu \to \omega_{\mu'} = \frac{\partial x^\mu}{\partial x'^{\mu'}} \omega_\mu$
- *tensors of type* (k,l): for any $\tau \in T(k,l)$, then
$$\tau^{\mu_1...\mu_k}_{\nu_1...\nu_l} \to \tau^{\mu'_1...\mu'_k}_{\nu'_1...\nu'_l} = \frac{\partial x'^{\mu'_1}}{\partial x^\mu_1} ... \frac{\partial x'^{\mu'_k}}{\partial x^{\mu_k}} \frac{\partial x^{\nu_1}}{\partial x'^{\nu'_1}} ... \frac{\partial x^{\mu_l}}{\partial x'^{\mu'_l}} \tau^{\mu_1...\mu_k}_{\nu_1...\nu_l}.$$

As for vectors, we can give a continuous assignment of a tensor at each point $p \in M$ and this yields a **tensor field** of type (k,l), a multilinear map $T : \underbrace{T^*M \times ... \times T^*M}_{k} \times \underbrace{TM \times ... \times TM}_{l} \to \mathbb{R}$.

Definition 1.1.7. *We call* **metric (tensor field)** *a multilinear map* $g : TM \times TM \to \mathbb{R}$ *such that*

- *it is symmetric, that is* $g(v, v') = g(v', v)$ *for all* $v, v' \in \mathfrak{X}(M)$*,*
- *it is non degenerate, that is* $g(v, v') = 0$ *for all* $v \in \mathfrak{X}(M)$ *implies* $v' = 0$

In a local chart, we can write $g = g_{\mu\nu} dx^\mu \otimes dx^\nu$ *or, equivalently,* $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ *and the components* $g_{\mu\nu}$ *are called* **metric coefficients***.*

This is a rather general definition of metric and it is possible to impose further requirements, such as, to quote the most relevant, smoothness. Yet, even with definition 1.1.7, it is possible to draw a few interesting consequences. First of all, since the vector space T_pM is now endowed with an inner product, it is always possible to choose its basis e_μ in such a way that

$$g(e_\mu, e_\nu) = 0 \;\; \forall \mu \neq \nu \quad \text{and} \quad g(e_\mu, e_\mu) = \pm 1 \;\; \forall \mu = 1, ... \dim T_pM.$$

The set of all "plus" and "minus" signs defines the **signature** of the metric of M at p and we distinguish between two important cases, here depicted in four dimensions:

- **Riemannian** $(+, +, +, +)$ or $(-, -, -, -)$ associated to Euclidean QFT
- **Lorentzian** $(-, +, +, +)$ or $(+, -, -, -)$ associated to standard QFT

Notice that, up to now, there is no reason to claim that the signature is the same across the whole manifold. As an example consider $\mathbb{R}^4$ endowed with the standard Cartesian coordinates (t, x, y, z). Then we can associate to it the following metric

$$ds^2 = \Theta(t) dt^2 + dx^2 + dy^2 + dz^2,$$

where $\Theta(t) = 1$ for $t \geq 0$ and -1 otherwise. It is clear that this example is not physical since the metric coefficients are not even continuous, yet it is not forbidden by our definitions. Hence:

Assumption: We require that the metric tensor field is smooth and of constant signature over the whole M.

Consequence 1.1.0. We learned that it is possible to associate a counterpart of $\eta_{\mu\nu}$ on a manifold in terms of the components of a smooth non-degenerate and symmetric tensor field of type $(0, 2)$. In this way we can give meaning to $g^{\mu\nu}$ and to the operation of "raising and lowering indices" as, for example, $v^\nu = g^{\mu\nu} v_\mu$ for all $v \in T^*_pM$ (and, of course, viceversa). Furthermore the notion of metric yields also a natural expression for a volume form to be used in place of d^4x in (1.1), that is, in the domain of definition of a coordinate system, $d^4x\sqrt{|g|}$ is the **volume form associate to** (M, g).

It is important to stress that the above discussion does not automatically allow us to define an integral of a function on a curved background, since we still need to discuss the domain of integration. It would be tempting to just replace $\mathbb{R}^4$ with M. Unfortunately, also due to the generality of our definition of M, this is not always possible. We sketch here how the correct procedure can be implemented:

Definition 1.1.8. *We say that*

- *a subset* U *of* $\mathbb{R}^n$ *has* **(Jordan) content zero** $c(U) = 0$ *if* $\forall \epsilon > 0$*, there exists a finite collection of n-cubes* $C_1, ... C_s$ *covering* U *and with volumes such that* $\sum_{i=1}^{s} C_i < \epsilon$*.*
- *for any manifold* M*, a subset* $\mathcal{O} \subseteq M$ *is said to have content zero* $c(\mathcal{O}) = 0$ *if it is contained in the finite union of compact subsets* A_i *each of which is contained in a coordinate neighbourhood* $\psi_i : \mathcal{O}_i \to U_i \subseteq \mathbb{R}^n$ *such that* $A_i \subset \mathcal{O}_i$ *and* $c(\psi_i(\mathcal{O}_i)) = 0$

- *for any manifold M, D is a* **domain of integration** *if it is relatively compact*[2] *and ∂D, the boundary of D has content zero $c(\partial D) = 0$.*

Notice that one can prove that, if D is a domain of integration, so is its closure and its interior. Furthermore the property is preserved under the operation of finite unions, finite intersections or action of a diffeomorphism. It is certainly possible that M itself is a domain of integration and, from now on, we shall always write $\int\limits_M d^4x\sqrt{|g|}$ although we implicitly think that we are considering a suitable domain of integration $D \subseteq M$.

It is interesting to remark that the outcome of our discussion is that, whenever a manifold M together with a smooth metric g is assigned, we have all the ingredients to define the natural counterpart of (1.1) in a curved background. We shall not do it now since we prefer to give also all the ingredients to discuss the counterpart of the Cauchy problem (1.2) before drawing any conclusion.

1.1.3 The covariant derivative: how do we globally derive tensors?

In section 1.1.1 we learned how it is possible to construct directional derivatives of smooth functions at each point $p \in M$. While in Minkowski spacetime, translational invariance allows to show that the definition is actually independent from the chosen point p, this has no natural counterpart in a curved manifold M with metric g.

Hence our goal is to seek a definition of a derivative which is tied to the geometrical structure of M, it reduces to ∂_μ in $(\mathbb{R}^n, \eta)$ and pointwise in M when acting on scalar smooth functions. In other words, we are looking for the following object:

Definition 1.1.9. *A* **covariant derivative** *in the direction of a vector field $v \in \mathfrak{X}(M)$ is an operator $\nabla_v : T(k,l) \to T(k, l+1)$ which satisfies the following requirements:*

- linearity*, i.e., for all $\tau, \tau' \in T(k,l)$ and for all $\alpha, \beta \in \mathbb{R}$,*

$$\nabla_v(\alpha\tau + \beta\tau') = \alpha\nabla_v\tau + \beta\nabla_v\tau',$$

- *the* Leibniz rule*, i.e., for all $\tau \in T(k,l)$ and $\tau' \in T(k', l')$,*

$$\nabla_v(\tau \boxtimes \tau') = \nabla_v\tau \boxtimes \tau' + \tau \boxtimes \nabla_v\tau',$$

- *the* compatibility with contraction*, i.e., for all $\tau \in T(k,l)$ with $k, l \geq 1$*

$$\nabla_v[C(\tau)] = C(\nabla_v\tau),$$

- consistency*, i.e., for $\tau \in T(0,1)$ and $f \in C^\infty(M)$*

$$\tau(f) = \tau^\mu\nabla_\mu(f) = t^\mu\partial_\mu(f).$$

In addition to the above properties it is also customary to require that ∇ is **torsion free***, i.e., for any $v, v' \in \mathfrak{X}(M)$,*

$$T(v, v') \doteq \nabla_v v' - \nabla_{v'} v - [v, v'] = 0,$$

where $[v, v']$ is the Lie-bracket between v and v', namely a vector field whose coefficients are pointwisely

$$[v, v']^\mu = v^\nu\partial_\nu(v')^\mu - (v')^\mu\partial_\mu(v).$$

[2] We recall that a subset $\mathcal{O}$ of a topological space M is called *relatively compact* if its closure with respect to the topology of X is compact.

Notice that the vanishing of the torsion tensor implies that, at a level of components and for all $f \in C^\infty(M)$, it holds

$$\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f - T^\rho_{\mu\nu} \nabla_\rho f = \nabla_\nu \nabla_\mu f.$$

The natural questions arising from definition 1.1.9 concern the existence and uniqueness of the covariant derivative. We shall not analyse this problem in detail, but a few comments are certainly necessary. To start with, it is immediate to notice that the directional derivative of section 1.1.1 fulfils all the above requirements and, thus, although existence is guaranteed, uniqueness seems impossible to achieve. As a matter of fact, if we take on (M, g) any two covariant derivatives, $\nabla, \widetilde{\nabla}$, it can be easily shown that the above definition entails, that for any $\omega \in T^*M$,

$$\nabla_\mu \omega_\nu = \widetilde{\nabla}_\mu \omega_\nu - C^\rho_{\mu\nu} \omega_\rho \quad \text{and} \quad C^\rho_{\mu\nu} = C^\rho_{\nu\mu}.$$

Notice that the first difference between two covariant derivatives can be seen only at a level of vectors and covectors since they must coincide on scalar smooth functions. Furthermore if $\widetilde{\nabla}_\mu = \partial_\mu$, then the coefficient $C^\rho_{\mu\nu}$ are called **Christoffel symbols** and they are often indicated as $\Gamma^\rho_{\mu\nu}$. Actually, on a manifold M endowed with a metric g, it is possible to provide an explicit expression for these coefficients provided a new concept is first introduced:

Definition 1.1.10. *Let $\gamma : I \subseteq \mathbb{R} \to M$ be a smooth curve with tangent vector t. We say that a vector v is parallel-transported along γ if and only if on γ it holds*

$$t^\mu \nabla_\mu v^\nu = 0.$$

Let us now consider two arbitrary vectors, v, w which are parallel-transported along a curve γ. If we require that also their inner product $g(v, w)$ is conserved along the curve, this yields

$$\nabla_t g(v, w) = (\nabla_t g)(v, w) + g(\nabla_t v, w) + g(v, \nabla_t w) = (\nabla_t g)(v, w) = 0,$$

where we used definition 1.1.10 in the second equality. Therefore, if we want that this last identity holds true for all possible smooth curves, it means that, at a level of coefficients,

$$\nabla_\rho g_{\mu\nu} = 0, \quad \forall \rho, \mu, \nu.$$

In this case it is said that ∇ is *metric compatible* and, in turn, this implies that the Christoffel symbols assume a specific form, namely (see theorem 3.1.1 in [57]):

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\delta} (\partial_\mu g_{\nu\delta} + \partial_\nu g_{\mu\delta} - \partial_\delta g_{\mu\nu}). \tag{1.3}$$

Consequence 1.1.0. We learned that it is possible to introduce a notion of covariant derivative acting on a generic tensor field. Furthermore if one imposes the additional requirement that such derivative must be torsion free and compatible with the metric, the Christoffel symbols assume the form (1.3) and we say that ∇ is the **Levi-Civita connection**. Furthermore the fundamental theorem of differential geometry states that this connection is unique. We have thus learned how to derive all kind of fields on a manifold M with metric g.

Consequence 1.1.0. Although in the construction of a Klein-Gordon action on a manifold M, the derivative operator coincides with the partial derivative since it acts on scalar functions, a substantial difference arises in the construction of the counterpart of the d'Alembert wave operator on M with metric g. As a matter of fact, if we keep in mind that in Minkowski spacetime $\Box_\eta f = \eta^{\mu\nu} \partial_\mu \partial_\nu f$ and that for all $f \in C^\infty(M)$, $\partial_\mu f$ are the coefficients of an element in $T(0, 1)$, we can define the **d'Alembert wave operator** as

$$\Box_g f \doteq g^{\mu\nu} \nabla_\mu \partial_\nu f = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} g^{\mu\nu} \partial_\nu) f, \quad \forall f \in C^\infty(M) \tag{1.4}$$

where the second identity derives from the explicit form of the Levi-Civita connection via (1.3).

1.1.4 Notable tensors

In the previous sections we constructed the main geometrical ingredient on a manifold M, namely the metric tensor field. Yet, it is possible to go one step further introducing further tensors which characterize the geometry of M; these are the building block in general relativity to construct the Einstein's tensor and equations, but they play a key role also in quantum field theory over curved backgrounds. As a matter of fact, as we shall comment at the beginning of next section, the absence in general of a large isometry group such as the Poincaré in Minkowski spacetime, implies that there is no natural way to unambiguously construct the equation(s) of motion of a free field. Hence the presence of new terms is indeed possible and they are all proportional to the (components of the) tensors we shall now introduce. Notice that all these new quantities vanish per construction if we work on Minkowski spacetime. We shall present all these new objects in terms of components since it is how we are going to use them, though they can be easily generalized.

• **the Riemann tensor** $R^{\delta}_{\mu\nu\rho}$ which is a map from $T(0,1)$ to $T(0,3)$, namely for every $\omega \in T^*_p M$,

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\, \omega_\rho = R^{\delta}_{\mu\nu\rho} \omega_\delta,$$

where ∇ is the Levi-Civita connection. Notice that the definition implies the following symmetries:

$$\begin{cases} R^{\delta}_{\mu\nu\rho} = R^{\delta}_{\nu\mu\rho} \\ R^{\delta}_{[\mu\nu\rho]} = 0 \\ R_{\mu\nu\rho\delta} = -R_{\rho\delta\mu\nu} \end{cases}, \tag{1.5}$$

where the symbol $[,]$ stands for total antisymmetrization of the enclosed indexes. Be aware that the last identity holds true only thanks to the requirement of metric compatibility. Furthermore a direct computation shows that the so-called *Bianchi identity* holds true:

$$\nabla_{[\gamma} R^{\delta}_{\mu\nu]\rho} = 0. \tag{1.6}$$

• **the Ricci tensor** $R_{\mu\nu}$ which is an element in $T(0,2)$, namely

$$R_{\mu\rho} \doteq R^{\delta}_{\mu\nu\rho} g^{\nu}_{\delta}. \tag{1.7}$$

Notice that

$$R_{\mu\rho} = R_{\rho\mu}.$$

• **the Ricci scalar / the scalar curvature** R which is a scalar function, that is an element of $T(0,0)$ such that

$$R \doteq R_{\mu\rho} g^{\mu\rho}. \tag{1.8}$$

We could infer many interesting properties of M from the above new objects, but this would bring us far from our goals. Yet, we wish at least to notice that, if one seeks a tensor field of type $(0,2)$ which is covariantly constant, then there are only two possibilities using the metric, namely g itself and the **Einstein's tensor** whose components are

$$G_{\mu\nu} = R_{\mu\nu} - \frac{g_{\mu\nu}}{2} R. \tag{1.9}$$

1.2 Causal structures and global hyperbolicity

The picture that emerges from the standard description of QFT over Minkowski spacetime is that a notion of causality and of causal structures emerges naturally thanks to the simple form of the flat metric η. In this way it is possible to understand how to implement the notion of "cause and effect" in the underlying theory and how to analyse it out of geometric quantities. Let us briefly review it: as a starting point, we consider $\mathbb{R}^4$ and an arbitrary point therein whose coordinates can be chosen as the origin of a Cartesian reference frame. Translational invariance implies that we are not losing of generality.

We can thus define the *light cone* with tip p as the set of points $p' \in \mathbb{R}^4$ of coordinates $(t, \vec{x}) = (t, x, y, z)$ such that $t^2 = |\vec{x}|^2$ where $|\cdot|$ is the Euclidean modulus. In other words p' and p are called **light-like separated** since they are joined by a curve whose tangent vector t satisfies at each point the identity $\eta^{\mu\nu} t_\mu t_\nu = 0$. The standard picture of special relativity tells us that the points causally connected to p are those lying within or on the boundary of the light cone. The former can be similarly characterized as being the set of points **timelike separated** from p, that is they are connected to p by a curve whose tangent vector t' is such that, at all points, $\eta_{\mu\nu} t'^\mu t'^\nu < 0$. Notice that we assume $\eta_{00} = -1$. To conclude, the points which are lying outside the light cone are said to be **spacelike separated** from p since, there exists a curve connecting them to p and whose tangent vector $\widetilde{t}$ is such that $\eta_{\mu\nu} \widetilde{t}^\mu \widetilde{t}^\nu > 0$ pointwisely.

The generalization to a curved backgrounds of these concepts is rather straightforward since the manifolds we choose are always endowed with a smooth non-degenerate metric g. Hence

Definition 1.2.1. *Given a manifold M with a smooth non-degenerate metric g of Lorentzian signature $(-, +, +, +)$, we say that a vector $v \in T_pM$ is*

- **timelike** *if and only if $g_{\mu\nu}|_p v^\mu v^\nu < 0$,*
- **lightlike** *if and only if $g_{\mu\nu}|_p v^\mu v^\nu = 0$,*
- **spacelike** *if and only if $g_{\mu\nu}|_p v^\mu v^\nu > 0$.*

Notice that ultimately we are adopting the same definition as in Minkowski spacetime, since at each point $p \in M$, it is always possible to find a coordinate system such that the Christoffel symbols (1.3) vanish identically at p and hence the metric becomes $\eta_{\mu\nu}$ exactly there. This special chart is also known as **normal coordinates** and we shall not dwell into this topic leaving an interested reader to definition 6.7 in [5] for the Euclidean scenario, the Lorentzian following suit. If we now try to generalize this concept to curves, we have to pay attention since we cannot just choose any possible smooth map $\gamma : I \subseteq \mathbb{R} \to M$, but we rather have to look for the one which is "the shortest one" between two given points in M. This leads to the following definition:

Definition 1.2.2. *Given a manifold M with a smooth metric g, a* **geodesic** *$\gamma : I \subseteq \mathbb{R} \to M$ is a smooth curve with affine parameter called λ and whose tangent vector T satisfies*

$$T^\mu \nabla_\mu T^\nu = \frac{dT^\nu}{d\lambda} + \Gamma^\nu_{\mu\rho} T^\mu T^\rho = 0. \tag{1.10}$$

In other words T is parallel transported along γ.

A further property of the light-cone in Minkowski spacetime is the possibility to establish whether a timelike vector $v \in T_p\mathbb{R}^4$ is future or past directed depending if it is contained in the upper or lower light-cone which originates from p. While, in this case, translational invariance guarantees that the concept can be made independent from the choice of p, in a generic manifold M with a smooth metric g the above concept can be slavishly adopted only pointwisely. Whenever we wish to vary the base point of the light cone and a continuous choice of an "upper" and "lower" light cone can be made, then we say that M is **time orientable**. We shall give later a characterization of a physically relevant class of manifolds where such notion plays a key role. We can now generalize all the above concepts at the level of curves:

Definition 1.2.3. *Let M be a smooth time-orientable manifold endowed with a smooth metric g. Then we call*

- **chronological future** *of $p \in M$, the set $I^+(p)$ of points $p' \in M$ such that there exists $\gamma : I \subseteq \mathbb{R} \to M$ connecting p and p' and such that the tangent vector of γ is everywhere timelike and future directed.*

- **causal future** *of* $p \in M$, *the set* $J^+(p)$ *of points* $p' \in M$ *such that there exists* $\gamma : I \subseteq \mathbb{R} \to M$ *connecting* p *and* p' *and such that the tangent vector of* γ *is everywhere either timelike or lightlike and future directed.*

An analogous definition holds for the chronological and causal past of p, *indicated as* $I^-(p)$ *and* $J^-(p)$ *respectively.*

Consequence 1.2.0. Definition 1.2.3 tells us that the geometric interpretation of causally connected points in a generic manifold M with a smooth metric g is the same as in Minkowski spacetime, namely events in p are influenced or influence the set $J^-(p) \cup J^+(p)$. Since in QFT observables are not pointwise objects, but they are rather localised in open sets $\mathcal{O} \subseteq M$, we can say that two regions $\mathcal{O}$ and $\mathcal{O}'$ are causally separated if and only if[3] $\mathcal{O}' \cap (J^-(\mathcal{O}) \cup J^+(\mathcal{O})) = \emptyset$ where $J^{\pm}(\mathcal{O}) = \bigcup\limits_{p \in \mathcal{O}} J^{\pm}(p)$.

1.2.1 Globally hyperbolic spacetimes: how to set-up a Cauchy problem?

In the previous analysis we have shown that it possible to generalize the notion of Minkowski spacetime introducing the concept of Hausdorff time-orientable manifold endowed with a smooth metric g. This operation allows us, on the one hand, to introduce a natural counterpart of a Klein-Gordon action on M and, on the other hand, to discuss important geometric-related concepts such as causally connected regions. Nonetheless, we are falling one step short of our final goal since, even if we construct a Lagrangian and we derive its Euler-Lagrange equations, we still need to identify an initial surface where to assign initial data out of which we can discuss existence and uniqueness of the solution[4]. In Minkowski spacetime, the natural choice are the constant time hypersurfaces which are three dimensional spacelike smooth manifolds. In a curved background the situation is far more complicated since there is a priori no reason to be able to select a foliation of the manifold in smooth spacelike hypersurface. As a matter of fact, in general the statement is that it is impossible; yet it is possible to select a distinguished class of manifolds where the general obstruction is circumvented and these are called *globally hyperbolic*, the natural candidates where a physically sensible classical and quantum (scalar) field theory can be set up[5].

As a starting point we wish to recollect the results of the previous sections as follows:

Definition 1.2.4. *A* **spacetime** (M, g) *is a four dimensional, smooth, Hausdorff, time-orientable, (arcwise) connected manifold* M *endowed with a smooth Lorentzian metric* g *of signature* $(-,+,+,+)$.

For a reader interested in the differential geometric properties of these manifolds, we stress that the above assumptions suffice to prove that M is also paracompact and second-countable (see [57] for a definition) as first shown by Geroch [28, 29]. Before continuing the above analysis, it is worthwhile to discuss a short example which shows why the above definition of spacetime is still incomplete and it leads to physically unacceptable scenarios.

Example: let us consider a spacetime (M, g) which is of "Gödel" type (see §5.7 in [32] though the metric

[3]Note that this condition is often stated also as follows: (the points of) $\mathcal{O}$ and $\mathcal{O}'$ are spacelike separated. This is true only when it is possible to prove that if two points are not causally related, they must be forcefully joined by a spacelike curve. This is certainly the case in Minkowski, but not for a generic manifold M with a smooth metric g, even when M is connected.

[4]In these notes we shall only consider Cauchy problems. It is indeed possible, for a given specific partial differential equation on a manifold M to discuss other ways to get a solution, such as for example the Goursat problem. In this case only one initial datum is needed on the characteristic surface of the chosen operator such as the light cone for the d'Alembert wave operator (a short survey can be found either in [12] or, with a more modern perspective, in [25]). We shall not discuss conditions of M to set up these problems, but the reader should keep in mind that such a possibility is available.

[5]A priori our discussion applies only to scalar fields. Whenever further structures, such as a spin structure are present, further restrictions might be present. It turns out that in four spacetime dimensions, this is not the case, but, in general, a reader is warned to pay attention if she/he does not want to treat only scalar fields.

therein is slightly different), namely M is topologically $\mathbb{R}^3 \times \mathbb{S}^1$ and the metric reads:

$$ds^2 = -dt^2 + dr^2 - \alpha r^2 dt d\varphi + (r^2 - \frac{\alpha^2}{4} r^4) d\varphi^2 + dz^2,$$

where $t, z \in \mathbb{R}$, $r > 0$ is a radial coordinate, $\varphi \in [0, 2\pi)$ an angular one while $\alpha \in \mathbb{R}$. Let us now consider the vector field v which pointwisely coincide with ∂_φ; then we can compute

$$g(v, v) = g(\partial_\varphi, \partial_\varphi) = r^2 \left(-\frac{\alpha^2}{4} r^2 + 1 \right),$$

which is vanishing if $r = \frac{2}{\alpha}$. That is, if $\alpha > 0$ the locus $\lambda \mapsto (t, \frac{2}{\alpha}, z, \varphi(\lambda))$ with t, z constant is a lightlike curve in M which, barring the need for a second chart to cover the whole $\mathbb{S}^1$ is closed, φ being the angular coordinate, and it is also a geodesic. Hence we have found a spacetime where a point p is strictly contained in its causal future and past. This is also known as a **closed lightlike curve** and it is the basic example of what it is believed to be physically meaningless.

We want thus to single out those spacetimes which contain closed timelike or lightlike curves, but also those with "almost" closed causal curves, that is, a small perturbation of the metric around a point leads to the formation of a closed causal curve. Such a request leads to the following definition:

Definition 1.2.5. *A spacetime (M, g) is called* **strongly causal** *if $\forall p \in M$ and for all open sets$\mathcal{O} \ni p$, there exists a subset $V \subseteq \mathcal{O}$ containing p and such that no causal curve intersects V more than once.*

Unfortunately this requirement does not suffice neither to single out the possibility that a perturbation of the metric around two or more points leads to closed causal curves nor to guarantee the existence of a splitting of M which singles out a family of hypersurfaces where to assign Cauchy initial data for a suitable field equation. Let us try to kill two birds with one stone and, to this avail, we need to introduce a few characterizations of hypersurfaces of codimension 1.

Definition 1.2.6. *Let Σ be a non-empty closed set of M. We say that Σ is* **achronal** *if $\nexists p, q \in \Sigma$ such that $q \in I^+(p)$. Furthermore we call*

- *the* **edge** *of Σ, the set of $p \in \Sigma$ such that, for all $\mathcal{O} \subset M$ containing p, there exist $q \in \mathcal{O} \cap I^+(p)$ and $r \in \mathcal{O} \cap I^-(p)$ and a timelike curve which joins q and r without intersecting Σ,*
- *the* **domain of dependence** *of Σ, $D(\Sigma)$, the set of all points $p \in M$ such that every inextensible curve through p intersects Σ.*

From this definition, it descends a proposition whose proof is available in chapter 8 of [57]:

Proposition 1.2.1. *Any non-empty closed achronal set $\Sigma \subset M$ with empty edge is an embedded C^0 submanifold of M of codimension 1.*

This proposition leads to the last definition of this "geometric chapter":

Definition 1.2.7. *A time-oriented spacetime (M, g) is called* **globally hyperbolic** *if and only if it contains a non-empty closed achronal set Σ with empty edge and $D(\Sigma) = M$. Furthermore Σ is called a* **Cauchy (hyper)surface** *of (M, g).*

It is absolutely immediate to recognize that the above definition is to say the least unwieldy, since we shall face scenarios where we know the metric g in a local chart together with the local topology of M and we will need to infer if (M, g) is globally hyperbolic. At the same time it might even be unclear the connection with definition 1.2.5. This can luckily be easily solved by the following lemma (see §8 in [57]):

Lemma 1.2.1. *Let (M, g) be a globally hyperbolic spacetime. Then it is strongly causal and for all $p, q \in M$ the intersection $J^+(p) \cap J^-(q)$ with the tips included is compact.*

A reader should be warned that, in more modern textbooks, strong causality and compactness of the so-called double cones is actually taken as the definition itself of a globally hyperbolic spacetime (see, for example, [1]). We want also to stress that the condition of strong causality might appear just a mathematical tool, but actually it has also a physical consequence when one thinks of a QFT over a curved backgrounds. As a matter of fact, it is possible to show that for every point p of a strongly causal spacetime (M, g), there always exists an open subset $\mathcal{O} \subset M$ containing p and such that $(\mathcal{O}, g)$ is globally hyperbolic. Since, as we shall see later, a quantum field theory is a local theory naturally defined on globally hyperbolic spacetimes, it is indeed conceivable to set a quantization scheme on a strongly causal spacetime, by requiring that it coincides with the standard one on every globally hyperbolic subset. In our opinion this is indeed an important possibility and, although we shall not pursue it here, it is certainly worth mentioning it.

Yet we still need to provide a practical characterization for globally hyperbolic spacetimes and it is somehow curious that this was not available until recently thanks to the work of Bernal and Sanchez who, starting from [30], proved in theorem 1.1. of [3] the following:

Theorem 1.2.1. *Let (M, g) be an four dimensional time-oriented spacetime. Then the following statements are equivalent:*

1. *there exists a Cauchy hypersurface Σ in M*
2. *(M, g) is isometric to $\mathbb{R} \times \Sigma$ with $ds^2 = -\beta dt^2 + h_{ij} dx^i dx^j$. Here (t, x_i) with $i = 1, ..., 3$ is a suitable coordinate system such that $\beta \in C^\infty(M; (0, \infty))$, h is a Riemannian metric on Σ depending smoothly on t and each locus $\{t = const\} \times \Sigma$ is a smooth spacelike Cauchy hypersurface embedded in M.*

It is clear that this theorem allows to conclude that a large class of spacetimes is globally hyperbolic by just inspecting their metric and the domain of definition of the given coordinate system. We shall now give some notable examples of globally hyperbolic spacetimes to acquaint the reader with some concrete cases.

Example: We shall list a few globally hyperbolic spacetimes which appear commonly in quantum field theory over curved backgrounds. As one can infer per direct inspection, they all fulfil the second condition of theorem 1.2.1:

- the prototype, that is *Minkowski spacetime* which is topologically $\mathbb{R}^4$ and, in a Cartesian frame,
$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2,$$
- *de Sitter spacetime* (dS), that is the maximally symmetric solution of Einstein's equations with a positive cosmological constant Λ. As a manifold it is topologically $\mathbb{R} \times \mathbb{S}^3$ and the metric reads:
$$ds^2 = -dt^2 + \frac{3}{\Lambda} \cosh^2 \left(\sqrt{\frac{\Lambda}{3}} t \right) \left[d\chi^2 + \sin^2 \chi \left(d\theta^2 + \sin^2 \theta d\varphi^2 \right) \right],$$
where $t \in \mathbb{R}$ while (χ, θ, φ) are the standard coordinates on $\mathbb{S}^3$. Notice that, on the contrary, *anti-de-Sitter spacetime* (AdS), the maximally symmetric solution of Einstein's equations with a negative cosmological constant is not globally hyperbolic since the time coordinate turns out to be periodic. There are possible ways to circumvent this problem, such us taking the so-called universal cover, and hence one can still set up a quantum field theory over an AdS spacetime. We shall not pursue this topic here, but the reader should keep in mind such possibility.
- *the Friedmann-Robertson-Walker spacetime*, *i.e.*, an isotropic and homogeneous manifold which is topologically $\mathbb{R} \times \Sigma$ and
$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\varphi^2) \right],$$

where k can be either 0 or ± 1. Hence, depending on such a choice, the local topology of Σ becomes that of a plane ($k = 0$), a sphere ($k = -1$) or of an hyperboloid ($k = 1$). The function $a(t)$ is smooth and positive valued, though, in principle t might range on an open interval $I \subseteq \mathbb{R}$. In this case it is always possible to find a new coordinate $t' = t'(t)$ which is defined on the whole real axis, so to fulfil condition 2. in theorem 1.2.1.

- *Schwarzschild spacetime, i.e.*, a stationary spherically symmetric solution of vacuum Einstein's equations, whose metric reads

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2M}{r}} + r^2(d\theta^2 + \sin^2\theta d\varphi^2).$$

  Here M is interpreted as the mass of the spherically symmetric source (a black hole, a star,...) and the domain of definition of the coordinates is $t \in \mathbb{R}$, $r \in (0, \infty)$ and $(\theta, \varphi) \in \mathbb{S}^2$. Notice that the locus $r = 2M$ is only an apparent singularity due to a "bad choice" of coordinates, while $r = 0$ is a true singularity. The spacetime is globally hyperbolic although the reader should pay attention not to conclude that the above expression fulfils criterion 2 in theorem 1.2.1 since the locus $t = const$ is not necessarily endowed with Riemannian metric, the coefficient g_{rr} being negative when $0 < r < 2M$. Yet the spacetime is globally hyperbolic and an analysis of the geometric properties of Schwarzschild spacetime can be found either in chapter 6 of [57] or in [38].

To conclude the section we can summarize the whole discussion as follows:

Consequence 1.2.0. The natural playground of a classical and of a quantum field theory is a four dimensional globally hyperbolic spacetime (M, g). Initial data for the underlying dynamical problem are thus assigned on any Cauchy surface Σ embedded in M.

Chapter 2

Classical field theory

2.1 The real Klein-Gordon field and its Cauchy problem

In the previous section we introduced all the ingredients needed to construct a full-fledged real scalar field theory on a four dimensional globally hyperbolic spacetime (M, g). Yet, before we write down the counterpart of (1.1) on (M, g) an important physical remark is in due course. In Minkowski spacetime $(\mathbb{R}^4, \eta)$ the presence of a large isometry group,$SO(3,1) \rtimes \mathbb{R}^4$, entails the existence of a well-defined mathematical procedure which allows for a classification of a free field as a map from $\mathbb{R}^4$ into a suitable finite dimensional target Hilbert space transforming under a unitary and irreducible representation of the Poincaré group. This procedure, first introduced by Wigner, has the net advantage that it does not only offer an unambiguous identification of a free field, but it also allows for the construction of their equation of motion once the squared mass and the sign of the energy have been fixed. Hence, in Minkowski spacetime, there is no arbitrariness when dealing with a classical free field theory, but, in a curved scenario, the situation is not so idyllic. As a matter of fact, the metric g can in principle possess no isometry at all and, thus, there is no hope to implement a Wigner like procedure on (M, g). Therefore, even in the simplest case of a real scalar field, there is no unambiguous way to construct the associated Lagrangian and our only guiding principles will be Occam's razor as well as the requirements that the theory is invariant under diffeomorphisms (see definition 1.1.3) and that, whenever $(M, g) = (\mathbb{R}^4, \eta)$ we recover the standard picture in Minkowski spacetime.

Out of these remarks, one can infer that, in the counterpart of (1.1) on (M, g), it is in principle possible to add any scalar function constructed out of the tensors in section 1.1.4 and this would vanish whenever $g = \eta$. Yet, if we stick to an Occam's razor perspective, we should discard this possibility with the the exception of a coupling to scalar curvature of the form $\xi R\phi^2$ in the Lagrangian. The reason is twofold: on the one hand, even if we omit this term, it will nonetheless reappear later when discussing the regularization of a scalar field theory in (M, g) (see also [26]) and, on the other hand, a notable property of the wave equation in Minkowski spacetime is the propagation of a solution along the light rays stemming from the support of the initial data. This field theoretical counterpart of the Huygens' principle holds true also in a curved background only if $\xi = \frac{1}{6}$ and $m^2 = 0$. Hence, it is customary to allow for $\xi \neq 0$ and, in this case, the field is called *non minimally coupled*. A third motivation arises a posteriori from the discussion of the quantization scheme of a real scalar field theory; as a matter of fact, it turns out that, under certain circumstances such as for example when the background is de Sitter spacetime, it is impossible to find a suitable ground state (*i.e.*, a low-energy state invariant under all background isometries) for a massless scalar field theory with no coupling to scalar curvature.

If we follow this line of reasoning, we shall call **action of a Klein-Gordon field** on a globally hyperbolic spacetime (M, g):

$$S[\Phi] = \int\limits_M d^4x \frac{\sqrt{|g|}}{2} \left[g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi + (m^2 + \xi R)\Phi^2\right], \tag{2.1}$$

where $\Phi : M \to \mathbb{R}$ and the integration over M has to be read along the lines of definition 1.1.8. We recall that $\xi \in \mathbb{R}$ and m^2 is a positive constant to be interpreted as the squared mass of the field. If one varies the above action, the associated Cauchy problem is the following:

$$\begin{cases} P\Phi \doteq \left(\Box_g - \xi R - m^2\right)\Phi = 0 \\ \Phi|_\Sigma = \Phi_0 \in C_0^\infty(\Sigma), \quad \partial_n \Phi|_\Sigma = \Phi_1 \in C_0^\infty(\Sigma) \end{cases}, \tag{2.2}$$

where Σ is a Cauchy surface whose normal vector is n and $\Box_g$ is the d'Alembert wave operator as in equation (1.4). This is the **Klein-Gordon equation** on (M, g) and, if $\xi = 0$ it is called *minimally* coupled to scalar curvature whereas, if $\xi = \frac{1}{6}$, *conformally* coupled. The choice of this word is not random since, it is possible to show that, whenever $m^2 = 0$, the conformal coupling guarantees that, to every solution of (2.2) on (M, g), it possible to associate a unique solution of the same problem on a conformally related globally hyperbolic spacetime (M, g') where $g' = \Omega^2 g$ and $\Omega \in C^\infty(M, (0, \infty))$. The new solution is constructed rescaling both the original one and its initial data on Σ by Ω^{-1}.

2.1.1 Existence and Uniqueness of the solution

The first question one should ask looking at (2.2) concerns the existence and the uniqueness of the solution, a problem to which this section will be devoted. From a mathematical point of view, if we write in a local coordinate system the equation of motion, we are interested in, we end up working with an hyperbolic second order partial differential equation (PDE). These have been extensively studied in the literature and we recommend an interested reader to tackle [54] or [37] for a deep analysis of all the technical tools and of the results which are involved in the analysis of these systems. Conversely, we shall mostly follow [1] where a reader can find all the proofs of the theorems we shall not demonstrate.

Proposition 2.1.1. *Let us consider on a globally hyperbolic spacetime (M, g) a second order hyperbolic PDE of metric principal type, that is, in a local coordinate system defined over an open subset $\mathcal{O} \subseteq M$, the operator reads*

$$P = g^{\mu\nu}(x)\partial_\mu \partial_\nu + A^\mu(x)\partial_\mu + B(x), \tag{2.3}$$

where both B and A^μ for each $\mu = 0, ..., 3$ are smooth functions on $\mathcal{O}$. Then, for any Cauchy surface Σ in M, the Cauchy problem

$$\begin{cases} P\Phi = 0 \\ \Phi|_\Sigma = \Phi_0 \in C_0^\infty(\Sigma), \quad \partial_n \Phi|_\Sigma = \Phi_1 \in C_0^\infty(\Sigma) \end{cases}, \tag{2.4}$$

admits a unique and smooth solution Φ in $\mathcal{O}$ such that

$$supp(\Phi) \subseteq J^+(supp(\Phi_0)) \cup J^+(supp(\Phi_1)) \cup J^-(supp(\Phi_0)) \cup J^-(supp(\Phi_1)). \tag{2.5}$$

This theorem basically solves all our problems since it guarantees us both the existence and the uniqueness of a solution of (2.2) and, furthermore, it also provides us with informations on the support of the solution itself. Within this respect, we note that the statement tells us a lot about the difference of the Klein-Gordon equation in a curved and in a flat spacetime. First of all (2.5) holds true irrespectively from the coefficients of the operator P and, thus, the obtained result is not always sharp. As an example, one can consider the wave equation in Minkowski spacetime where the information propagates exactly on the light rays (the Huygens' principle) and not on the interior of the causal future and past. Yet, it is possible to give concrete examples of spacetimes, such as the Schwarzschild black hole, where the solutions of $\Box_g \Phi = 0$ with smooth compactly supported initial data propagates really fulfil (2.5) with $=$ in place of $\subseteq$.

Furthermore, in the framework of quantum field theory over curved backgrounds, it is customary to use a different approach in order to construct the solutions of (2.2) and, due to its importance in the analysis of the quantum theory, we shall now sketch it. Overall this method is based on the following proposition (see §3.4 in [1]):

Proposition 2.1.2. *Let (M, g) be a time-oriented spacetime and let P be as in (2.3). Then there exists a unique* **advanced and retarded Green operator** $G^+ : C_0^\infty(M) \to C^\infty(M)$ *and* $G^- : C_0^\infty(M) \to C^\infty(M)$ *such that they are linear and*

1. $P \circ G^\pm = id : C_0^\infty(M) \to C_0^\infty(M)$,
2. $G^\pm \circ P|_{C_0^\infty(M)} = id : C_0^\infty(M) \to C_0^\infty(M)$,
3. $supp[G^\pm(f)] \subseteq J^\pm(supp(f))$ *for all* $f \in C_0^\infty(M)$.

The net advantage of the Green operators is that their difference is the so-called **causal propagator** $E = G^+ - G^- : C_0^\infty(M) \to C^\infty(M)$ which yields $\Phi_f \doteq E(f)$, a solution of (2.2) out of a smooth compactly supported functions over the spacetime M, *i.e.*, $PE(f) = 0$. Furthermore the support properties of Φ_f are exactly those of (2.5).

Yet the personal experience of the author with himself and with many students is that there is a certain difficulty in reconciling propositions 2.1.2 with 2.1.1. The standard criticism to this approach is that proposition 2.1.1 sounds more natural since it builds a solution of a second order PDE out of two initial data on a Cauchy surface as it is customary seen in the standard approaches to ordinary and partial differential equations (think for example of the harmonic oscillator). At the same time, the causal propagator needs just one function on the whole manifold to generate a solution and, thus, at first glance, it seems odd that the two approaches yield the same space of solutions. We deem thus absolutely necessary to fill in this gap and to prove the following:

Proposition 2.1.3. *The set of solutions of the Cauchy problem in proposition 2.1.1 is in one-to-one correspondence with the set of functions $\Phi_f \doteq E(f)$. Here f is chosen between the representatives of the equivalence classes of smooth compactly supported functions over M, such that $f \sim f'$ if and only if there exists $g \in C_0^\infty(M)$ such that $f - f' = Pg$.*

Proof. Let us start from $\Longleftarrow$. Let us consider any representative f, as per hypothesis, and the associated solution $\Phi_f = E(f)$. Since M is globally hyperbolic, theorem 1.2.1 guarantees that $M \sim \mathbb{R} \times \Sigma$ and, hence, we can always choose a Cauchy surface Σ in the future of *supp*(f) whose intersection with the support of Φ_f is non-empty. Furthermore $supp(\Phi_f) \cap \Sigma \subset J^+(supp(f)) \cap \Sigma$ must be compact too and the restrictions of both Φ_f and $\partial_t \Phi_f$ on Σ are smooth, being $\Phi_f \in C^\infty(M)$. Hence the pair $(\Phi_f|_\Sigma, \partial_t \Phi_f|_\Sigma)$ can be thought as an initial datum for (2.4) whose solution exists per proposition 2.1.2. Furthermore uniqueness and linearity of P also yields that it must coincide with Φ_f. Let us now look at $\Longrightarrow$. If we assign the initial data (Φ_0, Φ_1) of (2.4) we can always identify at least one $f \in C_0^\infty(M)$ whose restriction on Σ gives the initial data. The construction of f can be explicitly carried out for example using the function $e^{-\frac{1}{t^2}}$ as building block and we leave the details to the reader. Hence, slavishly following the reasoning of the first part of the proof, we conclude that $E(f)$ coincides with the solution generated out of the initial data on Σ, which leaves us with one last step to prove. Let us take a second $f' \in C_0^\infty(M)$ whose restriction on Σ yields the same initial conditions. Therefore $E(f) = E(f')$, or, in other words, per linearity $E(f-f') = G^+(f-f') - G^-(f-f') = 0$; if we call $h \doteq f - f'$ and we apply the operator P, it holds per proposition 2.1.2 that

$$PG^+(h) = PG^-(h) = h.$$

Furthermore, since $supp(G^\pm h) \subseteq J^\pm(h)$ and $G^+(h) = G^-(h)$, $supp(G^\pm(h)) \subseteq J^+(supp(h)) \cap J^-(supp(h))$ which is compact since $h \in C_0^\infty(M)$. To conclude, since P is a properly supported[1] operator, we can simply define $g \doteq G^+(h) \in C_0^\infty(M)$ and it holds $Pg = f - f'$. □

[1] A (pseudo)differential operator is called *properly supported* when it maps all the distributions supported on an arbitrary compact K into those supported on a second compact K'.

2.2 The structure of the space of solutions

The last bit of information that we need from the classical theory concerns the structure of the space of solutions of (2.2). This is a problem of interest on its own, but it shall play a pivotal role in the study of the quantum theory as we shall see in the next section, so we must spend a few lines on this topic. Hence we call

$$\mathcal{S}(M) = \{\Phi_f \in C^\infty(M) \mid \exists f \in C_0^\infty(M) \text{ and } \Phi_f = E(f)\},$$

where E is still the causal propagator. In order to understand more about $\mathcal{S}(M)$ we first need an auxiliary definition (for this topic we shall use the approach and the nomenclature of [53]):

Definition 2.2.1. *A possibly infinite dimensional vector space V is called* **symplectic** *if it is endowed with a map $\sigma : V \times V \to \mathbb{R}$ which is bilinear and antisymmetric. Furthermore σ induces a map $\sigma^0 : V \to V^*$ such that $\sigma^0(v) = \sigma(v, \cdot)$ for all $v \in V$. If σ^0 is an injective map, than we say that σ is weakly non-degenerate, whereas if σ^0 is an isomorphism, then σ is called strongly non-degenerate.*

The above distinction is meaningful only if V is not finite dimensional, since, in this latter case, every weakly non-degenerate symplectic form is automatically strongly non degenerate. In our case the candidate vector space is $\mathcal{S}(M)$ and so the best, we can hope for, is the following proposition:

Proposition 2.2.1. *The set $\mathcal{S}(M)$ is a weakly non-degenerate vector space if endowed with the form*

$$\sigma(\Phi_1, \Phi_2) = \int\limits_\Sigma (\Phi_1 \nabla_n \Phi_2 - \Phi_2 \nabla_n \Phi_1)\, d\mu(\Sigma), \tag{2.6}$$

where the fields and their derivatives are meant as restricted on the Cauchy surface Σ with normal n. Furthermore σ is independent from the choice of Σ.

Proof. We start noticing that $\mathcal{S}(M)$ is a vector space due to the linearity of P. Furthermore, per direct inspection, (2.6) is also bilinear and antisymmetric. We need to prove that the associated map $\sigma^0 : \mathcal{S}(M) \to \mathcal{S}(M)^*$ is injective, which is tantamount to show that $\sigma(\Phi_1, \Phi_2) = 0$ for all $\Phi_2 \in \mathcal{S}(M)$ implies that $\Phi_1 = 0$. Let us suppose, per absurd, that it is not true and that $\Phi_1 \neq 0$. Then we can choose Φ_2 as the solution generated by the initial data $(0, \Phi_1|_\Sigma)$. This is tantamount to say that

$$\sigma(\Phi_1, \Phi_2) = \int\limits_\Sigma \Phi_1^2 d\mu(\Sigma) = 0.$$

Since the restriction of an element of $\mathcal{S}(M)$ on Σ is smooth and compactly supported, it also holds that $\Phi|_\Sigma \in L^2(\Sigma, d\mu(\Sigma))$ and, hence, the above identity also reads $||\Phi_1|_\Sigma||_{L^2} = 0$, which implies $\Phi_1|_\Sigma = 0$ almost everywhere on Σ. The same result holds also for $\nabla_n \Phi_1$ on Σ repeating the same reasoning assuming that Φ_2 is generated by the initial data $(\nabla_n \Phi_1|_\Sigma, 0)$. Since P is linear, it holds that the function identically 0 is the only solution with vanishing initial data, which proves injectivity.

To conclude the proof we need to show independence of (2.6) from the choice of Σ. To this avail, let us choose any two Cauchy surfaces, say Σ and Σ' and, for any $\Phi_1, \Phi_2 \in \mathcal{S}(M)$, let us introduce $J_\mu = \Phi_1 \nabla_\mu \Phi_2 - \Phi_2 \nabla_\mu \Phi_1$. This is a current since

$$\nabla^\mu J_\mu = \Phi_1 \Box_g \Phi_2 - \Phi_2 \Box_g \Phi_1 = \Phi_1 P(\Phi_2) - \Phi_2 P(\Phi_1) = 0.$$

Hence we can take an integration domain $V \subset M$ sufficiently large (see figure 2.2.1) so that

$$0 = \int\limits_V d^4x \sqrt{|g|} \nabla^\mu J_\mu = \int\limits_\Sigma d\mu(\Sigma)\, n^\mu J_\mu - \int\limits_{\Sigma'} d\mu(\Sigma)\, n^\mu J_\mu,$$

which implies independence of the symplectic form from the choice of the Cauchy surface. □

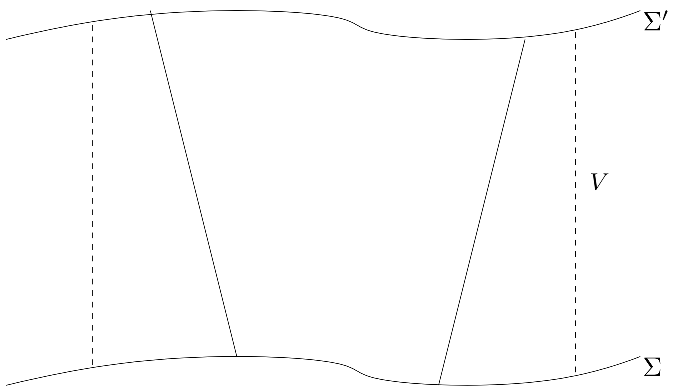

Figure 2.1: This is sketchy representation of the domain of integration V used in the proof of proposition 2.2.1. V is chosen in such a way to admit as a boundary an open neighbourhood of both Σ and Σ', two Cauchy surfaces. The oblique lines indicate that the union of the support of Φ_1 and Φ_2 propagates along the null geodesic from Σ to Σ' and thus V can always be chosen to include it.

2.2.1 Classical Observables

As it is widely known, one of the key points of a quantum theory is the discussion of the structure of its algebra of observables, an issue which is actually present also in classical mechanics although it is not always so emphasized. As a preparatory tool before quantizing the scalar field theory, it is certainly worth to spend a few words on the notion of classical observable. To start with, let us recall that a customary aspect of Hamiltonian mechanics is that the underlying phase space is actually symplectic and thus it is natural to follow the same procedure and identify $(\mathcal{S}(M), \sigma)$ as the phase space of the theory. We call a **classical observable** a (smooth) real function on such a phase space and, particularly, we can associate to each $f \in C_0^\infty(M)$ an observable as $F_f : \mathcal{S}(M) \to \mathbb{R}$ such that

$$F_f(\Phi) \doteq \Phi(f) = \int_M d^4x \sqrt{|g|}\Phi(x) f(x). \tag{2.7}$$

We notice that, on the left hand side of the equality, we are interpreting $\Phi \in C^\infty(M)$ as a distribution on $\mathscr{D}'(M)$ the set of continuous linear functionals from $C_0^\infty(M)$ into $\mathbb{R}$. Hence the just defined subclass of classical observable is notable since it satisfies the equation of motion in a distributional/weak sense:

$$F_f(P\Phi) = P\Phi(f) = \Phi(Pf) = F_{Pf}(\Phi) = 0. \quad \forall \Phi \in \mathcal{S}(M) \ \wedge \ \forall f \in C_0^\infty(M)$$

The next natural step is to endow the set of observables with a Poisson structure and, to this avail, it is important to recall that theorem 1.2.1 guarantees that we can always write locally the metric as $ds^2 = -\beta dt^2 + h_{ij} dx^i dx^j$. If we plug this form in (2.1), up to a global and irrelevant sign which depends on the convention of the signature, we end up with

$$S[\Phi] = \int\limits_{\mathbb{R}} dt L[\Phi] = \int\limits_{\mathbb{R}} dt \int\limits_{\Sigma} d^3x \frac{\sqrt{h}}{2} \left[\frac{(\partial_t \Phi)^2}{\sqrt{\beta}} - \sqrt{\beta} \left(h^{ij} \partial_i \Phi \partial_j \Phi + (m^2 + \xi R) \Phi^2 \right) \right].$$

Since the Lagrangian functional $L[\Phi]$ omits integration over time, and since the restriction of elements of $\mathcal{S}(M)$ on a Cauchy surface are compactly supported, we can switch from Φ to $\phi = \Phi|_\Sigma$, hence identifying as configuration space $C_0^\infty(\Sigma)$. Furthermore, since $L[\phi, \partial_t \phi, t]$ is a convex function, we can switch to an Hamiltonian description as:

$$\begin{cases} \Pi \doteq \frac{\delta L}{\delta(\partial_t \phi)} = \frac{\sqrt{h}}{\sqrt{\beta}} \partial_t \phi \\ H[\phi, \Pi] = \int\limits_{\Sigma} \frac{\sqrt{\beta}}{2} \left(\frac{\Pi^2}{\sqrt{h}} + \sqrt{h} h^{ij} \partial_i \phi \partial_j \phi + (m^2 + \xi R) \phi^2 \right) \end{cases}. \tag{2.8}$$

The reader should notice that we have omitted the integration measure over Σ because the (conjugate) momentum Π is actually a semi-density over $C_0^\infty(\Sigma)$. Furthermore, in the language of Hamiltonian mechanics, the phase space $\mathcal{P}$ of the theory would be naturally identified with the set of kinematically allowed configurations$(\phi, \frac{\Pi}{\sqrt{h}}) \equiv C_0^\infty(\Sigma) \times C_0^\infty(\Sigma)$ where the factor multiplying Π is present in order to have a function and not a semi-density. Notice also that this choice for $\mathcal{P}$ does not contradict what we claimed before since each element therein can be seen as the initial datum of the Hamilton's equations out of (2.8) whose solutions yield a unique element of $\mathcal{S}(M)$. Therefore a classical observable can also be equivalently seen as a smooth real function over $\mathcal{P}$.

We are now ready to introduce the **Poisson brackets** between two classical observables F and G as:

$$\{F, G\}(\phi, \Pi) = \int\limits_{\Sigma} \left(\frac{\delta F}{\delta \phi} \frac{\delta G}{\delta \Pi} - \frac{\delta F}{\delta \Pi} \frac{\delta G}{\delta \phi} \right), \tag{2.9}$$

where no integration measure is present since the variation with respect of ϕ actually yields a density and that with respect to Π a function. Let us now look at the Poisson brackets between two special observables, namely $F_f(\Phi)$ and $F_{f'}(\Phi)$ for $f \neq f' \in C_0^\infty(M)$. To this end we need a useful identity first proven in [21]:

$$F_f(\Phi) = \int\limits_{\Sigma} (\phi_f \Pi - \Pi_f \phi), \tag{2.10}$$

where $\phi_f \doteq E(f)|_\Sigma$, E being the causal propagator. The momentum Π_f is defined as in (2.8) plugging $E(f)$ in place of ϕ. That said, (2.9) yields:

$$\{F_f, F_{f'}\} = \int\limits_{\Sigma} (\phi_f \Pi_{f'} - \Pi_f \phi_{f'}) = F_f(E(f')) = E(f, f'), \tag{2.11}$$

where in the second equality we used (2.10) and, in the third, (2.7). Furthermore, since it will be convenient later, we also notice that (2.8) yields that we can reread the last equality as

$$\{F_f, F_{f'}\} = \int\limits_{\Sigma} d\mu(\Sigma) \left(\Phi_f \nabla_n \Phi_{f'} - \Phi_{f'} \nabla_n \Phi_f \right) = \sigma(\Phi_f, \Phi'_f).$$

Chapter 3

Quantization scheme - why the algebraic approach?

The goal of this chapter is only partly to introduce a quantization scheme for a free scalar field theory on a globally hyperbolic spacetime; actually our ultimate scope is to convince the reader that, whenever one deals with quantum issues on a curved background, the algebraic approach is a natural and powerful tool. At the same time we wish to strongly emphasize that his method is not an alternative to the usual schemes, but it is actually totally consistent with them, and its range of applicability is larger than the others. Of course there is a price to pay for this advantage, namely we are forced to accept an higher degree of abstractness and the mathematical tools, we shall need, can quickly become rather involved. Therefore I fear that a direct presentation of the algebraic approach might cause the opposite effect and push away a reader from it; hence we shall first consider a special class of manifolds where it is applicable a scheme similar to the one used in Minkowski spacetime to quantize the Klein-Gordon field. In reading the next section, we encourage the reader to pay attention above all to the limitations of the used arguments, namely to understand where and why they fail, so that it becomes clear that the adoption of a different point of view, albeit consistent when possible with the others, is a forced choice. In this enterprise we shall really benefit from the lecture notes of Chris Fewster, who gave a course on similar topics in Leipzig in 2008 [24].

As a starting point let us briefly recall that the quantization prescription first introduced by Dirac calls for associating to a classical observable F_f, as in (2.7), an operator $\widetilde{F}_f$ on a suitable Hilbert space $\mathcal{H}$ such that

$$[\widetilde{F}_f, \widetilde{F}_{f'}] \doteq i\,\{F_f, F_{f'}\}\,\mathbb{I} = iE(f, f')\mathbb{I},$$

where $\{,\}$ are the Poisson brackets, $\mathbb{I}$ is the identity operator and, in the third equality, we used (2.11). Notice that the support properties of the causal propagator out of proposition 2.1.2 guarantee the automatic implementation of the canonical idea that spacelike separated observables must commute. More generally if we even allow f to be complex valued, when quantizing, we look for an Hilbert space on which are acting operators constructed in such a way to fulfil the following properties for all $f, f' \in C_0^\infty(M;\mathbb{C})$:

1. the maps $f \mapsto \widetilde{F}_f \doteq \widetilde{F}_{Re(f)} + i\widetilde{F}_{Im(f)}$ is $\mathbb{C}$-linear,
2. $\widetilde{F}_f^* = \widetilde{F}_{\overline{f}}$,
3. $\widetilde{F}_{Pf} = 0$ where P is the Klein-Gordon operator as in (2.2),
4. $[\widetilde{F}_f, \widetilde{F}_{f'}] = iE(f, f')\mathbb{I}$.

We shall now show how this prescription can be directly implemented on a very special class of spacetimes.

3.1 Quantization on ultrastatic spacetimes

The content of this section summarizes the discussion of [24] and even more details can be found in [26]. As a starting point we look for a spacetime which is not too different from Minkowski, above all with respect to the dependence from the time variable. To make concrete this idea, we need "unfortunately" a new geometric concept:

Definition 3.1.1. *A vector field v is called a* **Killing field** *if $\nabla_{(\mu}v_{\nu)} = 0$ where $v_\nu = g_{\rho\nu}v^\rho$.*

From a pure geometrical point of view Killing fields are the local counterpart of an isometry, that is a map $\chi : M \to M$ such that $\chi^* g = g$ where χ^* is the pull-back map[1]. In this case we can construct v simply as $d\chi$. Yet the real problem when a metric g is assigned is to recognize whether or not it possesses any of these isometries and a reasonably useful way is a by-product of this lemma:

Lemma 3.1.1. *Suppose that, for a given spacetime (M, g) and a given coordinate system x_μ with $\mu = 1, ..., \dim M$, the metric coefficients do not depend upon one of the coordinates, say x_1. Then the vector field v which coincides pointwisely with $\frac{\partial}{\partial x_1}$ is a Killing field.*

Proof. Let us choose v as per hypothesis so that $v = \partial_{x_1}$. Then $2\nabla_{(\mu}v_{\nu)} = \partial_\mu v_\nu + \partial_\nu v_\mu - 2\Gamma^\rho_{\mu\nu}v_\rho$ where, if we use (1.3),

$$2\Gamma^\rho_{\mu\nu}v_\rho = g^{\rho\delta}\left(\partial_\mu g_{\nu\delta} + \partial_\nu g_{\mu\delta} - \partial_\delta g_{\mu\nu}\right)v_\rho = g^{\rho\delta}\left(\partial_\mu g_{\nu\delta} + \partial_\nu g_{\mu\delta}\right)v_\rho = \partial_\mu v_\nu + \partial_\nu v_\mu,$$

where, in the last equality, we used the definition of v and the independence of the metric from the coordinate x_1. □

With this result we understand a little bit more what is a spacetime with symmetries and we can therefore state the following.

Definition 3.1.2. *A four dimensional globally hyperbolic spacetime (M, g) is called* **ultrastatic** *if:*

- *it is static, that is, there exists a timelike Killing field v and an embedded spacelike 3-dimensional hypersurface Σ orthogonal to the orbits of the isometry giving rise to v,*
- *there exists a coordinate system $(t, x_i) = (t, x, y, z)$ - not necessarily Cartesian - such that the metric reads*

$$ds^2 = -dt^2 + h_{ij}(x, y, z)dx^i dx^j. \tag{3.1}$$

We can now quantize (2.2) on these backgrounds. Our discussion will be a sort of cooking recipe and so we will proceed step by step as in a cookbook:

Step 1) Since the spacetime is ultrastatic the Klein-Gordon operator P in (2.2) reads

$$P = -\frac{\partial^2}{\partial t^2} + K, \quad K = \nabla_h^2 - (m^2 + \xi R) = \frac{1}{\sqrt{h}}\partial_i(\sqrt{h}h^{ij}\partial_j) - (m^2 + \xi R),$$

where ∇_h^2 is the Laplace operator associated to h.

Step 2) Following Dirac's prescription we promote the classical observable F_f to an operator $\widetilde{F}_f$ which must solve weakly the equation of motion, that is $\widetilde{F}_{Pf} = 0$ for all $f \in C_0^\infty(\Sigma)$. With a slight abuse of notation, we

[1] Let M and N be smooth manifolds and $F \in C^\infty(M; N)$. Then the pull-back $F^* : T^*M \to T^*N$ is the dual map of the push-forward defined in proposition 1.1.1, namely, for all $\omega \in T^*N$, $F^*\omega(v) \doteq \omega(F_* v)$ for all $v \in TM$. The definition can be straightforwardly generalized to any tensor of type $(0, q)$.

can consider the observable at time $t = 0$ introducing $\widetilde{\phi}(x) \doteq \widetilde{F}_f(0, x)$ and the momentum (formally related to Π though without $\sqrt{h}$) $\widetilde{\pi}(x) \doteq (\partial_t \widetilde{F}_f)(0, x)$ so that the commutation relation

$$[\widetilde{F}_f, \widetilde{F}_{f'}] = iE(f, f')\mathbb{I} \tag{3.2}$$

can be also read as

$$[\widetilde{\phi}(f), \widetilde{\pi}(g)] = i(\overline{f}, g)_{L^2}\mathbb{I}, \quad [\widetilde{\phi}(f), \widetilde{\phi}(g)] = [\widetilde{\pi}(f), \widetilde{\pi}(g)] = 0,$$

for all $f, g \in C_0^\infty(\Sigma)$. Here $(,)$ stands for the inner product of $L^2(\Sigma)$.

Step 3) We *assume only for simplicity* that Σ is compact so that K has discrete eigenvalues and, thus we can found a basis of $L^2(\Sigma)$ constructed out of smooth functions ψ_j fulfilling $K\psi_j = \omega_j^2\psi_j$. Hence every function $\psi \in L^2(\Sigma)$ or even $\psi \in C^\infty(\Sigma)$ can be expanded as

$$\psi = \sum_j c_j \psi_j,$$

where the c_j are suitable coefficients and the index j runs over an infinite but countable set. Let us suppose for a moment that all the coefficients c_j are zero except one which is equal to 1 and that we consider $f = T\psi_j$ where T is a smooth compactly supported function depending only on the time variable t. Then the requirement that $\widetilde{F}_{Pf} = 0$ translates in $\widetilde{F}_{(\ddot{T}+\omega_j^2 T)\psi_j} = 0$ which admits a general solution as follows: there must exist two operators a_j and b_j such that

$$\widetilde{F}_{T\psi_j} = \int_{\mathbb{R}} \frac{dt}{\sqrt{2\omega_j}} T(t) e^{-i\omega_j t} b_j + \int_{\mathbb{R}} \frac{dt}{\sqrt{2\omega_j}} T(t) e^{i\omega_j t} a_j^*, \tag{3.3}$$

where we can express these new operators in terms of ϕ and π given at step 2) as follows:

$$b_j = \frac{1}{2\omega_j}\left(\omega_j \phi(\psi_j) + i\pi(\psi_j)\right),$$
$$a_j^* = \frac{1}{2\omega_j}\left(\omega_j \phi(\psi_j) - i\pi(\psi_j)\right).$$

Step 4) Let us now restore the sum over all possible indices, but, before proceeding, let us notice that $K\overline{f} = \overline{Kf}$, hence, if ψ_j is an eigenvector, so is also $\overline{\psi_j}$ and with the same eigenvalue. Thus, in the sum over all indices, for all j there must exist j' whose associated eigenvector is the complex conjugate of ψ_j and, without loss of generality, we shall write $\overline{\psi}_j = \psi_{\overline{j}}$ since this entails that we can write in (3.3) $a_{\overline{j}}$ in place of b_j. The commutation relations (3.2) become:

$$[a_j, a_k] = 0, \quad [a_j, a_k^*] = \delta_{jk}\mathbb{I}, \tag{3.4}$$

where δ is the Kronecker delta and the indices are running over all possible ones. These are exactly the **canonical commutation relations** (CCR). If we now consider a generic f which can be written as the product $f = TS$ where T is, as above, a smooth compactly supported function dependent only on t while $S \in C_0^\infty(\Sigma)$, then the above procedure generalizes and we obtain

$$\widetilde{F}_f = \int_{\mathbb{R}} dt\, T(t) \sum_j \frac{1}{2\sqrt{\omega_j}} \left((\overline{\psi}_j, S)_{L^2} e^{-i\omega_j t} a_{\overline{j}} + (\psi_j, S)_{L^2} e^{i\omega_j t} a_j^* \right),$$

which, relabelling the indices, becomes

$$\widetilde{F}_f = \int_{\mathbb{R}} dt\, T(t) \sum_j \frac{1}{2\sqrt{\omega_j}} \left((\overline{\psi}_j, S)_{L^2} e^{-i\omega_j t} a_j + (\psi_j, S)_{L^2} e^{i\omega_j t} a_j^* \right), \tag{3.5}$$

Step 5) At last we need to construct an Hilbert space where to realize the new operators and this is a simple step since we are dealing with creation and annihilation operators and thus we just need a Fock space.

Definition 3.1.3. *We call* **bosonic Fock space** *of an Hilbert space* $\mathcal{H}$

$$\mathscr{F}_s(\mathcal{H}) \doteq \sum_{n=0}^{\infty} \mathcal{H}^{\otimes_s n},$$

where $\mathcal{H}^0 \doteq \mathbb{C}$ *and* $\otimes_s$ *stands for the symmetric tensor product. If* $\mathcal{H} = L^2(\Sigma, d\mu(\Sigma))$ *in particular, then* $\mathcal{H}^{\otimes_s n}$ *coincides with the set of* $F^{(n)} \in L^2(\underbrace{\Sigma \times ... \times \Sigma}_{n})$ *which are symmetric in their arguments.*

Since the elements in $\mathscr{F}_s(\mathcal{H})$ can be represented as sequences of n-tuples $F^{(n)} \in \mathcal{H}^{\otimes_s n}$, we can define for any $f \in L^2(\Sigma)$ the action of the creation and annihilation operators as follows:

$$[a(f)F]^{(n)} \doteq \sqrt{n+1} \int_{\Sigma} d\mu(\Sigma)\overline{f}(x)F^{(n+1)}(x, x_1, ..., x_n), \tag{3.6}$$

$$[a^*(f)F]^{(n)} \doteq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(x_i)F^{(n-1)}(x_1, ..., \widehat{x}_i, ..., x_n), \tag{3.7}$$

where $\widehat{x}_i$ stands for the variable to be omitted. Notice that this definition is meaningful only on $\mathscr{F}_0$ the subset of the Fock space for which all but finitely many elements $F^{(n)}$ vanish.
These annihilation and creation operators obey indeed the canonical commutation relations since, per direct computation, one can see that, for all $f, g \in L^2(\Sigma)$

$$[a(f), a^*(g)] = (f, g)_{L^2}\mathbb{I}, \qquad [a(f), a(g)] = [a^*(f), a^*(g)] = 0. \tag{3.8}$$

To convince ourselves, let us show how the first commutator can be explicitly realized when applied to $\psi \in \mathcal{H}^{\otimes_s 1}$. Let us thus consider any $f, g \in L^2(\Sigma)$ and

$$a(f)a^*(g)\psi = \frac{a(f)}{\sqrt{2}}(g(x_1)\psi(x_2) + g(x_2)\psi(x_1)) = \int_{\Sigma} d\mu_\Sigma(x_1) \left[\overline{f}(x_1)g(x_1)\psi(x_2) + g(x_2)\overline{f}(x_1)\psi(x_1)\right] =$$

$$= (f, g)_{L^2}\psi(x_2) + a^*(g) \int_{\Sigma} d\mu_\Sigma(x_1)\, \overline{f}(x_1)\psi(x_1) = (f, g)_{L^2} + a^*(g)a(f)\psi,$$

where, in the various steps, we just used (3.7) and (3.6).

Step 6) The last step is to connect the above general procedure to identify a Fock space on $\mathcal{H} = L^2(\Sigma)$ to the particular system we have previously described. This can be achieved pertaining $\mathscr{F}_s(\mathcal{H})$ and noticing that we can read the various mode operators as $a_j \equiv a(\psi_j)$ and $a_j^* = a^*(\psi_j)$. Since the set of ψ_j forms an orthonormal basis of $L^2(\Sigma)$, the commutation relations (3.8) reduce to (3.4). Furthermore, if we consider the unit vector $\Omega \in \mathscr{F}_s(\mathcal{H})$ such that $\Omega^{(0)} = 1$ and $\Omega^{(n)} = 0$ for all $n \geq 1$, then (3.6) tells us that every annihilation operator acting on it gives 0 and thus we can interpret it as the *vacuum.* Finally we interpret each $a^*(\psi_j)$ as the creator of a pure mode while $a^*(f)$ with $f \in L^2(\Sigma)$ as the creator of a wave packet.

N.B. In the recipe we have discussed, it also common to call $\widetilde{F}_f$ the quantum field and to indicate it as $\widetilde{\Phi}(f)$ since it solves weakly (2.2). Accordingly, up to the choice of an orthonormal basis of $L^2(\Sigma)$, $\{\psi_j\}$, we can represent $\overline{\Phi}$ in the unsmeared form:

$$\widetilde{\Phi}(t, x) = \sum_j \frac{1}{\sqrt{2\omega_j}} \left(e^{-i\omega_j t}\psi_j(x)a_j + e^{i\omega_j t}\overline{\psi_j}(x)a_j^*\right),$$

where $K\psi_j = \omega_j^2\psi_j$ and $\omega_j \neq 0$. This is the standard decomposition of a quantum field in positive and negative frequency part that we are used to in Minkowski spacetime. Furthermore we can represent $\widetilde{\Phi}$ on the Fock space (or actually on $\mathscr{F}_0$) in a basis-independent form by spatial smearing as follows:

$$\int_\Sigma d\mu(\Sigma)\widetilde{\Phi}(t,x)f(x) = \frac{1}{\sqrt{2}}\left(a(e^{i\sqrt{K}t}K^{-\frac{1}{4}}\overline{f}) + a^*(e^{i\sqrt{K}t}K^{-\frac{1}{4}}f)\right),$$

where f must lie in the domain of $K^{-\frac{1}{4}}$. Notice that this prescription does not really require Σ to be compact and it boils down to the standard picture we know in four dimensional Minkowski spacetime.

Consequence 3.1.0. We have seen that it is possible to quantize a free Klein-Gordon field in an ultrastatic spacetime by means of a scheme which mimics the standard one in Minkowski spacetime, namely the identification of suitable quantum observables, particularly of a quantum field and the decomposition of the latter in positive and negative frequency parts. This allows to find a representation on a suitable bosonic Fock space constructed out of the space of square integrable functions over the Cauchy surface Σ.

This summary points out all the advantages but actually all the limitations of the employed method. Most notably the existence of a notion of a positive and negative frequency part arises intrinsically out of the metric coefficients being independent from t and thus each solution of (2.2) can be split in a time and in a spatial components. Hence, since a general globally hyperbolic spacetime (M, g) has metric coefficients depending on time (see for example the Friedmann-Robertson-Walker spacetime), how can we quantize a field theory defined thereon? The standard scheme cannot work and now we propose a suitable way out: **the algebraic approach**.

3.2 Algebraic Quantization

The previous section showed us that the "standard" approach to quantization can still work under very special circumstances, but it is ultimately doomed to failure in face of several obstacles such as:

- absence of a large isometry group such as the Poincaré, hence ultimately leading to a problem both in the construction of a classical free field theory and in the identification of a ground state for the quantum theory,
- absence of a way to make sense of concepts such as positive and negative frequency parts whenever the metric coefficients are explicitly time dependant,
- absence of a clear method to deal with observables represented on an Hilbert space which is different from the usual Fock one.

Yet our previous analysis has shed some light also on aspects of the quantization scheme which is desirable to pertain in every generalized method. More precisely it has been clarified that any approach to the quantization of the underlying theory must encompass at least two key steps:

Step 1) the recollection of all possible observables in a single body with the structure of an algebra,

Step 2) the identification of these observables as operators acting on a suitable Hilbert space.

The algebraic approach, which we shall now discuss, will indeed follow this logic hence its description can be separated in two parts implementing the two mentioned steps

3.2.1 The Borchers-Uhlmann and the Weyl algebra of observables

Since we ultimately want to quantize a free field theory, our starting point will still be the Klein-Gordon field (2.2), though we still stress that it should be intended as a toy model and not as the only case where the techniques described are applicable. Let us briefly recall that $\mathcal{S}(M)$ is the space of solutions of the Klein-Gordon equation out of smooth compactly supported initial data and that it is a symplectic space which can be interpreted as the phase space of the theory. Classical observables are maps from $\mathcal{S}(M)$ into $\mathbb{R}$, or equivalently from $C_0^\infty(M)$ into $\mathbb{R}$ if we recall the properties of E, the causal propagator. In between the classical observables we recall those of the form (2.7) whose Poisson brackets yield (2.11). In the Dirac prescription we simply promoted these objects to operators on a suitable Hilbert space in such a way that they satisfy a set of properties listed at the beginning of this chapter. Ultimately we want to keep this perspective although with some minor modifications.

As a matter of fact our quest will be to find a way to collect all our observables into an algebra $\mathcal{A}$ or, in other words, we look for a map $\varphi : C_0^\infty(M;\mathbb{C}) \to \mathcal{A}$ where the symbol φ is here used without any reference a priori with the field, which was previously indicated as Φ. The choice of $\mathcal{A}$ is not random but actually it must fulfil the requirements of the following definition:

Definition 3.2.1. *A collection $\mathcal{A}$ of elements is called a unital (associative) $*$-algebra if $\mathcal{A}$ is endowed with a (associative) product operation endowing it with an algebra structure and there exists $1 \in \mathcal{A}$ such that $a1 = 1a = a$ for all $a \in \mathcal{A}$. Furthermore it is also assigned $* : \mathcal{A} \to \mathcal{A}$ such that*

1. $(a + \lambda b)^* = a^* + \overline{\lambda} b^*$ *for all* $a, b \in \mathcal{A}$ *and for all* $\lambda \in \mathbb{C}$,
2. $(ab)^* = b^* a^*$ *for all* $a, b \in \mathcal{A}$,
3. $a^{**} = a$ *for all* $a \in \mathcal{A}$.

A further and useful specialization of this last definition is the following:

Definition 3.2.2. *A unital $*$-algebra $\mathcal{A}$ is called a C^*-algebra if it is endowed with a norm $\|\cdot\| : \mathcal{A} \to \mathbb{R}$ such that, for all $a, b \in \mathcal{A}$, $\|ab\| \leq \|a\|\|b\|$ and, furthermore, $\|a^* a\| = \|a\|^2$ for all $a \in \mathcal{A}$..*

A particular example of definition 3.2.1, say $\widetilde{\mathcal{A}}$, can be obtained constructing out of the set of generators $\{\varphi(f) \mid f \in C_0^\infty(M;\mathbb{C})\} \equiv \mathscr{D}(M)$. To wit, one introduces the free tensor algebra $T(\mathscr{D}(M)) \doteq \bigoplus_{n=0}^{\infty} \mathscr{D}(M)^{\otimes n}$, where the multiplication is individuated by the canonical isomorphism between $\mathscr{D}(M)^{\otimes k} \otimes \mathscr{D}(M)^{\otimes l}$ and $\mathscr{D}(M)^{\otimes(k+l)}$. The set of elements in $T(\mathscr{D}(M))$ should be understood only as finite sequences, *i.e.*, they are of the form $\bigoplus_{n=0}^{N<\infty} f_n$ where $f_n \in \mathscr{D}(M)^{\otimes n}$. Alas, this set is too big and, moreover, it has no information on the dynamics of the underlying field; to solve this problem, one can quotient $T(\mathscr{D}(M))$ by an ideal $\mathcal{I} \subset T(\mathscr{D}(M))$ which is the set of finite linear combinations of elements of the form abc with $a, c \in T(\mathscr{D}(M))$, while b (or b^*) is of one of the following forms:

- $\varphi(\lambda f + \mu f') - \lambda\varphi(f) - \mu\varphi(f')$,
- $\varphi(f)^* - \varphi(\overline{f})$,
- $\varphi(Pf)$ where P is as in (2.2),
- $[\varphi(f), \varphi(f')] - iE(f, f')id$.

All these elements can be recollected in the following lemma:

Lemma 3.2.1. *We have constructed the* **Borchers-Uhlmann**[2] *algebra, which is (homomorphic to) $\widetilde{\mathcal{A}} = \frac{T(\mathscr{D}(M))}{\mathcal{I}}$ where the $*$-operation is defined as $\varphi(f)^* = \varphi(\overline{f})$ and $\mathcal{I}$ is the $*$-ideal generated by the above*

[2] To be precise, the Borchers-Uhlmann algebra is usually meant only as the tensor algebra endowed with the $*$-operation and the quotient as the field algebra. Furthermore, in the original papers [4, 55] the space of testfunctions was $\mathcal{S}(\mathbb{R}^n)$, the set of rapidly decreasing functions on $\mathbb{R}^n$.

relations. Furthermore $\widetilde{\mathcal{A}}$ *is non trivial.*

From the above discussion, one can infer that $\widetilde{\mathcal{A}}$ plays indeed the role of a *field algebra* and thus the symbol φ can be correctly exchanged with Φ. Even though it is certainly rather easy to construct the Borchers-Uhlmann algebra and it fulfils all the properties we sought, we stress the existence of a second possible solution (the favourite of the author):

Definition 3.2.3. *Let us consider the symplectic space* $(\mathcal{S}(M),\sigma)$*. Then we call* **Weyl algebra** $\mathcal{W}(M)$ *the unique (up to* $*$*-isometries)* C^**-algebra which is unitary and whose generators are* $W(\Phi)$ *where* $\Phi\in\mathcal{S}(M)$ *and they fulfil*

$$W(\Phi)^* = W(-\Phi),\quad W(\Phi)W(\Phi') = e^{\frac{i}{2}\sigma(\Phi,\Phi')}W(\Phi+\Phi'),$$

for all $\Phi,\Phi'\in\mathcal{S}(M)$*.*

One could argue that the Weyl algebra is not really a map from $C_0^\infty(M)$ into suitable algebra elements. In order to solve this potential problem, we just recall that every $\Phi\in\mathcal{S}(M)$ can be written as $\Phi=E(f)$ for a suitable unique $f\in C_0^\infty(M)$. Furthermore if we recall (2.11) and the subsequent identity, we also know that $\sigma(\Phi,\Phi')=E(f,f')$ where $\Phi=E(f)$ and $\Phi'=E(f')$. In other words we have identified a $*$-homomorphism which associates to each generator $W(\Phi)$ another generator $V([f])$ such that

$$V([f])^* = V(-[f]),\quad V([f])V([f']) = e^{\frac{i}{2}E([f],[f'])}V([f]+[f']), \tag{3.9}$$

for all $[f],[f']\in C_0^\infty(M)/J$ where J stands for an equivalence relation such that $f\sim f'$ if there exists $g\in C_0^\infty(M)$ fulfilling $f-f'=Pg$. The formal interpretation of the generator $V([f])$ is of an exponentiated smeared field, *i.e.*, $e^{i\Phi([f])}$, but we stress that, in absence of an Hilbert space, this is only formal! Notice also that it is often customary to leave out the symbol of equivalence class and to simply write $V(f)$. We shall henceforth stick to this convention, because we feel that there is no risk for a potential confusion. Furthermore we will also work alternatively both with $W(\Phi)$ and with $V(f)$ depending on the convenience since it is understood that the two pictures are related by an algebra homomorphism.

The most important aspect of the Weyl algebra is possibly the set of all properties that it automatically embodies. Let us review them keeping in mind that, originally, in [21] these were first axiomatically required as necessary conditions that an algebra of observables should satisfy to be used in the quantization of a scalar field theory. Our perspective is instead different: we start from a classical field theory, we construct the space of solutions with its symplectic form and we associate to it a Weyl algebra. This procedure is unambiguous, it depends only on the global hyperbolicity of the spacetime (M,g) and it also yield:

1. for all open sets $\mathcal{O}\subset M$, there is an associated C^*-algebra $\mathcal{A}(\mathcal{O})$ constructed as the subalgebra of $\mathcal{W}(M)$ generated only by those $V(f)$ with *supp*$(f)\subset\mathcal{O}$. Hence **isotony** holds true, namely $\forall\mathcal{O}\subset\mathcal{O}'\subset M$, then $\mathcal{A}(\mathcal{O})\subset\mathcal{A}(\mathcal{O}')$ and the full algebra of observables is $\mathcal{A}=\bigcup\limits_{\mathcal{O}}\mathcal{A}(\mathcal{O})$.

2. $\mathcal{W}(M)$ encodes **causality** since if we consider $\mathcal{O},\mathcal{O}'\subset M$ such that $\mathcal{O}\cap(J^+(\mathcal{O}')\cup J^-(\mathcal{O}'))=\emptyset$, then $[\mathcal{A}(\mathcal{O}),\mathcal{A}(\mathcal{O}')]=0$, the reason being that $E(f,f')=0$, for *supp*$(f)\subset\mathcal{O}$ and *supp*$(f')\subset\mathcal{O}'$.

3. the algebra is **covariant** with respect to isometries (*i.e.*, these are implemented as algebra homomorphisms) since for any isometry $h:M\to M$, we can associate $\alpha_h:\mathcal{W}(M)\to\mathcal{W}(M)$ which maps $(\alpha_hV)(f)\doteq V(f\circ h^{-1})$. This is an algebra homomorphism since the symplectic form is constructed out of the metric and thus it is invariant under isometries.

4. the algebra is **primitive**, namely it admits an irreducible faithful representation. This holds true since all CCR-generated algebras fulfil this hypothesis and, hence, also $\mathcal{W}(M)$ thanks to our previous analysis.

We are left with just one last step: we need to represent our algebra on an Hilbert space!

3.2.2 States and the GNS theorem

The quest to find a representation either of the Weyl or of the Borchers-Uhlmann algebra on a suitable Hilbert space is the most difficult aspect of the algebraic approach to quantum field theory, even in the flat case. The example of ultrastatic spacetimes might lead to the false idea that, after all, one has just to engineer a suitable one-particle Hilbert space and, then, construct the associated Fock space via tensorialization. This would be the wrong interpretation, since, actually, one should read it in the other way round, namely we managed to find just one very specific example of an Hilbert space on which to represent the algebra of observables. On the contrary, there is really no reason why there should be a unique choice and one needs to take into account that the problem at hand might and, actually, must have many different solutions. At the same time, it is of course impossible to go on a quest to find all these possible Hilbert spaces one by one. To solve this impasse, it is possible to adopt an apparently completely different point of view, that is we shall not look directly for an Hilbert space but actually for a rather different object:

Definition 3.2.4. *We call (algebraic)* **state** *of a unital (topological) $*$-algebra $\mathcal{A}$ a continuous linear functional $\omega : \mathcal{A} \to \mathbb{C}$ such that*

$$a)\ \ \omega(e) = 1 \qquad\qquad b)\ \ \omega(a^*a) \geq 0 \quad \forall a \in \mathcal{A},$$

where e is the unit element of the algebra.

We stress that, in the given definition, condition *b)*, also known as *positivity* of the state, is a highly non linear condition and it is commonly the most difficult one to be verified. Therefore a reader should always pay attention to it and be warned that its implementation might not be straightforward. Barring this caveat, the real question concerns the relation of definition 3.2.4 with the picture of $\mathcal{A}$ being represented on an Hilbert space. An answer to this problem was first given by Naimark, Gelfand and Segal (GNS) and their theorem represents one if not the main building block of the whole algebraic approach. We shall now discuss its main aspects and, to this avail, we feel worth to separate it in some sub-propositions since we feel that the enunciation and the proof of a really "big fat theorem" might discourage a potential reader. Let us thus start showing that, whenever we are representing the algebra of observables on an Hilbert space, we are actually defining a state:

Lemma 3.2.2. *Let $\mathcal{A}$ be any topological $*$-algebra and $\mathcal{H}$ an Hilbert space with inner product $(,)$, such that there exists a faithful strongly continuous representation $\pi : \mathcal{A} \to \mathcal{L}(\mathscr{D})$ where $\mathscr{D}$ is a dense subspace of $\mathcal{H}$ and where $\pi(a^*) = \pi(a)^*$ for all $a \in \mathcal{A}$. Then this individuates a positive state $\omega : \mathcal{A} \to \mathbb{C}$.*

Proof. Let $(,)$ be the natural inner product over $\mathcal{H}$ and let $\psi \in \mathscr{D}$ be any element such that $\|\psi\|_{\mathcal{H}} = 1$. Then we can define $\omega_\psi(a) \doteq (\psi, \pi(a)\psi)$. Per construction it is linear and continuous since π is strongly continuous. Furthermore faithfulness and the standard property of an algebra representation that $\pi(ab) = \pi(a)\pi(b)$ yield that $\omega_\psi(e) = 1$. To conclude we notice that

$$\omega_\psi(a^*a) \doteq (\psi, \pi(a^*a)\psi) = (\psi, \pi(a)^*\pi(a)\psi) = \|\pi(a)\psi\|^2_{\mathcal{H}} > 0,$$

where in the before last equality we exploited that $\pi(a^*) = \pi(a)^*$.
□

As we can directly infer from the proof, the assignment of $(\psi, \pi, \mathcal{H})$ allows to construct unambiguously a state in the algebraic sense and the overall result depends strictly on the choice of ψ. Notice that one is not forced to choose a single element of norm 1 but it is possible to consider a linear combination of vectors in $\mathcal{H}$, say $\sum\limits_i \psi_i$ such that $\|\sum\limits_i \psi_i\| = 1$. Yet, the whole discussion would be moot, if the following theorem would not hold true:

Theorem 3.2.1. *Let ω be an algebraic state of a unital $*$-algebra $\mathcal{A}$. Then there exists $\mathscr{D}$, a dense subspace of an Hilbert space $\mathcal{H}$ with inner product $(,)$, as well as a representation $\pi : \mathcal{A} \to \mathcal{L}(\mathscr{D})$ and a norm 1 vector*

$\Omega \in \mathscr{D}$, *such that*

$$\omega(a) = (\Omega, \pi(a)\Omega), \qquad \mathscr{D} = \{\pi(a)\Omega, \ \ \forall a \in \mathcal{A}\}.$$

The set $(\mathscr{D}, \pi, \Omega)$ *is the* **GNS triple** *which is unambiguously determined up to unitary equivalence.*

Proof. The first step consists of endowing $\mathcal{A}$ with an inner product as

$$(a, b) \doteq \omega(a^* b). \qquad \forall a, b \in \mathcal{A}$$

This is per construction linear and positive semidefinite since $(a, a) = \omega(a^* a) \geq 0$. We thus need to check only the conjugate symmetry, namely that for all $a, b \in \mathcal{A}$ it holds $(a^*, b) = \overline{(b^*, a)}$. To this avail one needs to take into account the following two identities:

$$\begin{cases} 4a^*b = (a+b)^*(a+b) - (a-b)^*(a-b) - i(a+ib)^*(a+ib) + i(a-ib)^*(a-ib) \\ 4b^*a = (a+b)^*(a+b) - (a-b)^*(a-b) + i(a+ib)^*(a+ib) - i(a-ib)^*(a-ib) \end{cases}.$$

Since the positivity requirement for an algebraic state only yields ≥ 0, we need to single out the vanishing elements; hence we can introduce the subset $\mathcal{I} = \{a \in \mathcal{A} \mid \omega(a^*a) = 0\}$. This is a left ideal of the algebra $\mathcal{A}$ since

- it is closed, that is, for all $a, b \in \mathcal{I}$ and for all $\alpha, \beta \in \mathbb{C}$, also $\alpha a + \beta b \in \mathcal{I}$ since

  $$\omega\left((\alpha a + \beta b)^*(\alpha a + \beta b)\right) = |\alpha|^2 \omega(a^*a) + |\beta|^2 \omega(b^*b) + \overline{\beta}\alpha\omega(b^*a) + \overline{\alpha}\beta\omega(a^*b) = 0,$$

  where the last identity descends both from the definition of $\mathcal{I}$ and from the Cauchy-Schwartz inequality which, at a level of state, translates as

  $$|\omega(a^*b)|^2 = |(b^*a, a^*b)|^2 \leq (a^*, a)(b^*, b) = \omega(a^*a)\omega(b^*b) = 0.$$

- it holds the following: for all $a \in \mathcal{I}$ and $b \in \mathcal{A}$, then

  $$|\omega((ba)^*(ba))|^2 = |\omega(a^*b^*ba)|^2 \leq \omega(c^*c)\omega(a^*a) = 0,$$

  where $c \doteq a^*b^*b$. Hence $ba \in \mathcal{I}$.

We can thus define $\mathscr{D} \doteq \frac{\mathcal{A}}{\mathcal{I}}$ where the latter is the set of equivalence classes $[a]$ such that $a \sim a'$ if and only if there exists $b \in \mathcal{I}$ such that $a' = a + b$. Hence $\mathscr{D}$ is still a vector space and we can endow it with the positive definite scalar product $(,)_{\mathscr{D}}$ such that

$$([a], [b])_{\mathscr{D}} \doteq (a, b), \quad \forall [a], [b] \in \frac{\mathcal{A}}{\mathcal{I}}$$

where a and b are any representative of the equivalence classes $[a]$ and $[b]$ respectively. We shall henceforth omit the subscript in the inner product and we also notice that $(\mathscr{D}, (,))$ can be closed to an Hilbert space $\mathcal{H}$. The representation of the algebra can be induced via left multiplication namely we introduce $\pi : \mathcal{A} \to \mathcal{L}(\mathcal{H})$ such that, for all $b \in \mathcal{A}$ and for all $[a] \in \mathscr{D}$

$$\pi(b)[a] \doteq [ba].$$

Furthermore, since the algebra is unital, we can set up $\Omega \doteq [1]$ and, thus,

$$\omega(a) = \omega(1a) = ([1], [a]) = ([1], \pi(a)[1]) = (\Omega, \pi(a)\Omega),$$

which concludes the identification of the GNS triple. Let us now tackle uniqueness. Suppose that one can found another realization of ω as $(\mathscr{D}', \pi', \Omega')$ and let us introduce the operator $U : \mathscr{D} \to \mathscr{D}'$ such that

$$U(\pi([a])\Omega) \doteq \pi'([a])\Omega'.$$

This is well-defined since $\pi([a])\Omega = 0$ only if $\omega(a^*a) = 0$ which also yields $\pi'([a])\Omega' = 0$. Furthermore, per direct inspection of its definition, U preserves the scalar product and it is invertible on $\mathscr{D}$, thus it is extensible to a unitary operator from $\mathcal{H}$ to $\mathcal{H}'$, the Hilbert space constructed out of the second triple via closure with respect to $(,)$. In other words this means that $\Omega' = U^{-1}\Omega$ and that the defining relation for U can also be read as $\pi'([a]) = U\pi(a)U^{-1}$. This is nothing but the statement that the two representations, π and π' are unitary equivalent.
□

From the point of view of our analysis of quantum field theory over curved spacetimes it sounds thus reasonable to take the Weyl algebra $\mathcal{W}(M)$ as the natural counterpart of $\mathcal{A}$ in the last theorem and to ask ourselves if we can better characterize the class of states we associate to it. Actually, since we claimed several times that the algebraic approach is reproducing the results of the standard approaches to quantum field theory, the first step is to show that we can suitably choose ω in such a way that the GNS triple yields a Fock space playing the role of $\mathcal{H}$ and that, in a suitable sense, the representation of the observables thereon can be interpreted in terms of "creation" and "annihilation" operators. To this avail we start from an apparently different context, but it will be afterwards clear that, with the following steps, we are not far from solving our problems. Hence let us start with a slightly different scenario following §5 in [7], namely we consider:

- a complex Hilbert space $\widetilde{\mathcal{H}}$ endowed with the inner product $(,)$. Notice that, in principle we could just consider pre-Hilbert spaces,
- each complex Hilbert space is a strongly non-degenerate symplectic space, where $\sigma(\psi,\psi') \doteq Im(\psi,\psi')$ for all $\psi,\psi' \in \widetilde{\mathcal{H}}$,
- the natural Weyl algebra $\mathcal{W}(\widetilde{\mathcal{H}})$ associated to $(\widetilde{\mathcal{H}},(,))$ as the one whose generators $V(\psi)$ fulfil for all $\psi,\psi' \in \widetilde{\mathcal{H}}$
$$V(\psi)^* = V(\overline{\psi}), \quad V(\psi)V(\psi') = e^{\frac{i}{2}Im(\psi,\psi')}V(\psi+\psi').$$

As a preparatory tool we need the following

Definition 3.2.5. *Let ω be a state for $\mathcal{W}(\widetilde{\mathcal{H}})$ and let $(\pi,\mathscr{D},\Omega)$ be its GNS triple. Then the representation π is called* **regular** *if the one parameter family of operators $\pi[V(t\psi)]$ with $t \in \mathbb{R}$ and $V(\psi) \in \mathcal{W}(\widetilde{\mathcal{H}})$ are strongly continuous operators*[3] *for all $\psi \in \widetilde{\mathcal{H}}$.*

In other words the operators $\pi[V(t\psi)]$ admit self-adjoint generators which we can call $\Phi_\pi(\psi)$ and we can use them both to construct creation and annihilation operators and to make sense of the formal interpretation of the generators of the Weyl algebra as exponentiated fields. Yet this might not suffice because we have no guarantee that the cyclic vector Ω lies in the domain of definition of these operators and, in this case, we would be hindered from defining

$$\omega(\Phi_\pi(\psi)\Phi_\pi(\psi')) \doteq (\Omega, \Phi_\pi(\psi)\Phi_\pi(\psi')\Omega)_{\mathscr{D}},$$

which would allow us to give a meaning to objects such as the n-point function of a field even if we start from a Weyl algebra. Yet, we should not be too worried and we should proceed one step at a time. Hence, first of all we still need to show that we can provide a characterization of ω whose GNS triple yield a Fock space. The solution follows in this scenario still form [6, 7]:

Lemma 3.2.3. *Let us consider the Weyl algebra $\mathcal{W}(\widetilde{\mathcal{H}})$ built out of elements ψ which are also element of an Hilbert space $\widetilde{\mathcal{H}}$. Then the GNS triple $(\mathscr{F}(\widetilde{\mathcal{H}}),\pi,\Omega)$ with $\Omega = (1,0,...0)$ is the so-called Fock regular*

[3] Actually this condition can be relaxed and it suffices to require that the map $t \mapsto \omega(V(t\psi))$ is continuous for all $\psi \in \mathcal{H}$ since it holds
$$\|(\pi(V(t\psi)) - \mathbb{I})\pi(V(\psi'))\Omega\|^2 = 2 - e^{itIm(\psi,\psi')}\omega(V(t\psi)) - e^{itIm(\psi,\psi')}\omega(V(-t\psi)).$$

representation of the (regular) state

$$\omega(V(\psi)) = (\Omega, \pi(V(\psi))\Omega) \doteq e^{-\frac{||\psi||^2_{\widetilde{\mathcal{H}}}}{4}}. \quad \forall \psi \in \widetilde{\mathcal{H}} \tag{3.10}$$

Alas, we cannot content ourselves with the last lemma since, when dealing with a quantum field theory over curved backgrounds, we construct the Weyl algebra out of a symplectic space which does not a priori possess a pre-Hilbert space structure. Hence we should find a way to make a contact with lemma 3.2.3. Actually we shall now show that, under rather mild extra assumptions, there is a natural prescription and our presentation will be based mostly on [38].

Let us thus consider any complex vector space S endowed with a weakly non-degenerate symplectic form σ and let us call $\mathcal{W}(S)$ the associated Weyl algebra. Let us furthermore assume that there exists an inner product on S, namely a positive, symmetric bilinear map $\mu : S \times S \to \mathbb{R}$ such that the following inequality holds true:

$$|\sigma(s_1, s_2)| \leq 2|\mu(s_1, s_1)|^{\frac{1}{2}} |\mu(s_2, s_2)|^{\frac{1}{2}}, \qquad \forall s_1, s_2 \in S. \tag{3.11}$$

Then, a state for the Weyl algebra can be unambiguously constructed as $\omega_\mu : \mathcal{W}(S) \to \mathbb{C}$ in such a way that

$$\omega_\mu(W(s)) \doteq e^{-\frac{\mu(s,s)}{2}}, \tag{3.12}$$

which, from time to time, is also referred to as *Gaussian state*. This proposal looks reasonable at first glance, but it is hard to understand why (3.11) has been imposed. Actually, this condition is necessary to guarantee the positivity of the state as one can infer for example computing $\omega(a^*a)$ where $a \doteq W(s_1) - 1 + iW(s_2) - i$ for all possible $s_1, s_2 \in S$. The last doubt, which can possibly arise from (3.12), concerns the connection with lemma 3.2.3 and it can be dispelled thanks to the following theorem, whose proof can be found in appendix A of [38]:

Theorem 3.2.2. *Let S be a vector space on which it is defined both a weakly non-degenerate symplectic form σ and an inner product μ. Then, there always exists a complex Hilbert space $\widetilde{\mathcal{H}}$ with inner product $(,)$ together with a real linear map $K : S \to \widetilde{\mathcal{H}}$ such that*

- *the range of $K + iK$ is dense in $\widetilde{\mathcal{H}}$,*
- *$\mu(s_1, s_2) = Re(Ks_1, Ks_2)$ for all $s_1, s_2 \in S$,*
- *$\sigma(s_1, s_2) = 2Im(Ks_1, Ks_2)$ for all $s_1, s_2 \in S$.*

Moreover the pair $(K, \widetilde{\mathcal{H}})$ is determined uniquely up to equivalence, that is any other pair $(K', \widetilde{\mathcal{H}}')$ fulfilling the above properties is related to the original one by an isomorphism $U : \widetilde{\mathcal{H}} \to \widetilde{\mathcal{H}}'$ such that $K' = UK$.

It is now manifest that we have actually introduced a Fock regular state since the second consequence, which the theorem yields, guarantees us that $\omega(W(s)) = e^{-\frac{\mu(s,s)}{2}} = e^{-\frac{\|K(s)\|^2_{\widetilde{\mathcal{H}}}}{2}}$ which, up to an irrelevant multiplicative factor $\frac{1}{2}$, coincides with (3.10). The advantage of this theorem is that it totally clarifies the connection with the hypotheses and with the conclusions of lemma 3.2.3. Yet, we still need to clarify how we recover the notion of annihilation and of creation operators. This will be a two-step process. The first concerns the construction of these operators and of their domain of definition and this can be handled via the following theorem (the proof is in §5.2 of [7]), here adapted to the language of theorem 3.2.2:

Theorem 3.2.3. *Let $\mathcal{W}(S)$ be the Weyl algebra with a regular quasifree state ω whose one-particle structure is given by the pair $(K, \mathcal{H})$. Let $\phi_\pi(Ks)$ be the generators of $\pi[W(ts)]$. Then*

a) for all $s \in S$, $\phi_\pi(Ks)$ and $\phi_\pi(iKs)$ have a set of common analytic vectors for all finite-dimensional subspaces of $\mathcal{H}$, *b) the operators*

$$a_\pi(Ks) \doteq \frac{1}{\sqrt{2}}\left[\phi_\pi(Ks) + i\phi_\pi(iKs)\right], \tag{3.13}$$

$$a^*_\pi(Ks) \doteq \frac{1}{\sqrt{2}}\left[\phi_\pi(Ks) - i\phi_\pi(iKs)\right], \tag{3.14}$$

are the annihilation and creation operators defined for all f and

$$dom(a_\pi(Ks)) = dom(\phi_\pi(Ks)) \cap dom(\phi_\pi(iKs)) = dom(a^*_\pi(Ks))$$

c) both $a_\pi(Ks)$ and $a^*_\pi(Ks)$ are densely defined and closed. Furthermore $(a_\pi(Ks))^* = a^*_\pi(Ks)$ and for all $\widetilde{\psi} \in dom(a_\pi(Ks))$, it holds

$$\|\phi_\pi(Ks)\widetilde{\psi}\|^2 + \|\phi_\pi(iKs)\widetilde{\psi}\|^2 = 2\|a_\pi(Ks)\widetilde{\psi}\|^2 + \|Ks\|^2\|\widetilde{\psi}\|^2.$$

The last theorem is unfortunately not conclusive because it allows only to construct the annihilation and the creation operators, but as we mentioned before we want to make sense of the n-point function and hence of their action on the cyclic vector Ω of the GNS triple. To this avail, let us work under the hypotheses of theorem 3.2.3; then, we can additionally require that the state ω is of class C^m if the map $t \mapsto \omega(W(ts))$ is m-times strongly differentiable for all possible choices of $s \in S$. In this case it holds that $\pi[W(ts)]\Omega$ is m-times strongly differentiable and that $\Omega \in Dom(\phi_\pi(Ks)^m)$ for all $s \in S$. Particularly, for our purposes, it would be interesting to consider the case $m = \infty$, or in other words C^∞*-states*, since the following defining relation is meaningful:

$$\omega_n(\phi_\pi(Ks_1)...\phi_\pi(Ks_n)) \doteq (\Omega, \phi_\pi(Ks_1)...\phi_\pi(Ks_n)\Omega)_{\mathscr{D}}. \quad \forall n \in \mathbb{N}$$

In between the class of all the smooth states, there is a certain subset of particular interest since they lead to a rather useful notion, namely **quasi-free states**, which we shall now characterize. To this avail we first need the notion of a *truncated n-point* correlation function: let us assume that the hypotheses of definition 3.2.5 are fulfilled; hence the following expressions are meaningful:

$$\left.\begin{array}{r} \omega_T(\phi_\pi(Ks)) \doteq \omega(\phi_\pi(Ks)), \\ \omega_T(\phi_\pi(Ks_1)\phi_\pi(Ks_2)) \doteq \omega(\phi_\pi(Ks_1)\phi_\pi(Ks_2)) - \omega(\phi_\pi(Ks_1))\omega(\phi_\pi(Ks_2)), \\ ... \end{array}\right\}, \tag{3.15}$$

and so on and so forth for the higher n-point functions with $n \geq 2$.

Definition 3.2.6. *We say that a state ω on $\mathcal{W}(S)$ is* **quasi-free** *if it is analytic, that is the map $t \mapsto \omega(V(ts))$ is analytic in a neighbourhood of the origin for all $s \in S$, and its truncated n-point functions vanish for all $n > 2$.*

Consequence 3.2.0. We have learned that, whenever we have a symplectic space (S, σ) with σ weakly non degenerate we can associate a Weyl algebra $\mathcal{W}(S)$. $S = C^\infty_0(M)$ or, equivalently, $S = \mathcal{S}(M)$ are just two of the many possible cases. Furthermore, if we endow S with an inner product μ, then the state (3.12) is quasi-free (see [38]) and, thus, it admits a Fock representation. In this setting, we can make sense of the notion of n-point function and (3.15) guarantees us that the only contribution comes from the even terms. Furthermore, since ω is quasi-free all n-point functions with $n > 2$ can be computed out of the case $n = 2$, that is

$$\omega_2(\phi_\pi(Ks_1)\phi_\pi(Ks_2)) \doteq (\Omega, \phi_\pi(Ks_1)\phi_\pi(Ks_2)\Omega)_{\mathscr{F}(\mathcal{H})},$$

or, equivalently,

$$\lambda(s_1, s_2) \equiv \omega_2(\phi_\pi(Ks_1)\phi_\pi(Ks_2)) = \frac{\partial^2}{\partial t \partial r}\left(\omega(W(rs_1 + ts_2)e^{-\frac{irt}{2}\sigma(s_1,s_2)})\right)\Big|_{r=t=0}.$$

Hence, if we control the behaviour of the 2-point function for a quasi-free state on the Weyl algebra, we are actually controlling all possible cases and this will play a pivotal role in the next section when looking for physically sensible algebraic states.

3.3 Hadamard states

The analysis of the previous section has clearly underlined that the set of possible states of any fixed algebra of observables can indeed be enormous and it is certainly highly unlikely that all possible choices yield a GNS triple which we could call "physically reasonable". It is therefore necessary to find a suitable set of conditions to be imposed on ω in order to considered it as a viable physical state; it is not an exaggeration to claim that, from a technical point of view, this is possibly the most difficult aspect of the algebraic formulation of quantum field theory and its positive solution has been found only through a lot of hard work. Most of it has followed a sort "trial and error" pattern which lead to the recognition of some fixed requirements that ω must not violate.
From a general perspective, we can a priori require that a good physical state:

- yields finite quantum fluctuations of the expectation value of the measured observables, such as, to quote a relevant example, the smeared components of the stress-energy tensor $T_{\mu\nu} \doteq \frac{1}{\sqrt{|g|}}\frac{\delta S}{\delta g^{\mu\nu}}$,
- it mimics the ultraviolet behaviour of the Minkowski vacuum. From a very heuristic perspective, this condition translates at a level of quantum fields, the pictorial idea that, as we probe with higher energies (UV regime) smaller and smaller distances in a physical system, we are getting close to perform measurements at a fixed spacetime point, where it is known that it is always possible to find an orthonormal reference frame for which the metric is the flat Minkowskian one.

Yet this last pictorial property is simply recalling us that, when we evaluate the product of two or more observables, such as for example fields, inserted at two or more different spacetime points, say x and y, it is important to control the structure of the arising singularities when $x \to y$ when evaluated on ω. This is the building block for the construction of the so-called Wick powers of a field, the natural objects, one uses when one wants to discuss perturbative interactions. Although this is certainly not the standard picture, we concur with [24] that it can be yet a rather useful starting point to introduce the notion of "microlocal spectrum condition" and of "Hadamard states" as physical states. To this avail let us have a quick look at what happens on Minkowski spacetime $(\mathbb{R}^4, \eta)$ and let us consider a massless free scalar field theory thereon. This is a special case of what we have already done for the case of ultrastatic spacetimes in section 3.1 and, thus, from that analysis we know that, for all $x, y \in \mathbb{R}^4$,

$$\Phi(x)\Phi(y) = \int\limits_{\mathbb{R}^3\times\mathbb{R}^3} \frac{d^3k}{(2\pi)^{\frac{3}{2}}}\frac{d^3k'}{(2\pi)^{\frac{3}{2}}}\frac{1}{2\sqrt{\omega\omega'}}\left(a(k)e^{-ikx} + a^*(k)e^{ikx}\right)\left(a(k')e^{-ik'y} + a^*(k')e^{ik'y}\right),$$

where a and a^* are the annihilation and creation operators, here fulfilling the commutation relations

$$[a(k), a^*(k')] = i\delta(k - k')\mathbb{I}.$$

In order to construct the Wick polynomials, the first step consists of introducing the normal ordered fields which is tantamount to rewrite the above product putting the creation operator a^* to the rightmost side whenever possible. If we apply this procedure and we call the result $:\Phi(x)\Phi(y):$, we can realize via a simple algebraic manipulation that

$$\Phi(x)\Phi(y) =: \Phi(x)\Phi(y) : + \int\limits_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3}\frac{e^{ik(x-y)}}{2\omega}\mathbb{I}, \tag{3.16}$$

where we notice that whenever $k = k'$ also $\omega = \omega'$ since it holds the dispersion relation $|k|^2 = \omega^2$ where $|\cdot|$ stands for the Euclidean modulus. Notice that the integral term in the above expression is singular and, if we evaluate this expression on the Poincaré vacuum ω_P, we realize that the $\omega_P(:\Phi(x)\Phi(y):)$ becomes a meaningful expression also when $x \to y$ since it is actually smooth. Therefore one usually defines the squared field at a point x or, in other words, the squared Wick power of Φ

$$\omega_P(:\Phi^2(x):) \doteq \lim_{x\to y} \omega_P(:\Phi(x)\Phi(y):).$$

It thus looks appropriate to require that physical reasonable states are those where such a procedure can be implemented and the idea can be clearly also applied to a curved background. After all, the message of the previous example is that, if one controls the singular structure of the two-point function, seen as a distribution in $\mathscr{D}'(M \times M)$, M being the underlying spacetime, it is possible to remove all the unwanted pathologies out of a suitable subtraction (also known as regularization).

In order to make this fuzzy idea concrete, we need beforehand a mathematical tool which allows us to discuss the singular structure of a distribution and to this avail we need to introduce the concept of wavefront set. Our discussion will try to focus on the main structures and on the main operations, one can perform with these objects; if a reader is interested in the more subtle aspects of this technique, we cannot but suggest him to read carefully [36], particularly §8. This whole branch of mathematics goes under the name of *microlocal analysis* and it is based upon the idea that the singularities of a distribution u can be studied via the properties of its Fourier transform.

As a starting point we shall motivate this assertion with two notable examples:

Example 1) we consider the scenario where no singularity is present, namely $u \in C_0^\infty(\mathbb{R})$. Notice that this choice is not by chance, but it is calibrated on the fact that we shall later discuss curved backgrounds, where concepts such as rapidly decreasing functions are not present since the underlying manifold M only locally looks like $\mathbb{R}^4$ and, thus, it is meaningless the idea to control the behaviour of a distribution at infinity. On the other hand the notion of smooth and compactly supported function is perfectly well defined on any spacetime and thus, for any (M, g), it is legitimate to talk about $\mathscr{D}(M) \equiv C_0^\infty(M)$ and about $\mathscr{D}'(M)$ as the space of continuous linear functional from $\mathscr{D}(M)$ into $\mathbb{R}$. That said, for any $u \in C_0^\infty(\mathbb{R})$, we consider

$$(1 + |k|^{2n})|\widehat{u}|(k) = \left|\widehat{\left[1 + \left(-\frac{\partial^2}{\partial x^2}\right)^n\right] u}\right|(k) \leq \int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi}} \left|\left[1 + \left(-\frac{\partial^2}{\partial x^2}\right)^n\right] u\right|(x) < \infty,$$

where $n \in \mathbb{N}$ and all the (in)equalities arise out of the standard properties of the Fourier transform except the last one, which is a consequence of the smoothness and of the compactness of u and of its derivatives up an arbitrary order. This short calculation guarantees us that $\widehat{u} \in \mathcal{S}(\mathbb{R})$.

Example 2) let us consider the prototype of a singular element of $\mathscr{D}'(\mathbb{R})$ namely $\delta(x)$, which is a meaningful object even if we replace $\mathbb{R}$ with an arbitrary spacetime M. Once we test the Fourier transform of the delta function over an arbitrary $f \in C_0^\infty(M)$, we end up with

$$\widehat{\delta}(f) = \int_{\mathbb{R}} dx \int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi}} e^{ikx} \delta(x) f(k) = \int_{\mathbb{R}} \frac{dk}{\sqrt{2\pi}} f(k),$$

which implies that $\widehat{\delta} = 1$ which is smooth but nowhere rapidly decreasing function.

These two examples somehow suggest how to recollect and to slightly generalize to arbitrary dimensions the used notions in a single body:

Definition 3.3.1. *For any $u \in \mathscr{D}'(\mathcal{O})$ where $\mathcal{O}$ is an open set of $\mathbb{R}^n$, then a pair $(x, k) \in \mathbb{R}^{2n}$ with $k \neq 0$ is called a* **regular direction** *of u if there exists $\phi \in C_0^\infty(\mathcal{O})$ with $\phi \neq 0$, a conic neighbourhood[4] V of k and a constant C_n for all $n \in \mathbb{N}$ such that*

$$|\widehat{\phi u}(k)| < \frac{C_n}{1 + |k|^n}, \quad \forall k \in V \;\wedge\; \forall n \in \mathbb{N}$$

where $|\cdot|$ is the Euclidean modulus. In other words $\widehat{\phi u}$ is rapidly decreasing along k in the conic region V.

[4] Note that a smooth manifold M is called *conic* if there exists a smooth map $C : \mathbb{R}^+ \times M \to M$ such that, for any point $p \in M$, there is an open subset $\mathcal{O} \subset M$ with $p \in \mathcal{O}$ and with $C(\mathcal{O}) = \mathcal{O}$ and there is a diffeomorphism $\psi : \mathcal{O} \to U \subseteq \mathbb{R}^n$ such that U is a cone of $\mathbb{R}^n$ and $\psi \circ C(\mathcal{O}) = t\psi(\mathcal{O})$ for a $t \in \mathbb{R}^+$.

This definition characterizes the regular directions of a distribution u and it can be immediately generalized to a curved background because, per definition, every manifolds is locally diffeomorphic to $\mathbb{R}^n$. Hence, in order to consider $u \in \mathscr{D}'(M)$ it suffice to apply the above definition to every open subset $\mathcal{O}$ whose image, up to the choice of a local chart ψ, is an open subset of $\mathbb{R}^n$. For the moment let us just keep in mind this idea and let us carry on the analysis on $\mathbb{R}^n$, though we shall came back to this point later on. Let us thus introduce the main mathematical object used in the definition of Hadamard states in the modern days:

Definition 3.3.2. *Let us consider $u \in \mathscr{D}'(\mathcal{O})$ where $\mathcal{O}$ is an open subset of $\mathbb{R}^n$. We call* **wavefront set** *of the distribution u*

$$WF(u) = \{(x,k) \in \mathcal{O} \times (\mathbb{R}^n \setminus \{0\}) \mid (x,k) \textit{ is not a regular point}\}.$$

Furthermore we call **singular support** *of u the set*

$$singsupp(u) = \{x \in \mathcal{O} \mid \exists k \in \mathbb{R}^n \textit{ such that } (x,k) \in WF(u)\}.$$

Let us now apply this definition to the above given examples to have a feeling of how the wavefront set actually looks like:

1. in this scenario we consider $u \in C_0^\infty(\mathbb{R}) \subset \mathscr{D}'(\mathbb{R})$ and, hence, if we look at definition 3.3.1 and we take any $\phi \in C_0^\infty(\mathbb{R})$, it holds that ϕu is still smooth and compactly supported. As we have seen above, the Fourier transform is always a rapidly decreasing smooth function and hence this yields

$$WF(u) = \emptyset. \quad \forall u \in C_0^\infty(\mathbb{R})$$

   Notice that, since we wish to control the decay properties of ϕu, the same result holds true if we consider u just to be smooth without any assumption on its support.

2. in this case, the distribution was $\delta(x) \in \mathscr{D}'(\mathbb{R})$ and, if we follow once more definition 3.3.1, we must consider $\phi\delta$ for all $\phi \in C_0^\infty(\mathbb{R})$. We notice that if $0 \notin supp(\phi)$ then $\widehat{\phi\delta} = 0$, whereas, in the other case, $\phi\delta = \phi(0)$ and the Fourier transform of a constant is never rapidly decreasing. Hence, according to definition 3.3.2,

$$WF(\delta) = \left\{(0,k) \in \mathbb{R}^2 \mid k \neq 0\right\}.$$

The next natural step is to consider how the wavefront set transforms under change of coordinates and the answer is rather important (see §8 of [36] where the following behaviour is actually proven and not assumed) as we shall immediately realize. Let us thus consider a distribution $u \in \mathscr{D}'(M)$ and $\mathcal{O}$ an open subset of a manifold M. Then there exists $\psi : U \to \mathcal{O}$, a diffeomorphism from $U \subseteq \mathbb{R}^n$ to $\mathcal{O}$. Therefore we can define

$$\psi^* u(f) \doteq u(f \circ \psi^{-1}) \quad \forall f \in C_0^\infty(U),$$

and this yields that $\psi^* u \in \mathscr{D}'(U)$. Hence we can define the wavefront set of u supported in $\mathcal{O}$ as

$$WF(u|_{\mathcal{O}}) = \left\{(x, kd(\psi^{-1})|_x) \in \mathcal{O} \times (\mathbb{R}^n \setminus \{0\}) \text{ such that } (\psi^{-1}(x), k) \in WF(\psi^* u)\right\}. \tag{3.17}$$

This last formula allows us to extend all the previous definitions to a distribution on a generic spacetime M once a coordinate system $\{\mathcal{O}_\alpha, \psi_\alpha\}$ has been chosen since, in this case, for any $u \in \mathscr{D}'(M)$, we construct its wavefront set as

$$WF(u) = \bigcup_{\mathcal{O}} WF(u|_{\mathcal{O}}),$$

where $WF(u|_{\mathcal{O}})$ is as in (3.17). Furthermore, if recall that the tangent and the cotangent space of a manifold at a point can be identified with $\mathbb{R}^n$, and since the behaviour of the wavefront set under action of a diffeomorphism is such that each k behaves like a covector, we can identify the set $\mathcal{O} \times (\mathbb{R}^n \setminus \{0\})$ as $T^*\mathcal{O} \setminus \{0\}$, where 0 stands for the zero section of the cotangent bundle, that is all elements of the form $(x, 0)$. From now on we will follow this interpretation and we shall also omit the map ψ and its differential $d\psi$ from all formulas since it is implicit that the wavefront set of a distribution in curved spacetime is meant as in (3.17).

The powerfulness of the definition of wavefront set descends above all from the set of all properties that it enjoys. It is certainly a useful exercise to prove all of them, but this would bring us far from our final goal and we leave an interested reader to [36] for a more complete exposition. Instead we shall list them and comment on their usefulness:

- the wave front set of $u \in \mathscr{D}'(M)$ is *empty* if and only if $u \in C^\infty(M)$. The implication $\Longleftarrow$ has been proved in the examples above, while $\Longrightarrow$ is a consequence of definition 3.3.1 which guarantees that, if $WF(u) = \emptyset$, all (x,k) are regular directions and thus $\widehat{\phi u} \in \mathcal{S}(\mathbb{R}^n)$ for all choices of ϕ. Since the Fourier transform maps $\mathcal{S}(\mathbb{R}^n)$ into itself, u must be smooth. From a physical point of view, this condition can be though as follows: if we consider the counterpart of the product $\Phi(x)\Phi(y)$ in a generic spacetime, this yields an element, say ω_2 of $\mathscr{D}'(M \times M)$. Hence, in order to construct the counterpart of $:\Phi(x)\Phi(y):$ on (M,g) and in order to give sense to the limit of $x \to y$ for all choices of $x, y \in M$, we need to identify a second distribution $H \in \mathscr{D}'(M \times M)$ such that $\omega_2 - H$ is smooth. This is the basis of the so-called *Hadamard regularization*.

- For every $u, u' \in \mathscr{D}'(M)$ and for all $\lambda, \lambda' \in \mathbb{C}$ it holds

$$WF(\lambda u + \lambda' u') \subseteq WF(u) \cup WF(u').$$

  Notice that this inclusion tells us that our control of the wavefront set under sum of distributions can be far from ideal. To understand the meaning of this assertion, just consider the trivial example $u = -u' = \delta$ and $\lambda = \lambda' = 1$. Then

$$\emptyset = WF(\delta - \delta) \subseteq WF(\delta) \cup WF(\delta) = WF(\delta),$$

  which is clearly a not so interesting relation.

One of the most important properties of a wave front set arises when we apply a (pseudo)differential operator P to u. In this setting let us consider still a local reference frame for (M,g) and let us write P as $P = \sum\limits_{|a|\leq m} \alpha_a(x)(iD)^a$ where a is a multi-index, the coefficients α_a are taken smooth for simplicity and D indicates a directional derivative. We call *principal symbol* of P the polynomial $p_m(x)$ which is obtained substituting iD with k, that is $p_m(x,k) = \sum\limits_{|a|=m} \alpha_a(x)k^a$. Most notably we can associate to the principal symbol the so-called **characteristic set** of P:

$$char(P) = \{(x,k) \in T^*M \setminus \{0\} \mid p_m(k,x) = 0\}. \tag{3.18}$$

This new object plays a key role in this further property of the wavefront set:

- for all (pseudo)differential operators P with smooth principal symbol and for all $u \in \mathscr{D}'(M)$

$$WF(Pu) \subseteq WF(u) \subseteq WF(Pu) \cup Char(P). \tag{3.19}$$

  This is an extremely important property because the first inclusion tells us that the wavefront set is not enlarged under action of a derivative operator, while the second one allows us, for example, to control the wave front set of a solution of a partial differential equation out of the structure of the principal symbol of the associated operator. This property will find its main application in the proof of the so-called theorem of propagation of singularities.

Theorem 3.3.1. *Let P be a (pseudo)differential operator P whose principal symbol p_m is a real smooth homogeneous polynomial of degree m. Then if $u \in \mathscr{D}'(M)$ and $Pu = f$ with $f \in C^\infty(M)$, then $WF(u) \subseteq p_m^{-1}(0)$ and it is conserved along the Hamiltonian flow of*

$$H_{p_m} \doteq \sum_j \frac{\partial p_m}{\partial k_j}\frac{\partial}{\partial x_j} - \frac{\partial p_m}{\partial x_j}\frac{\partial}{\partial k_j},$$

where j runs from 1 up to the dimension of M.

Notice that the theorem has not been stated here in its most general form above all with respect to the role of the source term f, but this is the formulation which is more useful when dealing with quantum field theory over curved backgrounds. An interested reader can find the not so easy proof in §6 of [23]. We wish instead to underline the powerfulness of the theorem since it basically allows us to gather information of the wavefront set of a solution of a certain differential equation by looking both at the principal symbol and at the integral curves of an Hamiltonian flow. To better understand it, let us look at two concrete examples which play a role in quantum field theory over curved spacetimes:

Example 1) Let us consider $P = \Box_g + V(x)$ where $V \in C^\infty(M)$, which contains (2.2) as a particular sub-case. The principal symbol is independent from V being $p_m(x,k) = g^{\mu\nu}k_\mu k_\nu$ whose associated characteristic set is via (3.18)

$$char(P) \equiv \mathcal{N} = \{(x,k) \in T^*M \mid g^{\mu\nu}k_\mu k_\nu = 0\}, \tag{3.20}$$

which is a bundle of null rays at $T^*_x M$. If we are interested in the wavefront set of $u \in \mathscr{D}'(M)$ which solves in a weak sense $Pu = 0$ then we can first apply (3.19) to get

$$\emptyset = WF(Pu) \subseteq WF(u) \subseteq \mathcal{N} \setminus \{0\}.$$

Furthermore also the hypotheses of theorem 3.3.1 are met and thus we also can conclude that the wavefront set of u is conserved along the curves $\gamma(\lambda) \in T^*M$ solving

$$\begin{cases} \frac{dx^\mu}{d\lambda} = g^{\mu\nu}k_\nu \\ \frac{dk_\mu}{d\lambda} = 0 \end{cases}.$$

These two equations yield that the vector $\frac{dx^\mu}{d\lambda}$ is parallel transported along the projection of $\gamma(\lambda)$ on M and k_μ is cotangent to such projected curve. Hence, if we combine the two results together, we get that, <u>if we know</u> that a point (x,k) lies in $WF(u)$ (in other words we have an initial condition for the above system of ODEs), then we know that all the set of points $(x(\lambda), k(\lambda)) \in WF(u)$ where $(x(\lambda), k(\lambda))$ means that there exists a null geodesic passing through x and with tangent vector k. Notice that $k(\lambda)$ is the parallel-transport of k along the said geodesic.

Example 2) Let us now consider still P as in the previous example, but we now look for $\omega_2 \in \mathscr{D}'(M \times M)$ which fulfils $P_x\omega_2 = P_y\omega_2 = 0$, the subscripts x and y referring to the first and to the second entry respectively. We are essentially looking at the two-point function of a real scalar field theory. The operators P_x and P_y can be read as $P \otimes \mathbb{I}$ and $\mathbb{I} \otimes P$ respectively. The principal symbols are

$$\begin{cases} P \otimes \mathbb{I} \longrightarrow p_m(x, k_x, y, k_y) = g^{\mu\nu}(x)(k_x)_\mu(k_x)_\nu \\ \mathbb{I} \otimes P \longrightarrow p'_m(x, k_x, y, k_y) = g^{\mu\nu}(y)(k_y)_\mu(k_y)_\nu \end{cases}.$$

The characteristic sets can be computed as in the previous example thus being

$$\begin{cases} char(P \otimes \mathbb{I}) = \mathcal{N} \times T^*M \setminus \{0\} \\ char(\mathbb{I} \otimes P) = T^*M \setminus \{0\} \times \mathcal{N} \end{cases}.$$

We can now apply (3.18) to conclude that

$$WF(\omega_2) \subseteq (\mathcal{N} \times T^*M \setminus \{0\}) \cap (T^*M \setminus \{0\} \times \mathcal{N}) \subseteq \mathcal{N} \times \mathcal{N}.$$

Notice that, in this case, the use of the propagation of singularities theorem would not lead to an improvement of the control of the wavefront set of the distribution, we are interested in.

After these two physically oriented examples, we can once more consider the scenario of Minkowski spacetime and of the massless scalar field living thereon. As a first step let us rewrite in terms of wavefront set the conditions leading to a well defined normal ordering and notion of Wick polynomials. As a first step we consider the vacuum state ω which is quasi-free and thus we can only look at the associated two-point functions, ω_2. Let us notice that, since each solution of the wave equation can be written as $E(f)$ where E is the causal propagator and $f \in C_0^\infty(\mathbb{R}^4)$, we can interpret via the kernel theorem ω_2 as giving rise to an element of $\mathscr{D}'(\mathbb{R}^4 \times \mathbb{R}^4)$ which we shall indicate as $\omega_2(x,y)$. Since it satisfies the equation of motion in both entries we can also apply the results of example 2). Furthermore we recall that, in order to have a well-defined regularized squared scalar field, it was necessary to subtract the vacuum contribution which can be still seen as originating from an element of $\mathscr{D}'(\mathbb{R}^4 \times \mathbb{R}^4)$ which we shall indicate as $\omega_{vac}(x,y)$. Hence

$$WF(\omega_2 - \omega_{vac}) = \emptyset \Longrightarrow WF(\omega_2) = WF(\omega_{vac}),$$

and, from example 2),

$$WF(\omega_2) = WF(\omega_{vac}) \subseteq \mathcal{N} \times \mathcal{N}.$$

This inclusion can be slightly refined if we consider two arbitrary functions $\phi_1(x)$ and $\phi_2(y)$ both in $C_0^\infty(\mathbb{R}^4)$ and we construct $\phi(x,y) \doteq \phi_1(x)\phi_2(y)$ so that we can compute

$$\widehat{\phi\omega_{vac}}(k_x,k_y) = \int\limits_{\mathbb{R}^3} \frac{d^3\widetilde{k}}{(2\pi)^3}\frac{1}{2\omega}\widehat{\phi_1}(\widetilde{k}-k_x)\widehat{\phi_2}(\widetilde{k}+k_y),$$

where we inserted the expression of $\omega_{vac}(x,y)$ from (3.16) and where k is a future-pointing on-shell vector (*i.e.*, $\eta^{\mu\nu}k_\mu k_\nu = 0$). A direct inspection of the integrand shows that both ϕ_1 and ϕ_2 are rapidly decreasing functions and, thus, the main contribution to this integral comes form the points where $\widetilde{k}-k_x$ and $\widetilde{k}+k_y$ are contemporary small enough. In other words the integral must vanish as k_x and k_y diverge in an open conic neighbourhood of respectively of $\mathbb{R}^4_- \times \mathbb{R}^4$ and $\mathbb{R}^4_+ \times \mathbb{R}^4$, the subscripts $\pm$ respectively indicating that $k_0 > 0$ and $k_0 < 0$. Thus (x,k_x,y,k_y) is a regular direction if k_x is zero or past-directed ($k_x \lhd 0$) or if k_y either vanishes or is future directed ($k_y \rhd 0$). To summarize, if we introduce

$$\mathcal{N}^+ \doteq \{(x,k_x) \in \mathcal{N} \mid k_x \rhd 0\},$$

and

$$\mathcal{N}^- \doteq \{(y,k_y) \in \mathcal{N} \mid k_y \lhd 0\},$$

it holds that

$$WF(\omega_{vac}) \subset \mathcal{N}^+ \times \mathcal{N}^-.$$

Since all these considerations can be without effort translated into a curved background, it looks natural to extend them therein and thus we look for states whose two-point function mimics the ultraviolet behaviour of the Poincaré vacuum:

Definition 3.3.3. *A quasi-free state ω for a $*$-algebra of observables on a spacetime (M,g) is said to obey the μ**SC - microlocal spectrum condition** if its associated two-point function ω_2 seen as a distribution in $\mathscr{D}'(M \times M)$ fulfils*[5]

$$WF(\omega_2) \subset \mathcal{N}^+ \times \mathcal{N}^-.$$

From a physical perspective this definition has a clear interpretation since it mimics locally the idea that the first component of the two-point function of a ground state in a curved background must have a positive frequency while the second has a negative one. Yet, from a mathematical point of view, the microlocal spectrum condition appears as a necessary one to define a reasonable notion of a ground state on a spacetime (M,g) since the set of all states fulfilling this requirement might not behave well even under simple operations like the sum. To understand what it is meant, one can just think of a trivial example of two distributions

[5]To be precise the microlocal spectrum condition was originally formulated in [8] as a condition on the wavefront set of the n-point function which is automatically satisfied by all quasi-free Hadamard state. Yet, in [51], it was shown that this "enlarged" definition coincides with the one here given and thus the given statement is the most general possible.

$u = \lambda\delta \in \mathscr{D}'(\mathbb{R})$ and $v = \lambda'\delta \in \mathscr{D}'(\mathbb{R})$ with $\lambda, \lambda' \in \mathbb{R}$. While their wavefront set has already been computed at the beginning of the section, being the one of the δ-function, we notice that $WF(u+v)$ strongly depends on the chosen multiplicative constants, to the extent that only if $\lambda = -\lambda'$ the result is the empty wavefront set, while, in all other cases, it coincides again with that of the δ-function. If such an example mimics what happens summing up two states fulfilling the μSC, then definition 3.3.3 would be rather moot. Luckily enough this is not the case, as it was proven in two remarkable papers by Radzikowski [47, 48]:

Proposition 3.3.1. *Let ω and ω' be two quasifree states obeying the microlocal spectrum condition. Then*

$$WF(\omega_2 - \omega'_2) = \emptyset,$$

that is $\omega_2(x,y) - \omega'_2(x,y) \in C^\infty(M \times M)$.

This result is really remarkable because it entails that the singular structure of all states fulfilling the microlocal spectrum condition must be the same. Let us now try to exploit this proposition in the framework of quantum field theory. As a first step, let us recall that one of the building blocks of the quantization scheme we implemented (see the beginning of chapter 3 in particular) is that the commutator $[\phi(x), \phi(y)] = iE(x,y)$ where we omit the tilde since from now on we shall only speak about operators and no confusion can hence arise. For a quasifree state ω this means that

$$\omega_2(x,y) - \omega_2(y,x) = iE(x,y),$$

hence the antisymmetric part coincides with the causal propagator, whose wave front set was computed already in [23] (see also [47])

$$WF(E) = \{(x, k_x, y, k_y) \in T^*(M \times M) \setminus \{0\},\ |\ (x,k_x) \sim (y,k_y)\}\,, \tag{3.21}$$

where $\sim$ means that there exists a null geodesic connecting x and y whose tangent vector is k_x and k_y respectively and k_y is the obtained from k_x via parallel transport along the geodesic. Notice that there is no contradiction in talking about tangent vectors even if $k_x \in T^*_x M$ since we are simply considering the element of $T_x M$ whose components are $g^{\mu\nu}(k_x)_\mu$. Furthermore we can prove the following lemma:

Lemma 3.3.1. *Let ω be a quasifree state fulfilling the microlocal spectrum condition. Then*

$$WF(\omega_2) = WF(E) \cap (\mathcal{N}^+ \times \mathcal{N}^-).$$

Furthermore

$$WF(\omega_2) = \left\{(x, k_x, y, -k_y) \in T^*(M \times M) \setminus \{0\},\ |\ (x,k_x) \sim (y,k_y) \text{ and } (x,k_x) \in \mathcal{N}^+\right\}.$$

Proof. Let $\widetilde{\omega}_2(x,y) \doteq \omega_2(y,x)$. Then definition 3.3.3 entails that

$$WF(\widetilde{\omega}_2) \subseteq \mathcal{N}^- \times \mathcal{N}^+, \tag{3.22}$$

and we also know that $\omega_2 - \widetilde{\omega}_2 = iE$. This implies via the behaviour of the wavefront set under sum of distributions that

$$WF(E) \subseteq WF(\omega_2) \cup \mathcal{N}^- \times \mathcal{N}^+,$$

while, at the same time, we can also write

$$WF(E) \subseteq WF(\omega_2) \cup WF(\widetilde{\omega}_2) \subseteq WF(\widetilde{\omega}_2 + iE) \cup WF(\widetilde{\omega}_2) \subseteq WF(E) \cup \mathcal{N}^- \times \mathcal{N}^+.$$

The intersection of the first inclusion with $\mathcal{N}^+ \times \mathcal{N}^-$ yields $WF(E) \cap \mathcal{N}^+ \times \mathcal{N}^- \subseteq WF(\omega_2)$ while the intersection of the second with $\mathcal{N}^+ \times \mathcal{N}^-$ yields $WF(\omega_2) \subseteq WF(E) \cap \mathcal{N}^+ \times \mathcal{N}^-$ once we recall (3.22) and the μSC. This is indeed the first sought result, while the second arises from it plugging in (3.21). □

The importance of this lemma and, above all, of proposition 3.3.1 is that one can summarize the discussion claiming

Definition 3.3.4. *A quasifree state ω in a spacetime (M, g) is called a* **physical state** *or an* **Hadamard state** *if its two-point function obeys the microlocal spectrum condition*[6].

It is obvious to wonder why the name "Hadamard" and, more importantly, why an important concept such as that of a physical state came so late since the papers of Radzikowksi are of mid nineties. As a matter of fact the problem of finding "well-behaved" states in a curved background is an old one and the quest to characterize them is a story full of sorrows and of joys. We shall not summarize it here, though we wish to stress that an account of all the results, obtained before the papers of Radzikowski, can be found in [38] and we shall content ourselves with a few remarks. In earlier times it was believed that a physical state should satisfy the so-called *global Hadamard form* (see for example [38] or definition 3.4 in [48]), a complicated and unwieldy statement on the global structure of the two-point function of a quasifree state. From a merely practical point of view, such definition was almost impossible to check in concrete example and the proof in [47, 49] that it is actually equivalent to the microlocal spectrum condition allowed to use techniques of analysis to actually show that certain states were indeed Hadamard ground states(see for example [16, 17, 18, 19, 41, 42, 43, 50]).

3.3.1 Hadamard recursion relations

It is thus natural to wonder how people do and did explicit computations using Hadamard states and we shall devote this last section to this topic. The answer is actually quite simple, namely it was possible to show that a state obeying the global Hadamard form assumed locally a rather easy form which is the so-called *local Hadamard form*, which we shall now discuss. To this avail we need a few extra-ingredients:

- For every point p in a manifold M, it is possible to define an *exponential map* $\exp : T_pM \to M$ whose properties can be found for example in [39, 40] and, in the case of Lie groups it boils to the usual map from the algebra to the group itself. A most notable characteristic is that for every point p, there exists a neighbourhood $\mathcal{O} \ni p$ where exp is a local diffeomorphism. This is called the *normal neighbourhood* of p. Furthermore we also need that, for all points $q, r \in \mathcal{O}$, such that $q \in J^+(r)$, there exists a normal neighbourhood containing $J^+(r) \cap J^-(q)$. In this case $\mathcal{O}$ is called a *causal normal neighbourhood* and it always exists if (M, g) is globally hyperbolic (see Lemma 2.2 in [38]).

- in $\mathcal{O}$ it is possible to define a notion of geodesic distance also known as *Synge's function* (see for example the enlightening review of Poisson [46]), as

$$\sigma(x, y) = \int_{\lambda_0}^{\lambda_1} d\lambda \, \frac{\lambda_1 - \lambda_0}{2} g_{\mu\nu} t^\mu t^\nu,$$

  where t is the tangent vector of the geodesic $\lambda \mapsto \gamma(\lambda)$ connecting $x = \gamma(\lambda_0)$ and $y = \gamma(\lambda_1)$ under the hypothesis that $x, y \in \mathcal{O}$. Notice that σ is actually the halved squared geodesic distance, which is always well-defined between causally related points. Furthermore the factor $\frac{1}{2}$ is not a universally accepted convention and one should be careful to consult the relevant literature.

- whenever M is a globally hyperbolic spacetime, we can also find a global temporal function $T(x)$, that is a map $T : M \to \mathbb{R}$ which increases when evaluated on future directed causal curves. Its construction is discussed for example in [2].

[6] It is possible to prove that, for a given free field scalar theory, a Hadamard state can be constructed making a clever use of a deformation argument first introduced in [27]. A complete different problem arises if one wants also to implement an action of the background isometries on a Hadamard state. If successful, this procedure yields a natural and physical reasonable notion of *ground state*, though their existence is not a priori guaranteed. Actually, in four dimensional de Sitter spacetime, it is known that no such state can be individuated if one considers a massless minimally couple scalar field theory, *i.e.*, with $m = \xi = 0$.

With these new ingredients, we can write the local Hadamard form from the integral kernel of the two-point function $\omega_2 \in \mathscr{D}'(M \times M)$ of a quasifree state ω on a globally hyperbolic spacetime (M, g):

$$\omega_2(x,y) \doteq \frac{U(x,y)}{4\pi[\sigma(x,y) + i\epsilon(T(x)-T(y)) + \epsilon^2]} + V(x,y) \ln \frac{\sigma(x,y) + i\epsilon(T(x)-T(y)) + \epsilon^2}{\lambda^2} + W(x,y), \quad (3.23)$$

where ϵ is here meant as a regularization parameter, while λ is a reference length necessary to make the argument of the logarithm function dimensionless. The functions U, V, W are all elements in $C^\infty(\mathcal{O}\times\mathcal{O})$ There are some further mathematical subtleties in the definition of the local Hadamard form and we recommend a reader interested in them to consult [49, 56]. As far as our purposes are concerned, we split $\omega_2(x, y)$ in the sum of two components, namely $H(x, y) + W(x, y)$, where the first term is called the *Hadamard parametrix*:

$$H(x,y) \doteq \frac{U(x,y)}{4\pi[\sigma(x,y) + i\epsilon(T(x)-T(y)) + \epsilon^2]} + V(x,y) \ln \frac{\sigma(x,y) + i\epsilon(T(x)-T(y)) + \epsilon^2}{\lambda^2}. \quad (3.24)$$

This function encodes the full singular structure of the two-point function of a quasifree Hadamard state and furthermore we also know that

$$\omega_2(x,y) - \omega_2(y,x) = iE(x,y) = H(x,y) - H(y,x),$$

where the first equality descends from the employed quantization scheme, while the second can be seen as a sort of definition or of imposed condition on the Hadamard parametrix.

Consequence 3.3.0. A physically sensible quasifree ground state ω for a quantum field theory over a globally hyperbolic spacetime (M, g) has a two-point function ω_2 which satisfies the microlocal spectrum condition. This also implies that in a sufficiently small neighbourhood of any $p \in M$ the integral kernel $\omega_2(x, y)$ assumes the local Hadamard form (3.23).

From a practical point of view, also the local Hadamard form might look pretty much useless since it depends on three unknown function U, V, W whose explicit form is necessary in order to use them in concrete calculations. We shall now show that plenty of informations on U and V can be actually derived just from the classical equations of motion. More precisely let us recall that ω_2 satisfies in a weak sense

$$P_x\omega_2(x,y) = P_y\omega_2(x,y) = 0,$$

which also implies that the following chain of equalities holds

$$P_x\omega_2(x,y) + P_y\omega_2(x,y) = P_x\omega_2(x,y) + P_y\omega_2(y,x) + iP_yE(x,y) = P_x\omega_2(x,y) + P_y\omega_2(y,x) = 0, \quad (3.25)$$

where the last identity descends from the causal propagator satisfying the equation of motion. Furthermore, if we recall that $W(x, y)$ is taken to be smooth, also $P_xH(x, y)$ and $P_yH(x, y)$ must lie in $C^\infty(\mathcal{O} \times \mathcal{O})$. In order to exploit this remark we also need to <u>assume</u> that the function $V(x, y)$ in front of the logarithm can be expanded as a power series $V(x,y) = \sum_{n=0}^{\infty} v_n(x,y) \left(\frac{\sigma}{\lambda}\right)^n (x,y)$, where the dimensional constant is present since the left hand side is a scalar dimensionless function. That said, the strategy we shall follow is rather simple: we shall impose the equations of motion to the Hadamard parametrix and we shall obtain a series in powers of σ together with terms proportional to $\ln\sigma$. We shall thus impose that all the coefficients of the terms potentially diverging as $\sigma \to 0$ are actually identically 0.

Let us thus start from the first term in (3.24) and, for the sake of notational simplicity, we shall indicate the dependence neither on ϵ nor on (x, y) since they play no actual role in the forthcoming computation:

$$P_x\frac{U}{\sigma} = \frac{P_xU}{\sigma} - 2\frac{\sigma^\mu\nabla_\nu U}{\sigma^2} + \frac{2\sigma^\mu\sigma_\mu}{\sigma^3}U - \frac{\Box_x\sigma}{\sigma^2}U, \quad (3.26)$$

where $\sigma^\mu \doteq \nabla^\mu \sigma \equiv \partial^\mu \sigma$ since σ is a scalar function on $\mathcal{O} \times \mathcal{O}$ and where $\Box_x$ stands for $\Box_g$ applied to variable x. If we notice, that $\sigma^\mu \sigma_\mu = 2\sigma$ as computed in §2.1 of [46] and that the $\log \sigma$ term cannot contribute to a σ^{-2} term even if we expand V as power series, then we get the first equation, namely

$$2\sigma^\mu \nabla_\mu U + (\Box_x \sigma - 4)U = 0.$$

In order to solve it we need first to impose a suitable initial condition. In view of (3.25), it is sufficient to assign the behaviour of U when $x \to y$ and this is called the coinciding point limit, indicated as $[U] = \lim_{x \to y} U(x, y)$. Its value can be determined recalling that an Hadamard state behaves in the UV regime as the Minkowski vacuum and, thus, U must tend to value in Minkowski spacetime whenever $x \to y$. A long and tedious computation yield that $[U] = 1$, thus leading to

$$\begin{cases} 2\sigma^\mu \nabla_\mu U + (\Box_x \sigma - 4)U = 0 \\ [U] = 1 \end{cases}, \tag{3.27}$$

which completely determines $U(x, y)$.

Let us now focus on the function $V(x, y)$ and if we apply the operator P_x we get

$$P_x \left(V \frac{\ln \sigma}{\lambda^2} \right) = (P_x V) \ln \frac{\sigma}{\lambda^2} + 2 \frac{\nabla^\mu V \sigma_\mu}{\sigma} + V \frac{\Box_x \sigma}{\sigma} - 2 \frac{V}{\sigma}, \tag{3.28}$$

where we omitted the (x, y)-dependence and where we employed once more the identity $\sigma^\mu \sigma_\mu = 2\sigma$. Let us notice that the right hand side can be seen as the sum of two terms, one proportional to the logarithm and one proportional to σ^{-1}. Let us focus on the latter. If we plug in the expansion of V in power series, one can see that only the term with $n = 0$ survives, since all those with $n \geq 1$ are not divergent as $\sigma \to 0$. Hence the so-called v_0 contribution must compensate the one of the same order from (3.26) thus yielding the equation

$$P_x U + 2\nabla^\mu v_0 \sigma^\mu + (\Box_x \sigma - 2)v_0 = 0.$$

Since the U term has already been computed out of (3.27), we can read the last expression as a partial differential equation for v_0 with a source term, namely $P_x U$. In order to effectively determine v_0 we need to assign an initial condition and, for the same reasons as for U, we can just look at the behaviour as $x \to y$. If we take the coinciding point limit of the above expression, we get what we are seeking for:

$$[P_x U] + 2[v_0] = 0,$$

where we exploited that $[\sigma_\mu] = 0$ and $[\Box_x \sigma] = 4$ as computed in §2.3 of [46]. To summarize we get

$$\begin{cases} P_x U + 2\nabla^\mu (v_0) \sigma^\mu + (\Box_x \sigma - 2)v_0 = 0 \\ [v_0] = -\frac{1}{2} [P_x U] \end{cases}. \tag{3.29}$$

We are thus left with the contribution from the logarithmic term in (3.28), which yields $P_x V(x, y) = 0$. This is certainly true but, in order to solve it, we employ the same strategy and we actually expand V as a power series. After a not so lengthy but tedious computation one gets

$$P_x v_n + 2(n+1)\nabla_\mu v_{n+1} \sigma^\mu + [(n+1)\Box_x \sigma + 2n(n+1)]v_{n+1} = 0,$$

which can be read as a partial differential equation for v_{n+1} with a source term, namely $P_x v_n$. As before the initial condition can be fixed taking the coinciding point limit of this equality:

$$[P_x v_n] + 2(n+1)(n+2)[v_{n+1}] = 0.$$

Notice that the lowest term $n = 0$ is actually used to determine v_1 since v_0 is computed out of (3.29) and thus also the initial condition $[P_x(v_0)] + 4[v_1] = 0$ is meaningful. To summarize we get

$$\begin{cases} P_x v_n + 2(n+1)\nabla_\mu v_{n+1} \sigma^\mu + [(n+1)\Box_x \sigma + 2n(n+1)]v_{n+1} = 0 = 0 \\ [v_{n+1}] = -\frac{1}{(n+1)(n+2)} [P_x v_n] \end{cases}. \qquad \forall n \geq 1 \tag{3.30}$$

Consequence 3.3.0. The smooth functions $U(x,y)$ and $V(x,y)$ in (3.24) can be fully determined out of the **Hadamard recursion relations** given by (3.27), (3.29) and (3.30). The function V is constructed out of a power-series whose convergence is only guaranteed in an asymptotic sense. Most importantly all the involved partial differential equations depend only on quantities constructed out of the spacetime metric and on the constants appearing in the equation of motion, such as the squared mass for example. Thus both $U(x,y)$ and $V(x,y)$ can only depend on geometric quantities, contrary[7] to $W(x,y)$. The end game is thus the following

- the *singular structure* of the two-point function of an Hadamard state is fully determined out of the spacetime geometry,
- the *freedom in the choice* of an Hadamard state stands only in the assignment of the smooth part of (3.23), namely $W(x,y)$.

We conclude this section and the notes proposing an exercise which leads to compute an important quantity for concrete applications and the long path to the solution encompasses most if not all the tools we discussed:

Exercise: Let (M,g) be a globally hyperbolic spacetime and let $\phi : M \to \mathbb{R}$ be a scalar field satisfying the equation $P\phi = 0$ where $P = \Box_g - m^2 - \xi R$, where $\Box_g$ is the d'Alembert operator (1.4) and R is the scalar curvature. Show that, if one considers a quasifree Hadamard state $\omega : \mathcal{W}(M) \to \mathbb{C}$ for the associated Weyl algebra, then the Hadamard recursion relations of the associated two-point function are such that

$$[v_1] = \frac{1}{360}\left(C_{\mu\nu\rho\delta}C^{\mu\nu\rho\delta} + R_{\mu\nu}R^{\mu\nu} - \frac{R^2}{3} + \Box_g R\right) + \frac{1}{4}\left(\xi - \frac{1}{6}\right)^2 R^2 + \frac{m^4}{4} + \frac{1}{2}\left(\frac{1}{6} - \xi\right)\left(m^2 R + \frac{\Box_g R}{6}\right),$$

where $C_{\mu\nu\rho\delta}$ is the Weyl tensor, namely the traceless part of the Riemann tensor, which in four dimensions reads

$$C_{\mu\nu\rho\delta} \doteq R_{\mu\nu\rho\delta} - \frac{1}{2}(g_{\mu[\rho}R_{\delta]\nu} - g_{\nu[\rho}R_{\delta]\mu}) + \frac{R}{3}g_{\mu[\rho}g_{\delta]\nu}.$$

[7] A convincing and instructive exercise consists in repeating the procedure leading to the Hadamard recursion relations including also $W(x,y)$ and expanding it in a power series in terms of the ratio $\frac{\sigma}{\lambda}$. The final formulas show clearly that there is a freedom in the choice of $W(x,y)$ which thus cannot only be fixed out of the underlying geometry.

What's next?

As closure of these notes, we want to stress that we tried to put every reader in the condition to read with less efforts the modern literature on quantum field theory over curved backgrounds. This means that there are a lot of different and very active topics to choose from and this is certainly not the right place to describe all of them. We shall yet provide a brief description of a few of them, namely only those where the author has worked or is working on (in this way there is a fair chance that what it is written is meaningful!), so that a potentially interested reader has at least a bibliographic starting point. The order has no actual reference to the supposed relevance of the topic and its honestly speaking quite random (except may be for the first item):

- *The principle of general local covariance*: First formulated in the seminal paper [10], it leads to the realization of a quantum field theory as a covariant functor between the category of globally hyperbolic (four-dimensional) Lorentzian manifolds with isometric embeddings as morphisms and the category of C^*-algebras with invertible endomorphisms as morphisms. It is also remarkable the new interpretation of local fields as natural transformations from compactly supported smooth functions to suitable operators. This is a very important step forward in the understanding of the general structure and of the common features of a quantum field theory constructed over different backgrounds. It is substantially the natural modern language with which one should rephrase quantum field theory in presence of a non trivial spacetime and it leads to a great conceptual clarification when discussing advanced topic such as renormalization. In between the many applications, we would like to recall the extension to cope with spacetimes which are conformally related [45] and the recent development in [20] where a procedure to locally relate- free field theories on every strongly causal spacetime is set up thanks to the universal properties of the local causal structures of a differentiable Lorentzian manifold.

- *Renormalization*: The identification of the microlocal spectrum condition and, thus, of Hadamard states as the natural well-behaved physical states in a curved background has prompted a long series of research papers dealing first with the notion Wick polynomials for a field theory on a non trivial background. As we have seen for the scalar case in Minkowski spacetime, it is natural to define on a globally hyperbolic spacetime (M, g)

$$:\phi^2(x) :\doteq \lim_{x\to y}\left[\phi(x)\phi(y) - H(x,y)\right],$$

  where $H(x, y)$ is the Hadamard parametrix which depends only on geometric quantities as discussed in the previous section. Starting form this remark, one can carry on the analysis, first discussing the extended algebra of observables, encompassing the Wick polynomials, later introducing the notion of time ordering and ultimately developing a renormalization procedure which is applicable in the curved setting. This is an highly fascinating and physically relevant topic and many papers have been written on it [9, 33, 34, 35].

- *Dirac fields*: As it is crystal clear from these notes, the scalar field is a sort of prince of algebraic quantum field theory, but it is absolutely wrong to think that these methods do not apply also to other fields. Already in the eighties it has been shown that the algebraic scheme can be applied successfully to Dirac fields in [22] though this field played a sort of ancillary role and it was analysed only in few papers, for example [49]. Quite recently we have witnessed a small revival of the interest in spinors also due to their potential relevance in cosmological scenarios; particularly we stress the fact that the analysis of Dirac has been recast in the language of general covariance, the relative Cauchy evolution has been proven [51], the notion of Wick polynomials has been introduced and the conformal anomaly rigorously computed via the Hadamard regularization [15].

- *AQFT and Cosmology*: One of the most important classes of solutions of Einstein's equations is the one which arises imposing homogeneity and isotropy of the underlying background. The end point is the class of Friedmann-Robertson-Walker metrics, which is fully determined up to a positive smooth function depending only on time. These backgrounds are thought to be of high relevance in the description of the Universe as we observe it and therefore the development of a rigorous and full-fledged quantum field theory thereon is somehow desirable, if not even compulsory. In the past three years, there have been several attempts in this direction and very interesting results are available both at the level of field theory [13], at the level of construction of Hadamard states [17, 18, 43], and at a level of semiclassical Einstein's equations [14].

and, of course, around there is much much more ....

Bibliography

[1] C. Bär, N. Ginoux and F. Pfaffle, *"Wave equations on Lorentzian manifolds and quantization"*, Zürich, Switzerland: Eur. Math. Soc. (2007) 194 p.

[2] A. N. Bernal and M. Sanchez, *"Smooth globally hyperbolic splittings and temporal functions,"* arXiv:gr-qc/0404084.

[3] A. N. Bernal and M. Sanchez, *"Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes,"* Commun. Math. Phys. **257** (2005) 43 [arXiv:gr-qc/0401112].

[4] H. J. Borchers *"On structure of the field algebra"*, Nuovo Cimento **24** (1962) 214.

[5] W. M. Boothy *"An introduction to differentiable Manifolds and Riemannian Geometry"*, 2 ed. (1986) Academic Press Inc.

[6] O. Bratteli and D. W. Robinson, *"Operator Algebras And Quantum Statistical Mechanics. 1. C* And W* Algebras, Symmetry Groups, Decomposition Of States," New York, Usa: Springer-verl.(1979) 500 P.(Texts and Monographs In Physics).*

[7] O. Bratteli and D. W. Robinson, *"Operator algebras and quantum statistical mechanics. Vol. 2: Equilibrium states. Models in quantum statistical mechanics," Berlin, Germany: Springer (1996) 517 p.*

[8] R. Brunetti, K. Fredenhagen and M. Kohler, *"The microlocal spectrum condition and Wick polynomials of free fields on curved spacetimes,"* Commun. Math. Phys. **180** (1996) 633 [arXiv:gr-qc/9510056].

[9] R. Brunetti and K. Fredenhagen, *"Microlocal analysis and interacting quantum field theories: Renormalization on physical backgrounds,"* Commun. Math. Phys. **208**, 623 (2000) [arXiv:math-ph/9903028].

[10] R. Brunetti, K. Fredenhagen and R. Verch, *"The generally covariant locality principle: A new paradigm for local quantum physics,"* Commun. Math. Phys. **237** (2003) 31 [arXiv:math-ph/0112041].

[11] S. M. Carroll, *"Spacetime and geometry: An introduction to general relativity,"* Addison-Wesley (2004) 513 p.

[12] R. Courant and D. Hilbert *"Methods of mathematical physics. Vol. II. Partial differential equations"* (1962) Wiley-Interscience.

[13] A. Degner and R. Verch, *"Cosmological particle creation in states of low energy,"* arXiv:0904.1273 [gr-qc].

[14] C. Dappiaggi, K. Fredenhagen and N. Pinamonti, *"Stable cosmological models driven by a free quantum scalar field,"* Phys. Rev. D **77**, 104015 (2008) [arXiv:0801.2850 [gr-qc]].

[15] C. Dappiaggi, T. P. Hack and N. Pinamonti, *"The extended algebra of observables for Dirac fields and the trace anomaly of their stress-energy tensor,"* Rev. Math. Phys. **21**, 1241 (2009) [arXiv:0904.0612 [math-ph]].

[16] C. Dappiaggi, V. Moretti and N. Pinamonti, *"Rigorous steps towards holography in asymptotically flat spacetimes,"* Rev. Math. Phys. **18** (2006) 349 [arXiv:gr-qc/0506069].

[17] C. Dappiaggi, V. Moretti and N. Pinamonti, *"Cosmological horizons and reconstruction of quantum field theories,"* Commun. Math. Phys. **285** (2009) 1129 [arXiv:0712.1770 [gr-qc]].

[18] C. Dappiaggi, V. Moretti and N. Pinamonti, *"Distinguished quantum states in a class of cosmological spacetimes and their Hadamard property,"* J. Math. Phys. **50**, (2009) 062304 arXiv:0812.4033 [gr-qc].

[19] C. Dappiaggi, V. Moretti and N. Pinamonti, *"Rigorous construction and Hadamard property of the Unruh state in Schwarzschild spacetime"* arXiv:0907.1034 [gr-qc].

[20] C. Dappiaggi, N. Pinamonti and M. Porrmann, *"Local causal structures, Hadamard states and the principle of local covariance in quantum field theory,"* arXiv:1001.0858 [hep-th].

[21] J. Dimock *"Algebras of Local Observables on a Manifold"* Comm. Math. Phys. **77** (1980) 219.

[22] J. Dimock *" Dirac quantum fields on a manifold"* Trans. Amer. Math. Soc. **269** (1982), 133.

[23] J.J. Duistermaat and L. Hörmander, *"Fourier integral operators II"* Acta Math. **128** (1972), 183.

[24] C. J. Fewster: *"Lectures on quantum field theory in curved spacetime"* available at `http://www.mis.mpg.de/preprints/ln/lecturenote-3908.pdf` - (2008) Leipzig.

[25] F. G. Friedlander *"The wave equation on a curved space-time* (1975) Cambridge University Press.

[26] S. A. Fulling, *"Aspects of quantum field theory in curved space-time,"* London Math. Soc. Student Texts **17** (1989) 1.

[27] S. A. Fulling, F. J. Narcowich and R. M. Wald, *"Singularity Structure Of The Two Point Function In Quantum Field Theory In Curved Space-Time. II,"* Annals Phys. **136** (1981) 243.

[28] R. P. Geroch, *"Spinor Structure of Space-Times in General Relativity. I"* J. Math. Phys. **9** (1968) 1739.

[29] R. P. Geroch, *"Spinor Structure Of Space-Times In General Relativity. II,"* J. Math. Phys. **11**, (1970) 343.

[30] R. P. Geroch, *"Domain of dependence"* J. Math. Phys. **11**, (1970) 437.

[31] R. Haag and D. Kastler, *"An Algebraic approach to quantum field theory,"* J. Math. Phys. **5** (1964) 848.

[32] S. W. Hawking and G. F. R. Ellis, *"The Large scale structure of space-time,"* Cambridge University Press, Cambridge, 1973.

[33] S. Hollands and R. M. Wald, *"Local Wick polynomials and time ordered products of quantum fields in curved spacetime,"* Commun. Math. Phys. **223**, 289 (2001) [arXiv:gr-qc/0103074].

[34] S. Hollands and R. M. Wald, *"Existence of local covariant time ordered products of quantum fields in curved spacetime,"* Commun. Math. Phys. **231**, 309 (2002) [arXiv:gr-qc/0111108].

[35] S. Hollands and R. M. Wald, *"On the renormalization group in curved spacetime,"* Commun. Math. Phys. **237**, 123 (2003) [arXiv:gr-qc/0209029].

[36] L. Hörmander, *"The Analysis of Linear Partial Differential Operators I"* Spinger (2000).

[37] L. Hörmander, *"The Analysis of Linear Partial Differential Operators III"* Spinger (1994).

[38] B. S. Kay and R. M. Wald, *"Theorems on the Uniqueness and Thermal Properties of Stationary, Nonsingular, Quasifree States on Space-Times with a Bifurcate Killing Horizon,"* Phys. Rept. **207**, 49 (1991).

[39] S. Kobayashi, K. Nomizu, *"Foundations of differential geometry. Vol. I."* (1963) Wiley-Interscience Publication.

[40] S. Kobayashi, K. Nomizu, *"Foundations of differential geometry. Vol. II."* (1969) Wiley-Interscience Publication.

[41] V. Moretti, *"Uniqueness theorem for BMS-invariant states of scalar QFT on the null boundary of asymptotically flat spacetimes and bulk-boundary observable algebra correspondence,"* Commun. Math. Phys. **268**, 727 (2006) [arXiv:gr-qc/0512049].

[42] V. Moretti, *"Quantum ground states holographically induced by asymptotic flatness: Invariance under spacetime symmetries, energy positivity and Hadamard property,"* Commun. Math. Phys. **279**, 31 (2008) [arXiv:gr-qc/0610143].

[43] H. Olbermann, *"States of Low Energy on Robertson-Walker Spacetimes,"* Class. Quant. Grav. **24**, (2007) 5011.

[44] B. O'Neill, *"Semi-Riemannian Geometry"* (1983) Academic Press.

[45] N. Pinamonti, *"Conformal generally covariant quantum field theory: The scalar field and its Wick products,"* Commun. Math. Phys. **288** (2009) 1117 [arXiv:0806.0803 [math-ph]].

[46] E. Poisson, *"The motion of point particles in curved spacetime,"* Living Rev. Rel. **7** (2004) 6 [arXiv:gr-qc/0306052].

[47] M. J. Radzikowski, *"A Local to global singularity theorem for quantum field theory on curved space-time,"* Commun. Math. Phys. **180** (1996) 1.

[48] M. J. Radzikowski, *"Micro-Local Approach To The Hadamard Condition In Quantum Field Theory On Curved Space-Time,"* Commun. Math. Phys. **179** (1996) 529.

[49] H. Sahlmann and R. Verch, *"Microlocal spectrum condition and Hadamard form for vector valued quantum fields in curved space-time,"* Rev. Math. Phys. **13** (2001) 1203 [arXiv:math-ph/0008029].

[50] H. Sahlmann and R. Verch, *"Passivity and microlocal spectrum condition,"* Commun. Math. Phys. **214** (2000), 705.

[51] K. Sanders, *"Equivalence of the (generalised) Hadamard and microlocal spectrum condition for (generalised) free fields in curved spacetime,"* arXiv:0903.1021 [math-ph], to appear on Comm. Math. Phys.

[52] K. Sanders, *"The locally covariant Dirac field"* arXiv:0911.1304 [math-ph].

[53] R. Schmid, *"Infinite Dimensional Hamiltonian Systems"*, (1987) Bibliopolis.

[54] M. E. Taylor, *"Partial differential equations. I. Basic theory"* in Applied Mathematical Sciences **115** (1996) Springer Verlag.

[55] A. Uhlmann *"Über die Definition der Quantenfelder nach Wightmann und Haag"*, Wiss. Zeitschr. Karl-Marx Univ. Leipzig **11** (1962), 213.

[56] R. Verch, *"Wavefront sets in algebraic quantum field theory,"* Commun. Math. Phys. **205** (1999) 337 [arXiv:math-ph/9807022].

[57] R. M. Wald, *"General Relativity"* (1984) Chicago University Press.

[58] R. M. Wald, *"Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics"* (1994) Chicago Lectures in Physics.