Max-Planck-Institut für Gravitationsphysik
Dr. Frederic P. Schuller

Institut für Physik und Astronomie
Prof. Dr. Martin Wilkens

Tensorial spacetime geometries carrying predictive, interpretable and quantizable matter dynamics

Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Mathematische Gravitationsphysik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Sergio Rivera Hernández

Potsdam, den 15. Februar 2012

Max Planck Institute
for Gravitational Physics
Albert Einstein Institute

Universität Potsdam

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2012/6186/
URN urn:nbn:de:kobv:517-opus-61869
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-61869

A mis padres,
con mucho amor y admiración.

Contents

Acknowledgements

Dear Frederic, this work would not have been possible without your tremendous support in science and in all other aspects, and I have not enough space here to thank you for everything. Thank you for having supervised my doctoral research, I have learnt a lot of physics (but also about other things) from you. In particular, I thank you for conveying the aim to really understand and improve things. I really appreciate that you were always open to discuss about any subject and ideas. I thank you for the opportunity to come to this beautiful country, for your trust, for your advices, for your patience, for your help, and, of course, for the coffees.

Next, I would like to thank my PhD-colleagues Dennis Rätzel and Christof Witte. I enjoyed a lot the time we spent working together. Dear Dennis, it was a great time working in our first paper together with Frederic, and also in the quantization of matter fields. Dear Christof, I thank you for the insightful discussions, and it was a pleassure to have worked with you at the "Deutsche Schülerakademie".

I would also like to thank Prof. Dr. Martin Wilkens for acting as my supervisor at the university of Potsdam and Prof. Daniel Sudarsky for his understanding about finishing my PhD in Germany.

I thank my wife Vicky and my son Sergito for their love and support over the last years. It is wonderful to share my life with both of you. I also thank my parents, Fernando and Alicia, for their love and support in all circumstances, and my brother Fernando for all experiences we have had together.

I would also like to thank Fritz Henke, Frank Michalet, Natividad Batista, Wolfgang and Ingrid Christoph, and my friends Odir Rodriguez and Ivette Cruz for helping me (or my family) in several ways.

Finally, I would like to thank the Albert Einstein Institute for its hospitality and support where this doctoral work has been done, and the finacial support of the German Academic Exchange Service (DAAD).

CHAPTER 1

Introduction to spacetime geometries

In this doctoral work, we will determine the requirements that a pair (M, G) of a finite-dimensional smooth manifold M and an a priori arbitrary tensor field G must satisfy in order to provide a classical spacetime geometry. This introductory chapter will be devoted to motivate our investigation, to make precise which problem we will solve, and to describe the logical path we will follow. In the first section, we present a review on general relativity based on an idealized historical logic for its construction, which indeed mirrors how we will determine the conditions on (M, G). In the second section, we shortly review the main observations interpreted in the framework of general relativity and the standard model of particle physics. We will argue that these suggest new tensorial geometries to replace Lorentzian metrics as an appropriate mathematical structure modelling the physical spacetime structure. Finally, in the third section, we state key conditions that any pair (M, G) must satisfy in order to provide a classical spacetime geometry. The enterprise undertaken in this thesis is to make these conditions mathematically precise, derive rigorous conclusions and apply them to physics.

1.1. The conceptual path to general relativity

Historically, Einstein did not directly postulate that spacetime is a pair (M, g) of a finite-dimensional smooth manifold M equipped with a Lorentzian metric g and then made physical constructions on it, as sometimes it is the standard textbook presentation of general relativity. It was rather the study of a matter field, namely the electromagnetic field, that ultimately led Einstein to conclude that the geometry of spacetime could be encoded in a Lorentzian metric [1]. Here it will be precisely this logical path followed by Einstein, namely starting from a study of matter fields, that allows for the identification of the pairs (M, G) that can provide a classical spacetime structure. So let us review, using modern mathematical language, how Einstein constructed general relativity by studying Maxwell electrodynamics.

We begin by considering a one-form field A, to be identified with the electromagnetic potential, on a finite-dimensional smooth manifold M. If the manifold M on which the field A lives is equipped with a metric g, a priori not necessarily of Lorentzian signature, we can stipulate the well-known Maxwell action

$$S[A, g] = -\frac{1}{4} \int d^4x \sqrt{|\det(g)|}\, g^{ab} g^{cd} F_{ac} F_{bd} \tag{1.1}$$

to provide dynamics for the field A, where $F = dA$ is defined as the field strength, and the components of g^{ab} and F_{ab} are chosen with respect to some coordinate system. Variation of this action with respect to the electromagnetic degrees of freedom A yields a consistent set of partial differential equations, a priori for any signature of the metric. But if we want to do physics, we want to predict "future" values of the field A, i.e., we want the field equations for the field A to be predictive. More precisely, given initial values of the field A and its first derivatives on an "adequate" hypersurface Σ_0 of M, which is all we have access to measure, we want the dynamics arising from the action above to be able to reconstruct in a unique way the values of the field A at a "later" hypersurface Σ_t.

It is a well-known result from the theory of partial differential equations [2, 3] that predictivity in this sense can only be possible if the field equations arising from the considered action are of hyperbolic type. This already amounts to the requirement that the inverse metric g^{-1} in the action, and therefore the metric g itself, be of Lorentzian signature (and for definiteness, we choose the mainly minus signature $(+-\cdots-)$ in the following). Thus the *predictivity* requirement on the dynamics of the matter field A already restricts a metric manifold (M, g) to be Lorentzian. The adequate hypersurfaces Σ on which initial data can be provided are then recognized as those whose co-normals $q \in T_x^*M$ at each point x of the hypersurface Σ (which are determined up to a non-zero scale) are contained in one of the convex cones C or $-C$, which are defined as the set of covectors q that are positive with respect to g^{-1}, e.g. the cones defined by

$$g_x^{-1}(q, q) > 0\,, \tag{1.2}$$

see figure 1.1. The physical interpretation of the covectors k on the boundary of the cones C and $-C$, i.e., those satisfying $g^{-1}(k, k) = 0$, is obtained by studying the geometric optical limit of the field equations arising from the considered action. This study proceeds by considering wave-like approximate solutions of the form $e^{iS(x)/\lambda}$ for the field equations obeyed by A, where S is a scalar real function on M, and studying these for arbitrary small wavelength λ. These approximate solutions then correspond to high-frequency propagating waves whenever the wavelength is negligible in comparison to other relevant length scales. The study of quantities in this limit is *geometric optics*, where we therefore can describe the electromagnetic field by light rays. Crucially, one finds that the function S must satisfy the eikonal equation

$$g^{-1}(\partial S, \partial S) = 0\,, \tag{1.3}$$

such that the g^{-1}-null covectors correspond to co-gradients of the wave front hypersurfaces defined by $S(x) = $ constant. Equation (1.3) is therefore identified as the massless dispersion relation and the covectors satisfying this equation are called massless covectors. Notice that this picture, beginning with the study of a matter field first, has been so far completely developed in cotangent space, not tangent space. But we now want to find the equations for the worldline of a free massless particle, which will yield, as a corollary, the tangent space analogue to equation (1.3) that the tangent vector to a massless particle worldline must satisfy. Indeed, enforcing the massless dispersion relation by considering a pure constraint Hamiltonian $H(x, k) = g^{-1}(x)(k, k)$, the particle action takes the Helmholtz form

$$S[\mu, x, k] = \int d\tau\ \left[\dot{x}^a k_a - \mu\, g^{-1}(x)(k, k)\right]\ , \tag{1.4}$$

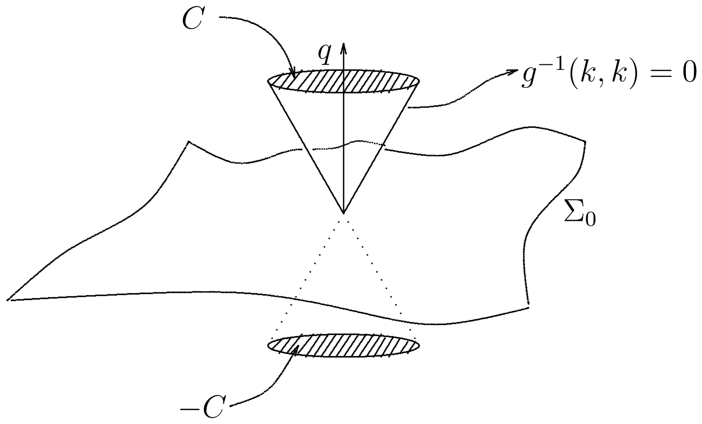

FIGURE 1.1. Suitable initial value hypersurfaces Σ_0 for Maxwell electrodynamics are those whose co-normals q at each point of the hypersurface are contained in one of the convex cones C or $-C$ (dashed parts in the figure) defined in each cotangent space.

where μ is a function of τ. Our aim is to eliminate the massless momentum k from the above action and thus obtain an action whose Lagrangian only depends on the configuration variables x and the multiplier μ. In the metric case considered in this introduction, this is easy: variation of the action with respect to k_a yields

$$\dot{x}^a = 2\mu\, g^{ab} k_b\,,$$

which expression we now want to invert in order to obtain k as function of $\dot{x}$. This is only straightforward for the obtained metric massless dispersion relation $g^{-1}(k,k)=0$, since here one can use the invertibility properties of the metric in order to obtain

$$k_a = \frac{1}{2\mu}\, g_{ab}\dot{x}^b\,. \tag{1.5}$$

However, as we will see, for the more general massless dispersion relations we are faced to consider later in this thesis, the map between k and $\dot{x}$ is not linear anymore. There, execution of this step requires a deep understanding of the mathematics behind the association between massless momenta with the corresponding worldline tangent vectors: real algebraic geometry. But for the simple case we now consider, using (1.5) one finally obtains the action

$$S[\mu, x] = \int d\tau\, \frac{1}{4\mu}\, g_{ab}\dot{x}^a\dot{x}^b$$

equivalent to the Helmholtz action (1.7), where variation of this action with respect to μ now shows that the worldline tangent vectors $\dot{x}$ associated with massless momenta satisfy the condition

$$g(\dot{x},\dot{x}) = 0\,.$$

In other words, tangent vectors $\dot{x}$ to worldlines of massless particles associated with massless momenta must lie in the vanishing set, or "light cone", defined by the metric g in each tangent space T_xM. One then prescribes a time-orientation on the manifold M by choosing a vector field T lying at each point of the manifold in one of the convex cones defined now in tangent space as the set of all vectors v such that $g(v,v) > 0$. The cone of future observers is then defined in tangent space as the convex cone containing the time-orientation vector T, see figure

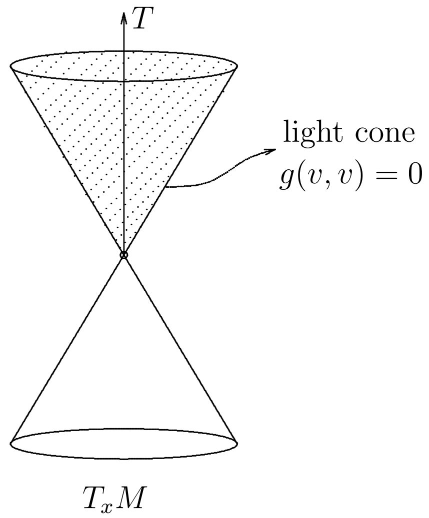

FIGURE 1.2. Once a time-orientation vector field T has been chosen, the cone of future observers is defined in tangent space as the set of all vectors contained in the connected component (dashed part in the figure) containing the time-orientation T.

1.2. Thus if Maxwell field dynamics is to be *interpretable*, the Lorentzian manifold (M, g) must be *time-orientable* in order to allow for an unambiguous notion of future-moving observers. This we will see to be true as well for the more general tensorial manifolds (M, G) considered later in this thesis.
Furthermore, with the notion of future observers at hand, positive energy covectors q are identified as those satisfying $q_0 = e_0(q) > 0$ for all future observer vectors e_0. This allows to distinguish C before $-C$ in figure 1.1 since only C contains positive energy covectors.

Turning now to the study of massive particles, one stipulates that the momenta q of positive energy massive particles are those contained in the convex cone C, which contains only positive energy covectors. The so defined set of positive energy massive momenta is then foliated by the positive number m identified with the mass of the particle in equation

$$g^{-1}(q, q) = m^2 \,, \qquad m > 0 \,, \tag{1.6}$$

which is therefore called the massive dispersion relation. Clearly, the massless momenta on the boundary of the cone C of positive energy massive momenta are also recognized to have positive energy. Thus all observers agree on the sign of the energy of massive as well as massless momenta. This *energy-distinguishability* property of Lorentzian spacetimes is crucial for Maxwell electrodynamics to be quantizable, since it allows to perform positive/negative frequency splits of fundamental solutions of Maxwell field equations in canonical quantization. This energy-distinguishability property is automatic for Lorentzian metrics, but must be imposed explicitly for any of the other tensorial spacetime geometries (M, G) considered later in this thesis.

With the knowledge of the massive dispersion relation we can obtain the trajectory of a free massive particle by considering the Helmholtz action

$$S[\lambda, x, q] = \int d\tau \, \left[\dot{x}^a q_a - \lambda(g^{-1}(x)(q, q) - m^2)\right] \,. \tag{1.7}$$

Variation of the action with respect to q_a yields

$$\dot{x}^a = 2\lambda\, g^{ab} q_b\,, \tag{1.8}$$

and using the invertibility properties of the metric we get (superficially seen: using similar mathematics as before) the relation

$$q_a = \frac{1}{2\lambda}\, g_{ab}\dot{x}^b\,. \tag{1.9}$$

However, as we will see, for massive dispersion relations associated with more general tensorial spacetimes, the map between q and $\dot{x}$ is not linear anymore, and solving (1.8) for the now massive momentum q requires a mathematical theory very different from the projective algebraic geometry we will use for massless momenta, namely convex analysis. But for the simple case we now consider, the simple inversion (1.9) suffices and one obtains the action

$$S[\lambda, x] = \int d\tau\, \left[\frac{1}{4\lambda}\, g_{ab}\dot{x}^a\dot{x}^b + \lambda m^2\right]$$

equivalent to the Helmholtz action (1.7), and further variation with respect to λ yields $\lambda = \sqrt{g(\dot{x},\dot{x})}/2m$, so that one finally obtains the action for a massive particle

$$S[x] = \int d\tau\, m\sqrt{g_{ab}\dot{x}^a\dot{x}^b}\,.$$

This action is reparametrization invariant, which allows to consider the natural parametrization

$$g(\dot{x},\dot{x}) = 1\,, \tag{1.10}$$

which is to correspond to the time shown by any clock which travels along the particle worldline, up to an affine reparametrization.

The final ingredient to interpret measurements is the introduction of observer frames at any point x of spacetime, which is defined as a basis $\{e_0, e_\alpha\}$ of the tangent space T_xM, where $\alpha = 1, \dots, \dim M - 1$, such that (i) e_0 is contained in the cone of observers (dashed part in figure 1.2), and (ii) the frame vectors $\{e_0, e_\alpha\}$ are ortho-normalized as

$$g_x(e_a, e_b) = \eta_{ab}\,, \tag{1.11}$$

where $\eta_{ab} = \mathrm{diag}(1, -1, \dots, -1)$ is the Minkowski normal form of the Lorentzian metric g at the point x. The space spanned by the $(\dim M - 1)$ vectors $\{e_\alpha\}$ is interpreted as the purely spatial directions seen by the observer. The introduction of observer frames allows now to interpret covariant quantities, such as the field strength F, in terms of quantities that can be measured by observers, such as the electric field $E_\alpha = F(e_0, e_\alpha)$.

This is essentially the entire kinematical apparatus of general relativity. Our presentation here deliberately focussed on its foundations in the *hyperbolicity, time-orientability and energy-distinguishability* conditions on the geometry (M, g) in order to make Maxwell electrodynamics *predictive, interpretable and quantizable*, and therefore to qualify as a spacetime geometry. This was instructive in so far as we will prove in this thesis that any pair (M, G) satisfying predictivity, time-orientability and energy-distinguishing conditions can carry predictive, interpretable and quantizable matter field dynamics and allows for an exactly analogous construction of the entire kinematical apparatus of general relativity.

We have thus found properties of the geometry (M, g) that ensure that matter propagates causally and that can be interpreted in terms of quantities measured by observers. The coefficients of matter field equations are completely defined by the values of the geometry at every point x of spacetime. Thus the next logical question is what determines the values of the geometry and, therefore, the coefficients of matter field equations? Remarkably, it has been shown that the developed kinematical setting also determines, under rather weak physical assumptions, also the dynamics of the geometry. Indeed, Hojman, Kuchar and Teitelboim [4], by studying the deformation algebra of initial data hypersurfaces in Lorentzian manifolds, reduced the problem of finding the dynamics for the Lorentzian metric to the problem of finding a minimal representation of the commutators of deformation operators of hypersurfaces in terms of geometric phase space variables defined on a hypersurface. They found a representation which uniquely led to Einstein's equations of general relativity,

$$R_{ab} - \frac{1}{2} g_{ab} R = 8\pi G\, T_{ab} + \Lambda g_{ab}\,. \tag{1.12}$$

where R_{ab} is the Ricci tensor, R the Ricci scalar, Λ a cosmological constant and T_{ab} the energy momentum tensor. This then completes the picture of classical spacetime in general relativity and we will summarize in the next section how well the theory fares when confronted with observations.

1.2. Observations

General relativity has been very successful in making predictions at scales of the solar system[1]. However, at cosmological scales, if one assumes that the source T_{ab} in equations (1.12) of the gravitational field (encoded in the Lorentzian metric g) can only be constituted by standard model matter, observations do not correspond to the predictions of general relativity. The main first two issues are only solved by stipulating vast amount of so-called dark matter and dark energy in addition to the standard model matter.

More precisely, in the simplest possible but realistic cosmology model, one assumes that the four dimensional spacetime manifold M is foliated by spacelike slices such that each three dimensional slice is spatially homogeneous and isotropic. The metric tensor g is therefore encoded in an FRW spacetime line element

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-\kappa r^2} + r^2 d\Omega^2\right] \tag{1.13}$$

with respect to coordinates $\{t, r, \theta, \phi\}$, where $a(t)$ is a dimensionless scale factor, κ encodes a constant curvature $\{-1, 0, 1\}$ of the spatial slices and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. On the other hand, the distribution of matter and energy in spacetime is modelled, at cosmological scales, as a perfect fluid with density and pressure parameters ρ and p, respectively; for instance, for non-interacting matter or dust, one takes $p_M = 0$. A universe in which most of the energy density is in the form of dust is called matter-dominated. Radiation is modelled by the equation of state $p_R = 1/3\rho_R$, and a universe in which most of the energy density is in the form of

[1] In fact, there are still some observations that must be explained within the solar system, see [5].

radiation is called radiation-dominated. Finally, the contribution of a cosmological constant Λ to the energy contents in the universe can be modelled by an equation of state $p_\Lambda = -\rho_\Lambda$. A universe in which most of the energy density is in the form of a cosmological constant is called energy-dominated. In particular, one finds that in our late universe, the contribution of radiation to the energy contents in the universe is negligible in comparison with the contribution of matter and a cosmological constant and, therefore, will not be considered in this short discussion, see [6, 7] for further details.

Thus, at late times, only dust and a cosmological constant as matter source for the gravitational field may be considered, so that from Einstein's equations one obtains the Friedmann equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}(\rho_M + \rho_\Lambda) - \frac{\kappa}{a^2} \qquad \text{and} \qquad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_M - 2\rho_\Lambda)\,. \tag{1.14}$$

It is then convinient to define $H = \dot{a}/a$ as the Hubble parameter whose present value $H_0 \simeq 72\,(km/s)/MpC$ is known as the Hubble constant, such that the first of the above equations can be written as

$$\Omega_\kappa = 1 - \Omega_M - \Omega_\Lambda\,, \tag{1.15}$$

where

$$\Omega_M = \frac{8\pi G}{3H^2}\rho_M\,, \quad \Omega_\Lambda = \frac{8\pi G}{3H^2}\rho_\Lambda\,, \quad \Omega_\kappa = \frac{\kappa}{a^2 H^2}\,.$$

One then has the following observational results about the late universe

- Experiments have shown that the expansion of the universe accelerates [8, 9]; this is meant in the sense that the scale factor a satisfies $\ddot{a} > 0$. From the second of the Friedmann equations in (1.14), one sees that this can only be possible if there is a sufficient contribution of ρ_Λ to the energy contents of the universe in comparisson to the matter contribution ρ_M.
- Detailed observations of the power spectrum of fluctuations in the cosmic microwave background [6] set an upper limit $\Omega_\kappa < |0.05|$, which means that the universe is spatially almost flat and that

  $$\Omega_M + \Omega_\Lambda = 1 \pm 0.05\,,$$

  so that the model $\kappa = 0$ seems distinguished.
- The matter density ρ_M itself is inferred, for instance, by investigating the gravitational effects of cluster matter. For example, one investigates the necessary matter contents in order to obtain the correct rotation curves of spiral galaxies [6, 10]. In any case, the result is

  $$\Omega_M \approx 0.3\,, \qquad \text{and thus} \qquad \Omega_\Lambda \approx 0.7\,.$$

  Hence, approximately, only 30% of the energy contents of the universe is due to matter and the other 70% is of cosmological constant (dark energy) origin.

Unfortunately for the idea of a universe governed by general relativity and filled with standard model matter,

- consistency with the cosmic microwave background spectrum and agreement with predictions of the abundance of light elements for big bang nucleosynthesis imply that the baryonic matter contents of the universe $(\Omega_M)_{\text{baryonic}}$ is a mere

$$(\Omega_M)_{\text{baryonic}} = 0.04 \pm 0.02\,,$$

  which represents, approximately, only the 15 % of matter contents in the universe because $\Omega_M \approx 0.3$, and so only 4 % of the matter contribution to the energy of the universe can be provided by standard model matter,
- a naive estimate of the vacuum energy density ρ_Λ, as it would originate from the vacuum of matter fields in the standard model of particle physics results in a number 120 orders of magnitude larger than the measured one.

Thus, if one insists on the validity of general relativity and the standard model of particle physics, one is forced to conclude that (i) approximately 85 % of the matter contents in the universe must be of non-baryonic (i.e, dark matter) origin, and (ii) that the energy contents in the universe ρ_Λ must be of other (i.e. dark energy) origin than the vacuum energy of matter fields in the standard model of particle physics. The existence of dark matter and dark energy is the overwhelming majority opinion on the resolution of the above described discrepancy between observations and gravity theory sourced by only standard model matter. Indeed, this view is a logical possibility. The lack of knowledge concerning the precise nature of dark matter and dark energy, however, justifies the somewhat more sober view that the amounts of 75 % of dark energy and 21% of dark matter are a pure parametrization of our ignorance of what precisely is going on. The only thing we know for sure is that there is something we do not understand about matter, gravity, or both.

But then we may face the daunting possibility to abandon general relativity and/or the standard model of particle physics in its present form. This view is now greatly supported by a recent experimental result, namely the observation of neutrino propagation at speeds faster than that of light: the OPERA collaboration [11] measured, at a reported 6σ confidence level, that muon neutrinos with mean energy of 17.5 GeV travel 730 km from CERN to the OPERA detector in the Gran Sasso laboratory arriving 60 nanoseconds less than expected. The conclusion is that these neutrinos exceed the velocity of light by 25 parts in a million. However, as seen in the previous section and shown in [12], physical particles cannot propagate faster than light on Lorentzian manifolds. This observation, if confirmed, shows that the very foundations of Einstein's theory, a Lorentzian metric as the spacetime geometry, is contradicted by experiment.

1.3. New tensorial geometries

General relativity and the standard model of particle physics are both built on the assumption that spacetime is a pair (M, g) of a finite-dimensional smooth manifold M and a Lorentzian metric g. However, predictions, as shortly reviewed in the previous section, are not effortless in agreement with observations. In particular, the reported faster-than-light propagation of neutrinos is irreconcilable with the idea that spacetime is a Lorentzian manifold. But this inevitably raises the hard question

What other tensor fields can serve as a spacetime geometry?

which we aim at answering rather comprehensively in this thesis. More precisely, we aim at identifying the pairs (M, G) of a finite-dimensional smooth manifold M and an a priori arbitrary smooth tensor field G that can provide a classical spacetime geometry, following the same logical path as for the construction of general relativity presented in section 1.1. We will thus select the pairs (M, G) that can provide a classical spacetime geometry by probing matter fields[2] on (M, G) (precisely as Lorentzian manifolds were selected from the set of metric manifolds by studying the electromagnetic field) and requiring their dynamics to be

I. *Predictive* (= *hyperbolicity of the geometry*). Predicting the "future" values of the matter fields propagating in spacetime is all what classical physics is about. If a fundamental theory of spacetime lacks this predictive power, it must be discarded as a viable theory. We must therefore require that in order to provide a classical spacetime geometry, the pair (M, G) must be such that the dynamics for the matter fields propagating in and probing the geometry of (M, G) must be predictive.

II. *Interpretable* (= *time-orientability of the geometry*). There is no future to predict if we cannot say what future is. Thus, in order to provide a classical spacetime geometry, we require that the pair (M, G) must allow for the choice of a time-orientation. We will see that this condition also allows to provide an unambiguous notion of observers and, therefore, to interpret spacetime quantities in terms of those measured in any observer's frame.

III. *Quantizable* (= *energy-distinguishability of the geometry*). In order to quantize a classical theory, it is crucial to unambiguously distinguish between the positive and the negative energy solutions of matter field equations. Thus, in order to provide a classical spacetime geometry that can carry quantizable fields, the pair (M, G) must satisfy the additional condition of allowing a notion of positive energy of particles on which all observers agree. This condition will play a crucial classical rôle as it was the case for metric geometry in section 1.1.

---

[2]Physically, this is all we can do; all conclusions about the structure of spacetime are obtained only by probing matter on it.

All kinematical constructions known in general relativity will be seen to have an analogue on spacetime backgrounds (M, G) satisfying the above three fundamental conditions. Hence, any pair (M, G) satisfying these conditions indeed deserves to be called a spacetime geometry.

The next three technical chapters are devoted to translate the stated predictivity, interpretability and quantizability conditions on the dynamics of matter fields probing the geometry of (M, G) into precise mathematical requirements on the tensor field G. In particular, in chapter 2, the predictivity condition will be mathematically translated into the requirement that the differential equations governing the dynamics of matter fields must be of hyperbolic type. This condition alone will identify a hyperbolic polynomial P in each cotangent space as the tensorial structure encoding the geometry seen by massless and massive point particles in each cotangent space. In chapter 3, the interpretability conditon on matter field dynamics, which is equivalent to the condition that the geometry of (M, G) be time-orientable, will be translated into the condition that the identified hyperbolic polynomial P in each cotangent space must correspond to a hyperbolic polynomial $P^\#$, the so-called dual polynomial of P, in each tangent space. The polynomial $P^\#$ will be shown to provide the geometry seen by massless particles in each tangent space. A choice of a time-orientation and an unambiguous notion of observers will then be provided in terms of the tangent bundle function $P^\#$. In chapter 4, the quantizability condition of matter field dynamics, which translates into the the requirement that the geometry (M, G) be energy-distinguishing, will be formulated as the condition that all observers, defined in terms of the cotangent bundle function $P^\#$, must agree on the sign of the energy of massless particles. This will also imply that they agree on the sign of the energy of massive particles and allow to develop a complete theory of massive particles. A generically non-polynomial tangent bundle function P^* will then be shown to encode the geometry seen by massive particles in each tangent space.

Thus, following the predictivity, interpretability and quantizability conditions on matter field dynamics, we will arrive at the main result of this work (as presented in [13]), namely that

A pair (M, G) of a finite-dimensional smooth manifold M and a tensor field G can present a classical spacetime geometry for given matter dynamics coupled to G only if the principal polynomial P of the matter dynamics is a hyperbolic, time-orientable and energy-distinguishing homogeneous polynomial in each cotangent fibre of M.

The explicit reference to matter needed here to qualify a geometry as a valid spacetime structure is to be seen not as a weakness, but rather as an insight; Lorentzian manifolds are a good spacetime structure for Maxwell electrodynamics but fail to be so if superluminal neutrinos are added.

Based on this result, the remaining technical chapters (chapters 5 to 9) will be devoted to investigate generic classical and quantum properties that can be inferred for any spacetime geometry (M, G). In particular, in chapter 5, we will introduce observer frames and construct transformations connecting observers. Using the introduced observer frames, we will then be able to present the temporal-spatial split of physical dispersion relations, and show that some proposals do not qualify as viable modified dispersion relations. We will finally show that refined spacetime geometries (M, G) not only generically admit superluminal propagation of massive matter but also that infraluminal motion is what particles always tend to, since they can radiate off Cherenkov radiation in vacuo, but only until they reach infraluminal speed. In chapter 6, we will show how the developed theory works for concrete proposals of spacetime geometries. First, we will show how purely based on the general theory developed in chapters 2 to 5, all known kinematical constructions on Lorentzian manifolds are recovered. Apart from being a consistency check, the study of Lorentzian manifolds in the framework of our general theory of spacetime geometries allows us to clearly recognize the different rôles that the metric plays in general relativity. We will then study the more complicated case of area metric manifolds, which illustrates how the predictivity, interpretability and quantizability conditions on matter field dynamics restrict a concrete non-metric geometry. We then proceed with the study of quantum matter, in particular, chapter 7 will be concerned with the quantization of the electromagnetic field on flat area metric backgrounds. The aim of this chapter is two-fold; to show how the conditions imposed on the geometry are important even at the quantum level in order to guarantee a positive definite Hamiltonian, and to provide an example of how quantization proceeds for a field satisfying a modified dispersion relation. In chapter 8, we study possible sources for the electromagnetic field on flat area metric backgrounds. First, we find a covariant propagator for Maxwell's equations on area metric spacetimes, and then study point charges coupled to the electromagnetic field. Thus we can perform a first quantization of free massive particles, and see that the particle-antiparticle interpretation holds on generic spacetime geometries. Finally, we study modifications of the Dirac equation and provide an explicit example of a modified Dirac equation for the case of a bimetric spacetime. Finally, the last technical chapter (chapter 9) will be devoted to review, how using the theory to be presented in this work, it was achieved [14, 15] to reduce the problem of finding the gravitational dynamics of any spacetime geometry (G, M) to a problem of representation theory, and finally to the problem of solving a homogeneous linear system of partial differential equations.

CHAPTER 2

Condition I: Hyperbolicity

In this first technical chapter, we study tensorial matter field equations on spacetime geometries defined in terms of a (a priori arbitrary) tensor field. The requirement that the dynamics of matter fields coupled to the geometry be predictive implies that a particular polynomial constructed in cotangent space from the coefficients of the field equations (and thus the underlying geometric tensor field) must be hyperbolic. Important first results for the cotangent space geometry of a spacetime manifold are then collected.

2.1. Initial value problem and hyperbolicity

As explained in the introduction, we wish to investigate which pairs (M, G) of a finite-dimensional smooth manifold M and a smooth tensor field G of arbitrary valence can provide a spacetime geometry. In this and the following two chapters we will find necessary restrictions on G for (M, G) to provide a consistent classical spacetime structure in the sense that the spacetime (M, G) can carry predictive, interpretable and quantizable matter field dynamics. Along the way, we will see that for any such spacetime (M, G), the full kinematical apparatus known from general relativity can be constructed.

This chapter, in particular, is concerned with the first condition, predictivity, which mathematically refers to the requirement of the well-posedness of the initial value problem for the dynamics of matter fields propagating on the geometric background (M, G). So we consider a field Φ (the "matter") taking values in some tensor representation space V. The matter field Φ probes the geometry of the background (M, G) by coupling to the tensor G (the "geometry") via the action

$$S[\Phi, G] = \int d^{\dim M}x\, \mathcal{L}(\Phi, \partial\Phi, \partial\partial\Phi, \dots, G, \partial G, \partial\partial G, \dots)\,,$$

where the Lagrangian density $\mathcal{L}$ may depend on the tensor field G, on the field Φ and on finitely many partial derivatives of both. Notice that the matter field dynamics arising from the variation of the action $S[\Phi, G]$ with respect to the field variables Φ^A (with $A = 1, \dots, \dim V$) may be restricted from start to those giving rise to linear field equations, since only these can serve as test matter probing the geometry[1]. Moreover, the linearity of the matter field equations also ensures that the superposition principle, which is crucial to perform a quantization of free matter fields, is satisfied. More precisely, the linear field equations obtained by variation of the action

[1]This is the case because considering non-linear field equations, in most cases one cannot disentangle the properties of the underlying geometry from the properties of the matter field.

can be assumed to take the general form of the s-th order linear partial differential equations,

$$\left[\sum_{n=0}^{s} Q[G]_{MN}^{i_1\cdots i_n}(x)\,\partial_{i_1}\cdots\partial_{i_n}\right]\Phi^N(x) = 0\,, \tag{2.1}$$

where small spacetime latin indices range from 0 to $\dim M - 1$ and the matrix coefficients Q at all orders depend only on the tensor field G and its partial derivatives, but not on the value of the fields Φ^N; note that this is really only true for genuinely linear field equations, and does not even hold for the linearization of fundamentally non-linear dynamics. The fact that the above field equation is derived by variation with respect to the tensor field Φ from a scalar action implies that the entire equation transforms as a tensor. The mathematically weakest form of the predictivity requirement is then that the field equations have a well-posed initial value problem.

Definition. *We say that the Cauchy problem (or initial value problem) for the system of field equations (2.1) is well-posed if*

(1) for initial data on suitable initial value surfaces there exist a unique solution to the system of field equations, and
(2) the solutions depend continuously on the initial data.

The key mathematical structure needed to formulate a workable criterion for well-posedness of linear field equations is provided in the following

Definition. *The principal symbol* [16, 17] *of the differential field equations (2.1) is the cotangent bundle function*

$$P : T^*M \to \mathbb{R}\,, \qquad (x,q) \mapsto P(x,q)$$

defined by

$$P(x,q) = \rho \det_{M,N}\left[Q_{MN}^{i_1\cdots i_s}(x)\, q_{i_1}\cdots q_{i_s}\right]\,. \tag{2.2}$$

This expression is defined solely in terms of the leading order coefficients $Q_{MN}^{i_1\cdots i_s}(x)$ since of all coefficients Q in (2.1), only these are certain to transform like a tensor. Thus the determinant is a tensor density whose weight depends on the precise form of the field equations as well as on the geometry, and is to be countered by the weight ρ, which must be constructed from the geometry G such that the principal symbol indeed becomes a function, not a function density, on the cotangent bundle.

An important remark has to be made at this point. When gauge symmetries are present in the action, it typically happens that the principal symbol $P(x,q)$ is identically zero. In this case, one either has to first fix the gauge in the action and may only then compute the principal symbol of the corresponding gauge fixed field equations, or one needs to resort to some other method to eliminate gauge ambiguities, such as re-writing the field equations in terms of field strengths and constructing P only from the evolution equations [18].

Provided the function P has been appropriately constructed in some way, we can now state the main theorem of this section, which already provides the first condition on the geometry.

THEOREM 2.1.1. (theorem 2.1 and 3.1 of [16], and theorem 1.2.1 of [17]) *Assume that the Cauchy problem for the field equations (2.1) is well-posed in a region of spacetime. Then the principal symbol P defines a homogeneous hyperbolic polynomial P_x at every point x of the considered region. More precisely,*

$$P_x : T_x^*M \to \mathbb{R}\,, \qquad q \mapsto P_x(q) = P(x,q)$$

must be a homogeneous hyperbolic polynomial. Moreover, suitable initial value surfaces must have co-normals hyperbolic with respect to P_x.

The suitable initial value surfaces determined by P may justifiedly be called P–spacelike hypersurfaces.

Since the notion of a hyperbolic polynomial will play a crucial rôle throughout our investigation, we present the relevant definition and main properties, in a way congenial to our further developments, in the following section.

2.2. Hyperbolic polynomials

As we have just seen, in order to provide a predictive geometric background, the pair (M, G) must be such that the cotangent bundle function P (constructed in terms of the geometry G from the matter field equations) must provide a homogeneous hyperbolic polynomial P_x in each cotangent space T_x^*M.
Homogeneity simply means that $P_x(\lambda q) = \lambda^{\deg P} P_x(q)$ for all $q \in T_x^*M$ and $\lambda \in \mathbb{R}$, and it is convenient to introduce the vanishing set

$$V(P_x) = \{k \in T_x^*M \mid P_x(k) = 0\}\,, \tag{2.3}$$

of a homogeneous polynomial P_x, which has the structure of a cone[2] exactly because of the homogeneity of P_x. We further have the

DEFINITION. *A homogeneous polynomial P_x is called hyperbolic with respect to some covector h if $P_x(h) \neq 0$ and for every covector q, any λ solving*

$$P_x(q + \lambda h) = 0 \tag{2.4}$$

is real.

For any polynomial P that is hyperbolic with respect to one covector h one can always arrange for $P(h) > 0$ by choosing the appropriate sign for the density ρ in equation (2.2), and we agree to make precisely this choice in the following[3].

Example. (Hyperbolicity of a Lorentzian polynomial) Consider an inverse Lorentzian metric $g^{-1} : T_x^*M \times T_x^*M \to \mathbb{R}$ and basis covectors $\{\epsilon^a\}$ for T_x^*M such that $g^{ab} = g(\epsilon^a, \epsilon^b) = \mathrm{diag}(1, -1, \cdots, -1)$. Then the Lorentzian polynomial $P_x(q) = g_x^{-1}(q,q)$ is hyperbolic with respect to $h = (1, 0, \cdots, 0)$ represented in the same basis $\{\epsilon^a\}$. This can be seen by taking an arbitrary covector $q = (q^0, \cdots, q^{\dim M - 1})$ because then the discriminant of the second order equation

[2]A subset C of a real vector space is called a cone if $\lambda v \in C$ for any $v \in C$ with λ real and positive.

[3]This corresponds, in the familiar metric case, to a choice of mainly minus signature.

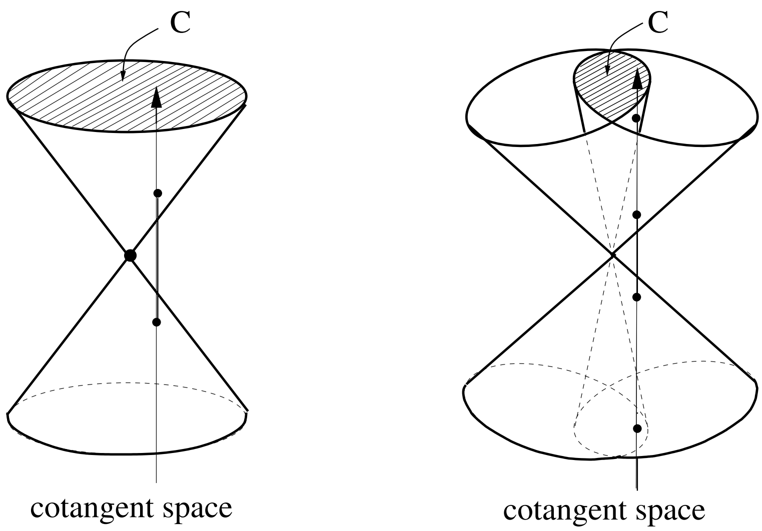

Figure 2.1. Hyperbolicity cones for two prototypical polynomials. On the left, the familiar second degree Lorentzian cone; on the right, a fourth degree cone defined, for simplicity, by a product of two Lorentzian metrics.

in λ, $g_x(q+\lambda h, q+\lambda h)=0$, is simply given by the positive quantity $(q^0)^2+\cdots+(q^{\dim M-1})^2$. This implies that the λ-roots are always real, as required from the definition of hyperbolicity.

The definition of a homogeneous hyperbolic polynomial is easy to understand in geometric terms. It simply means that there is at least one covector h such that every affine line in cotangent space in the direction of h intersects the cone $V(P_x)$ defined by P_x in precisely $\deg P$ points, see figure 2.1, counting algebraic rather than geometric multiplicities. Any such covector h identifying P_x as hyperbolic is itself called a hyperbolic covector with respect to P at the point x.

The various connected sets of hyperbolic covectors in the same cotangent space, as for instance the upper (shaded) cones in figure 2.1, are called the hyperbolicity cones of P at x. More precisely,

Definition. *Let h be a covector hyperbolic with respect to P_x. Then the hyperbolicity cone $C(P_x, h)$ containing h is constituted by all covectors q with the property that all λ satisfying*

$$P_x(q-\lambda h)=0$$

are positive, that is, $C(P_x,h)=\{q\in T_x^*M \,|\, P_x(q-\lambda h)=0 \text{ has only positive roots } \lambda\}$.

From this definition, it follows immediately that $C(P_x,h)$ is indeed a cone. The remarkable properties of hyperbolicity cones, which underlie all further constructions, have been mainly elucidated by Gårding [19] a long time ago. Recalling that a subset $C\subset T_x^*M$ is called convex if $(1-\lambda)u+\lambda v\in C$ for all $u,v\in C$ and $\lambda\in[0,1]$, Gårding proved the following theorem.

Theorem 2.2.1. (Gårding [19] and theorem 3.1 of [20]) *The hyperbolicity cone $C(P_x,h)$ is open and convex, the hyperbolic polynomial P_x is strictly non-zero on $C(P_x,h)$ and is equal to zero on the boundary of $C(P_x,h)$. Furthermore, $C(P_x,-h)=-C(P_x,h)$ and P_x is hyperbolic with respect to any $h'\in C(P_x,h)$.*

It is often useful to take a more global point of view and consider a smooth distribution C of hyperbolicity cones C_x over all spacetime points x, which one simply may think of as the cone of all smooth covector fields h for which $h_x \in C_x$. More precisely, let h be a covector field hyperbolic with respect to P, that is h defines a hyperbolic covector at every spacetime point. Then the hyperbolicity cone $C(P, h)$ containing h is constituted by all covector fields q with the property that all functions λ on M satisfying

$$P(x, q(x) - \lambda(x)h(x)) = 0 \tag{2.5}$$

are positive everywhere on M. The cone $C(P, h)$ induces a cone $C_x(P, h)$ in each cotangent space T_x^*M, consisting of the values $q(x)$ of all $q \in C(P, h)$ evaluated at x, which is called the hyperbolicity cone of P with respect to h at x. Clearly, $C_x(P, h)$ only depends on the value of h at x, and thus one may think of $C(P, h)$ simply as the said distribution of the $C_x(P, h)$ over all $x \in M$. The somewhat implicit definition of hyperbolicity cones, both the local one in (2.4) and the global one in (2.5), can be cast into the explicit form of $\deg P$ polynomial inequalities, as follows from the following

THEOREM 2.2.2. (Theorem 5.3 of [20]) *Let P be hyperbolic with respect to h and $P(h) > 0$. Then the hyperbolicity cone is described by the* $\deg P$ *inequalities*

$$\det H_i(v, h) > 0 \qquad \textit{for all } i = 1, \ldots, \deg P\,, \tag{2.6}$$

where the matrices $H_1, H_2, \ldots, H_{\deg P}$ are constructed as

$$H_i(v,h) = \begin{bmatrix} h_1 & h_3 & h_5 & \ldots & h_{2i-1} \\ h_0 & h_2 & h_4 & \ldots & h_{2i-2} \\ 0 & h_1 & h_3 & \ldots & h_{2i-3} \\ 0 & h_0 & h_2 & \ldots & h_{2i-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & h_i \end{bmatrix}_{i \times i} \qquad \textit{where } h_j \textit{ is set to } 0 \textit{ for } j > i \tag{2.7}$$

from the coefficients of the expansion

$$P(x, v + \lambda h) = h_0(x, v, h)\,\lambda^{\deg P} + h_1(x, v, h)\,\lambda^{\deg P - 1} + \cdots + h_{\deg P}(x, v, h)\,. \tag{2.8}$$

There is a rather elaborate theory of semi-algebraic sets [21, 22] defined by polynomial inequalities, of which the hyperbolicity cones are, according to the above theorem, a particular instance, and we will have the opportunity to use some results from this theory in the proof of lemmas in chapter 3.

The calculation of hyperbolicity cones is significantly simplified if the polynomial P is factorizable as

$$P_x = P_{1\,x}^{a_1}\, P_{2\,x}^{a_2} \cdots P_{f\,x}^{a_f}\,.$$

In this case, P is hyperbolic with respect to h if and only if each of its individual factors is hyperbolic with respect to h and the corresponding hyperbolicity cones have non-empty intersection with each other. For such reducible P_x, the determination of the hyperbolicity cone with respect to some hyperbolic covector h is reduced to the determination of the hyperbolicity cones of the individual factors, since

$$C_x(P_x, h_x) = C_x(P_{1\,x}, h_x) \cap \cdots \cap C_x(P_{f\,x}, h_x)\,. \tag{2.9}$$

Thus it is not a coincidence that the hyperbolicity cone indicated in the right half of figure 2.1 is the intersection of the hyperbolicity cones of the corresponding two Lorentzian cones. Another mathematical concept which will be of fundamental relevance in chapter 4 is provided in the following

Definition. *A homogeneous hyperbolic polynomial P_x is called complete if the lineality space*

$$L(P_x) = \{q \in T_x^*M \,|\, \textit{for all } v \in T_x^*M \textit{ and } \lambda \in \mathbb{R} : P_x(v + \lambda q) = P_x(v)\} \tag{2.10}$$

only contains the zero covector.

In other words, in order to be complete, P must depend on all covector components in any chosen basis. Geometrically, completeness can be read off from the closure of the hyperbolicity cones since it is equivalent [23] to

$$\text{closure}(C(P_x, h)) \;\cap\; \text{closure}(C(P_x, -h)) = \{0\}\,. \tag{2.11}$$

However, completeness does not need to be postulated here since it will follow in conjunction with the energy-distinguishing condition we have good reason to impose in chapter 4. We finish this section by stating the following theorem for complete hyperbolic polynomials, which will play a significant rôle in chapter 4.

Theorem 2.2.3. *Defining the tensor*

$$P_x(q_1, \ldots, q_{\deg P}) = \frac{1}{(\deg P)!} \prod_{J=1}^{\deg P} \left(\sum_{i=1}^{\dim T_x^*M} (q_J)_i \frac{\partial}{\partial q_i} \right) P_x(q)\,, \tag{2.12}$$

as the totally symmetric polarization of the polynomial P, it follows that if P_x is complete,

(1) the reverse triangle inequality

$$P_x^{1/\deg P}(q_1 + q_2) \geq P_x^{1/\deg P}(q_1) + P_x^{1/\deg P}(q_2) \tag{2.13}$$

is satisfied for all q_1 and q_2 in the same hyperbolicity cone C_x and

(2) a reverse Cauchy-Schwarz inequality

$$P_x(q_1, \ldots, q_{\deg P}) \geq P_x(q_1)^{1/\deg P} \cdots P_x(q_{\deg P})^{1/\deg P} \tag{2.14}$$

is satisfied for all $q_1, \ldots, q_{\deg P}$ in the same hyperbolicity cone C_x, where equality holds if and only if all arguments q_i are proportional to each other.

Thus we have learnt in this chapter that for a geometry (M, G) to provide a feasible background for predictive matter field equations, the principal polynomial constructed from the latter in terms of the former must be a homogeneous hyperbolic polynomial P_x in each cotangent space. This is clearly a property that cannot be weakened. This is a rather well-known result from the theory of partial differential equations, which has previously been recognized as a condition for matter field dynamics [24]. But in the course of the next two chapters, we will greatly extend this result in order to have a geometry that does not only allow for predictivity of matter field equations, but also for their interpretability from the point of view of observers.

CHAPTER 3

Condition II: Time-orientability

In this chapter, we first show that the principal polynomial P associated with predictive matter field equations determines the dispersion relation for massless point particles. Based on that insight, we develop the appropriate duality theory that associates velocity vectors with massless particle momenta. The key result of this chapter, from a practical point of view, is then the identification of the hyperbolicity of a dual polynomial $P^{\#}$ to the principal polynomial P as a requirement for the time-orientability of (M, G) and, therefore, for the interpretability condition to hold.

3.1. The short wave approximation and the massless dispersion relation

The central result of the previous chapter was that the principal symbol of the field equations governing the dynamics of matter fields must be a homogeneous hyperbolic polynomial P_x in each cotangent space T^*_xM in order to make the theory predictive. Here, we will see that the principal symbol plays a second important rôle; it provides the massless dispersion relation which arises in the geometric optical limit of equations (2.1) as the solvability condition

$$P(x, q) = 0\,. \tag{3.1}$$

This is seen by considering wave-like matter solutions of equation (2.1) taking the form of the formal series

$$\Phi^N(x, \lambda) = e^{i\frac{S(x)}{\lambda}} \sum_{j=0}^{\infty} \phi_j^N(x)\lambda^j \tag{3.2}$$

and then producing an approximate solution by considering the case of very small wavelength λ. In the above expansion, $\phi_j^N(x)$ is a tuple of functions for each j, and the quantity S/λ in the phase of the considered wave-like solution is required to be real-valued. The function $S(x)$ is known as the eikonal function or the wave front surface. Physically, considering very small λ can be thought of as the limit of very short wave lengths, or of waves propagating with extremely high frequencies. Then the ansatz (3.2) is known as the short wave approximation or the limit of geometric optics [3, 25, 26, 27].

Substituting the formal series (3.2) into the field equations (2.1), one finds

$$e^{i\frac{S(x)}{\lambda}}\lambda^{-s}\left[Q_{MN}(x)^{i_1\cdots i_s}\partial_{i_1}S(x)\cdots\partial_{i_s}S(x)\,\phi_0^N(x) + \sum_{j=1}^{\infty} v_{Mj}(x)\lambda^j\right] = 0\,, \tag{3.3}$$

where each of the $v_{Mj}(x)$ terms depends on some of the matrix coefficients Q of the differential equation (2.1), on the coefficients $\phi_j^N(x)$ of the expansion (3.2) and on the eikonal function S and its derivatives of lower than the highest order s. For $\Phi^N(x)$ to be a solution after any truncation of the series (3.3), the latter has to vanish order by order in λ. Clearly, the first term $e^{i\frac{S(x)}{\lambda}}Q_{MN}(x)^{i_1\cdots i_s}\partial_{i_1}S(x)\cdots\partial_{i_s}S(x)\,\phi_0^N(x)$, corresponding to the power λ^{-s}, vanishes with non-trivial ϕ_0^N only if the eikonal function S satisfies the differential equation

$$\frac{P(x,dS)}{\rho} = \det\Big(Q[G]^{i_1\cdots i_s}(x)\partial_{i_1}S(x)\cdots\partial_{i_s}S(x)\Big) = 0\,, \tag{3.4}$$

where $P(x,dS)$ is recognized to be the principal symbol of equation (2.1) evaluated at the point (x,dS) in cotangent space. Equation (3.4) is known as the eikonal equation and represents the solvability condition for the first term in (3.3). The relevance of this equation and therefore of the geometric optical limit, beyond the rôle it plays for us here, is that having this lower order approximate solution, one can generate higher order approximate solutions converging to the actual solution.

In case P is a reducible polynomial in each fibre, i.e., a product

$$P(x,q) = P_1(x,q)^{a_1}\cdots P_f(x,q)^{a_f}$$

of irreducible[1] factors $P_1,\ldots,P_f$ with positive integer exponents $a_1,\ldots,a_f$, we agree to take as the cotangent bundle function P the *reduced* polynomial

$$P(x,q) = P_1(x,q)\cdots P_f(x,q). \tag{3.5}$$

In other words, one must remove repeated factors in the original polynomial. Thus no information is lost but degeneracies are removed, see the comment further below. We will have more to say about the relation between properties of a reduced polynomial and those of its individual factors.

Definition. *The set N_x of massless momenta at a spacetime point x is defined as*

$$N_x = \{k\in T_x^*M\,|\,P(x,k)=0 \text{ with } P \text{ reduced}\}\,, \tag{3.6}$$

which due to the homogeneity of P_x constitutes a cone, i.e., every positive real multiple of a massless momentum is again a massless momentum. For technical reasons, we also define the subcone N_x^{smooth} of massless momenta

$$N_x^{smooth} = \{k\in N_x\,|\,DP(x,k)\neq 0\}\,, \tag{3.7}$$

where DP denotes the derivative of P with respect to the cotangent fibre.

So the cotangent bundle function P determines the (smooth) massless momentum cone. Clearly, if the principal polynomial were not reduced and one used it to define the set of massless momenta, the obtained set of massless momenta would coincide with the vanishing set $V(P_x)$ defined in equation (2.3) at a spacetime point x, but not with the set N_x^{smooth} of smooth massless momenta.

We finish this section by proving two key properties of hyperbolic polynomials, which we will use repeatedly throughout this work.

[1]A non-constant real polynomial is irreducible if it cannot be written as a product of two non-constant polynomials. There is no known algorithm to decide the irreducibility of real polynomials in several real variables; a case by case analysis is required.

First Lemma. For a reduced homogeneous hyperbolic polynomial P_x, the set N_x^{smooth} is a dense subset of the cone N_x of massless momenta.

Proof. Since the set of massless momenta N_x is generated from a single polynomial P , it follows from Definition 3.3.4 of [22] that the set of singular points is $\text{Sing}(N) = N \setminus N^{\text{smooth}}$. But then $\dim \text{Sing}(N) < \dim N = \dim M - 1$, where the inequality is Proposition 3.3.14 of [22] and the equality follows from the hyperbolicity of P [28]. Thus we know that the singular set is at most of dimension $\dim M - 2$. Further, we know from the first remark in 3.4.7 of [21] that $\text{Sing}(N)$, being a real algebraic set, can be expressed as a finite union of analytic semialgebraic manifolds S_i and that every such manifold has a finite number of connected components. From the propositions 2.8.5 and 2.8.14 of [22] we thus obtain that $\dim \text{Sing}(N) = \max(\dim(S_i)) = \max(d(S_i))$, where $d(S_i)$ is the topological dimension of the semialgebraic submanifold $S_i \subset T_x^*M$. Since $\dim \text{Sing}(N) \leq \dim M - 2$ we conclude that $\text{Sing}(N)$ consists of only finitely many submanifolds of $\mathbb{R}^n$ of topological dimension less or equal to $\dim M - 2$. Thus its complement $N^{\text{smooth}} = N \setminus \text{Sing}(N)$ is dense in N.

Second Lemma. If P_x is a reduced homogeneous hyperbolic polynomial with hyperbolicity cone C_x at some point $x \in M$, then for all covectors $s \in T_x^*M \backslash \text{closure}(C_x)$ there exists a massless covector r on the boundary ∂C_x of the hyperbolicity cone such that $s(DP_x(r)) < 0$.

Proof. It is clear that if $y \in C_x$ and $s \notin \text{closure}(C_x)$, the line $y + \lambda s$ intersects the boundary ∂C_x at some $r_0 = y + \lambda_0 s$ for some positive λ_0. Thus $P_x(r_0) = 0$ and, since $P_x(C_x) > 0$, we have $P_x(r_0 - \epsilon s) > 0$ for sufficiently small positive ϵ. Now we must distinguish two cases: First assume that $P_x(r_0 + \epsilon s) < 0$, from which it follows that $\frac{d}{d\epsilon} P_x(r_0 + \epsilon s)|_{\epsilon=0} = s(DP_x(r_0)) < 0$, which proves the lemma with $r := r_0$; Second, assume that $P_x(r_0 + \epsilon s) > 0$ which is equivalent to $\frac{d}{d\epsilon} P_x(r_0 + \epsilon s)|_{\epsilon=0} = s(DP_x(r_0)) = 0$ which in turn holds if and only if $DP_x(r_0) = 0$ (to see the latter equivalence assume that, to the contrary, $s(DP_x(r_0)) = 0$ and $DP_x(r_0) \neq 0$; this implies that s must be tangential to ∂C_x at r_0, but since y lies in C_x and C_x is a convex cone $y + \lambda s$ could then not intersect ∂C_x at r_0, which we however assumed). So to prove the lemma in this second case, we need to construct another $r_0' \in \partial C_x$ that satisfies the condition $s(DP_x(r_0')) < 0$. Now since the First Lemma guarantees that the set N_x^{smooth}, on which DP_x is non-zero, lies dense in N_x, we can find in every open neighbourhood U around r_0 a vector $r_0' \in \partial C_x$ such that $DP_x(r_0') \neq 0$. We define $z := r_0' - r_0$ and $y' := y + z$. Since C_x is an open cone, y' lies in C_x if we choose the neighbourhood small enough, and the line $y' + \lambda s$ intersects ∂C_x at r_0'. Finally since $r_0' \in \partial C_x$ we know that $P_x(r_0') = 0$ and $P_x(r_0' - \epsilon s) > 0$. We conclude that $s(DP_x(r_0')) < 0$. This proves the second lemma with $r := r_0'$.

3.2. Solution to the eikonal equation

In the previous section, we have seen that the eikonal equation determines the massless dispersion relation and that its solution guarantees the construction of an approximate solution of equation (2.1) to any desired order. However, it still remains to guarantee the existence of a solution to the eikonal equation. Based on [29], we will see in this section that for our case, with P hyperbolic, a solution can always be constructed. In fact, the problem is reduced to find a smooth family of solutions to the system of Hamilton's equations

$$\dot{x}^a = \frac{\partial P_x(q)}{\partial q^a}\,, \qquad \dot{q}^a = -\frac{\partial P_x(q)}{\partial x^a}\,, \tag{3.8}$$

with $a = 0, \ldots, \dim M - 1$ and initial data $(x(0), q(0) = (\partial S/\partial x)(0))$ under the condition $q(0) \in N^{\text{smooth}}_{x(0)}$.

A solution $\gamma(\tau) = (x(\tau), q(\tau))$ of these equations describes a curve in the cotangent bundle T^*M which is called a *bicharacteristic*. The projection of $\gamma(\tau)$ onto the base manifold M is then called the *ray trajectory* associated with that bicharacteristic.

To see that it is indeed sufficient to find a solution to the equations (3.8), let us assume that $S(x)$ is a solution of the eikonal equation in an open region of M, and $\gamma(\tau)$ a bicharacteristic passing through the point $\gamma_0 = (x(0), (\partial S/\partial x)(0))$. Since the Hamiltonian $H(x, q) = P_x(q)$ does not depend explicitly on τ, and by assumption $P_{x(0)}(q(0)) = 0$, it follows that $P_{x(\tau)}(q(\tau)) = 0$ for all values of the parameter τ. Moreover, we have

$$q_a = \frac{\partial S}{\partial x^a}$$

on the ray $x(\tau)$, so that

$$\frac{dS}{d\tau} = \dot{x}^a \frac{\partial S}{\partial x^a} = q_a \dot{x}^a\,.$$

If the value $S(x(0))$ is given, by integrating the above equation we obtain

$$S(x(\tau)) = S(x(0)) + \int_{\tau=0}^{\tau} q_a(\tau')dx^a(\tau')\,,$$

which expression provides the value of the eikonal function S along the ray.

Keeping this in mind, we now consider a $(\dim M - 1)$–dimensional manifold Σ embedded in M by the map

$$\Sigma \hookrightarrow M\,, \quad y \to x(y)\,,$$

where $y^\alpha = (y^1, \ldots, y^{\dim M - 1})$ are local coordinates for Σ. We also provide the values of the eikonal function S and its gradients

$$S(x(y)) = S_0(y) \quad \text{and} \quad \frac{\partial S}{\partial x}(x(y)) = q_0(y) \tag{3.9}$$

along the initial manifold Σ. The conditions

$$P(x(y), q_0(y)) = 0 \quad \text{and} \quad (dS)_0(y) = q_0(y)\, dx(y) \tag{3.10}$$

must clearly be satisfied. A crucial assumption is now that $q_0(y) \in N^{\text{smooth}}_x(y)$ for all $x(y)$, which can be made because we know from the first lemma in the preceding section that $N^{\text{smooth}}_x(y)$ is a dense set.

The initial data (3.9) induce the $y-$family of initial data

$$x_0 = x(\tau = 0) = x(y)\,, \quad q_0 = q(\tau = 0) = q_0(y) \tag{3.11}$$

for equations (3.8). Hence, by solving these equations, one obtains the family of bicharacteristics $\gamma(\tau, y) = (x(\tau, y), q(\tau, y))$, and by integrating the second equation in (3.10) one finds

$$S(\tau, y) = S_0(y) + \int_{\tau=0}^{\tau} q_a(\tau', y) dx^a(\tau', y) \,. \tag{3.12}$$

Finally, we assume that Σ is given such that the vectors $\{\dot{x}(y), \partial x/\partial y^\alpha\}$ provide a basis of $T_{x(y)}M$ at all points of Σ (notice that this could never be the case if we did not consider $q_0(y) \in N_x^{\text{smooth}}(y)$). Because only if the last condition is met, are the functions $x(\tau, y)$ locally invertible and we obtain $\tau = \tau(x)$ and $y = y(x)$, and thus the solution $S = S(x)$ to the eikonal equation by replacing the obtained $(\tau(x), y(x))$ into (3.12).

We therefore conclude that a solution to the eikonal equation can always be constructed under the assumption that P is hyperbolic, because then we can guarantee that the set N_x^{smooth} at each point of the initial value hypersurface Σ is non-empty.

3.3. Massless duality theory: the Gauss map and action for massless particles

As we saw in section 3.1, the cotangent bundle function P determines the (smooth) massless momentum cone. The converse question, namely under which conditions the massless momentum cone N_x at a point x determines the polynomial P_x up to a constant factor, is subtle, but of central importance. The vanishing sets associated with polynomials are the subject of study of algebraic geometry (to be studied in the next section) and we will indeed have opportunity to employ some elaborate theorems of real algebraic geometry. We begin by clarifying the relation between vanishing sets of real polynomials and the principal ideals that these polynomials generate, since this will be relevant for our study of massless particles. Recall that an ideal $I \subset R$ in a ring R is a subset that is closed under addition and under multiplication with an arbitrary ring element. Concretely, R is here the ring of real polynomials on T_x^*M in $\dim M$ real variables. Now on the one hand, we may consider the situation where we are given an ideal I and define the vanishing set $\mathcal{V}(I)$ as the set of cotangent vectors that are common zeros to all polynomials in I. On the other hand, we may be given a subset S of cotangent space and consider the set $\mathcal{I}(S)$ of all polynomials in R that vanish on all members of that set S. Now it can be shown that $\mathcal{I}(S)$ is an ideal in the ring of polynomials on cotangent space, and that one always has the inclusion

$$\mathcal{I}(\mathcal{V}(I)) \supseteq I \,. \tag{3.13}$$

The question under which conditions equality holds is studied in the Nullstellensätze of algebraic geometry. While this is a relatively straightforward question for polynomials over algebraically closed fields [30], such as the complex numbers, for the real numbers underlying our study here, one needs to employ a string of theorems that were originally developed in order to solve Hilbert's seventeenth problem. The central result for our purposes is

PROPOSITION 3.3.1. *Let P_x be a reduced homogeneous hyperbolic polynomial on T_x^*M, then the equality*

$$\mathcal{I}(N_x) = \langle P_x \rangle \tag{3.14}$$

holds.

Proof. From the first lemma in section 3.1, we know that $N_x^{\text{smooth}} \neq \emptyset$ (as follows from the hyperbolicity of P_x). Here $\langle P_x \rangle$ denotes the ideal containing all polynomials that have P_x as a factor. Drawing on the said results from real algebraic geometry, this is seen as follows. Let $P_{x\,i}$ be the ith irreducible factor of P_x. Then there exists a $q \in N^{\text{smooth}}(P_{x\,i})$ so that corollary 2.9 of [31] shows that $P_{x\,i}$ generates a real ideal, i.e., $\mathcal{I}(N(P_{x\,i})) = \langle P_{x\,i} \rangle$. According to corollary 2.8 of [31], the reduced polynomial P_x thus also generates a real ideal since it does not contain repeated factors. Finally, theorem 4.5.1 of [22] yields the claim.

The equality (3.14) will play a significant technical rôle in ensuring that we can determine the vector duals of massless momenta using the elimination theory that will be presented in the following section.

In order to associate velocity vectors with massless particle momenta in physically meaningful fashion, we employ the dynamics of free massless point particles. Their dynamics, in turn, are uniquely determined by the dispersion relation, because the Helmholtz action

$$I_0[x, q, \lambda] = \int d\tau \, [q_a \dot{x}^a + \lambda P(x, q)] \tag{3.15}$$

corresponding to the pure constraint Hamiltonian λP obviously describes free massless particles. In the following, we wish to eliminate the momentum q and the Lagrange multiplier λ to obtain an equivalent action in terms of the particle trajectory x only. Variation of the Helmholtz action with respect to λ of course enforces the null condition for the particle momentum. Now variation with respect to q yields $\dot{x} = \lambda \, DP_x(q)$ for all $q \in N^{\text{smooth}}$, which implies the weaker equation

$$[DP_x(q)] = [\frac{\dot{x}}{\lambda}] \,, \tag{3.16}$$

where $[X]$ denotes the projective equivalence class of all vectors collinear with the vector X.

In order to solve (3.16) for q, we need the inverse of the projective map $[DP]$. We will now derive that this inverse is given by the gradient of a so-called dual polynomial. Indeed, the image $N_x^{\#}$ of the massless covector cone N_x under the gradient map DP is again described by a homogeneous polynomial $P_x^{\#}$, albeit of generically different degree than P. More precisely, for an irreducible cotangent bundle function P, we look for a likewise irreducible tangent bundle function $P^{\#}$ that is uniquely determined up to a real constant factor at each point x of the manifold by the equation

$$P_x^{\#}(DP_x(N_x^{\text{smooth}})) = 0 \,. \tag{3.17}$$

The polynomials P_x and $P_x^{\#}$ given by P and $P^{\#}$ at any given point x of the base manifold are then called dual to each other, and it is convenient to also call the corresponding cotangent bundle function P and tangent bundle function $P^{\#}$ dual to each other. For a cotangent bundle function P that is reducible into irreducible factors

$$P(x, k) = P_1(x, k) \cdots P_f(x, k) \,, \tag{3.18}$$

we define the dual tangent bundle function as the product

$$P^{\#}(x, v) = P_1^{\#}(x, v) \cdots P_f^{\#}(x, v) \,, \tag{3.19}$$

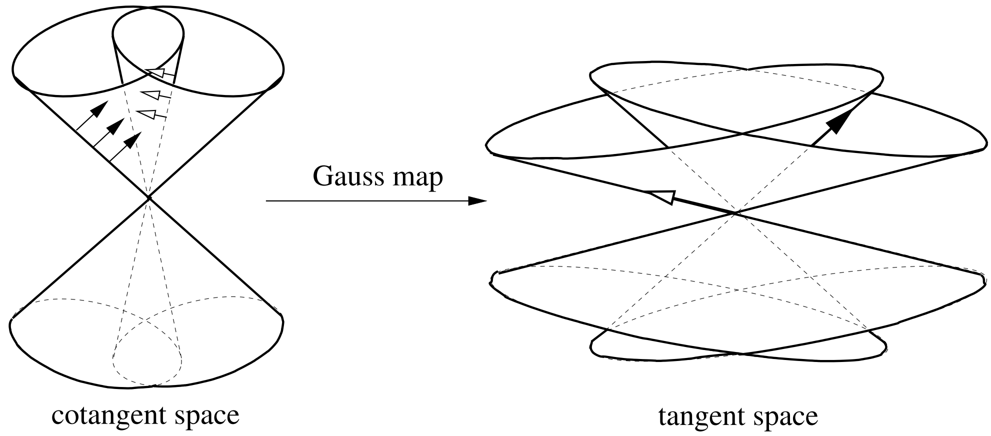

FIGURE 3.1. Gauss map sending the zero locus of a polynomial to the zero locus of a dual polynomial

where the $P_i^\#$ are the irreducible duals of the irreducible P_i determined by equation (3.17). Thus $P^\#$ is uniquely determined up to a real factor function on M and satisfies again equation (3.17), as one easily sees from application of the product rule. The existence of a dual $P_x^\#$, and indeed its algorithmic computability for any reduced hyperbolic polynomial P will be the subject of the next section. For the remaining part of the present section, we assume that a dual polynomial $P_x^\#$ indeed exists.

Thus equipped with the notion of the dual polynomial, we may now return to the projective gradient map

$$[DP_x] : [N_x^{\text{smooth}}] \to [N_x^\#] , \qquad [q] \mapsto [DP_x(q)] \tag{3.20}$$

first encountered in (3.16), where the brackets denote projective equivalence classes, identifying parallel vectors (respectively covectors), but not antiparallel ones, and $N_x^\#$ is the image of N_x^{smooth} under DP_x. The projective map $[DP_x]$ is well-defined due to the homogeneity of P_x, and will be referred to as the Gauss map. The problem of inverting the Gauss map is now solved by definition of the dual Gauss map $[DP_x^\#]$ in terms of the dual polynomial $P_x^\#$,

$$[DP_x^\#] : [N_x^{\#\ \text{smooth}}] \to [N_x] , \qquad [X] \mapsto [DP_x^\#(X)] , \tag{3.21}$$

since we then have for null covectors $k \in N_x^{\text{smooth}}$ that

$$[DP_x^\#]([DP_x]([k])) = [k] \qquad \text{if } \det(DDP_x)(k) \neq 0 , \tag{3.22}$$

so that the dual Gauss map $[DP^\#]$ acts as the inverse of the Gauss map on the images of all covectors k satisfying the above determinantal non-degeneracy condition. That relation (3.22) holds is most easily seen from rewriting the duality condition (3.17) in the form

$$P^\#(x, DP(x,k)) = Q(k)P(k) \qquad \text{for all covectors } k , \tag{3.23}$$

since this form does not require an explicit restriction to null covectors. Thus differentiation with respect to k yields, by application of the chain rule and then of Euler's theorem[2] on the right

[2]Euler's theorem asserts the simple fact that for any function f that is homogeneous of degree $\deg f$, the relation $Df(v)v = (\deg f)f(v)$ holds for any v in the domain of f.

hand side, for any null covector k satisfying the non-degeneracy condition in (3.22) that

$$DP^{\#}(x, DP(x,k)) = \frac{Q(x,k)}{\deg P - 1} k \,, \tag{3.24}$$

which in projective language takes the form (3.22). In particular, we may thus solve the projective equation (3.16) for

$$[q] = [DP^{\#}_x]([\dot{x}/\lambda]) \,. \tag{3.25}$$

Obviously, the homogeneity of $DP^{\#}_x$ in conjunction with the projection brackets allows to disregard the function λ altogether. However, another undetermined function μ appears when translating this result back to non-projective language,

$$q = \mu \, DP^{\#}_x(\dot{x}) \,. \tag{3.26}$$

Now we may replace the momentum in (3.15) by this expression and use again Euler's theorem applied to the homogeneous polynomial $P^{\#}_x$ to finally obtain the massless point particle action

$$I_0[x,\mu] = \int d\tau \mu \, P^{\#}(x,\dot{x}) \,. \tag{3.27}$$

Relations (3.16) and (3.25) reveal the physical meaning of the Gauss map $[DP_x]$ and its inverse $[DP^{\#}_x]$: up to some irrelevant conformal factor, they associate null particle momenta in N_x^{smooth} with the associated null particle velocities in $N_x^{\#\,\text{smooth}}$. The automatic appearance of a final Lagrange multiplier μ in (3.27) also hardly comes as a surprise, since it is needed to enforce the null constraint $P^{\#}_x(\dot{x}) = 0$. This reveals the direct physical relevance of the dual tangent bundle function $P^{\#}$ as the tangent space geometry seen by massless particles.

Furthermore, comparing equation (3.8) with equation (3.16), we also conclude that the problem of finding the equations which describe the trajectory of rays (and hence of the bicharacteristics in the tangent bundle) in the geometric optical limit is reduced to finding the dual polynomial of the principal symbol of the associated field equations. This is of high relevance when studying Maxwell electrodynamics on metric and area metric manifolds (which we will study in chapter 6) since this describes how light rays propagate in linear optical media.

3.4. Existence and computability of the dual polynomial

In the previous section, we have seen that in order to develop the dual theory associating null particle momenta in N_x^{smooth} with null particle velocities in $N_x^{\#\,\text{smooth}}$, the dual polynomial $P^{\#}$ defined by (3.17) must exist. Proposition 3.4.3 below will indeed imply the existence of a dual polynomial $P^{\#}$ for any hyperbolic polynomial P. Moreover, we describe Buchberger's algorithm, which allows to compute the dual polynomial explicitly.

This will require to develop some techniques from algebraic geometry, which we quickly review following [30]. So we consider the ring $\mathbb{R}[p_1, \dots, p_{\dim M}]$ of polynomials in the variables $p_1, \dots, p_{\dim M}$ having coefficients in the real field $\mathbb{R}$ and its respective ideals. We will also use the multi-index notation $\alpha = (\alpha_1, \dots, \alpha_{\dim M})$ such that an arbitrary element f of $\mathbb{R}[p_1, \dots, p_{\dim M}]$ is written as $f = \sum_\alpha c_\alpha p^\alpha$ with $c_\alpha = c_{\alpha_1 \dots \alpha_{\dim M}} \in \mathbb{R}$ and $x^\alpha = p_1^{\alpha_1} \cdots p_{\dim M}^{\alpha_{\dim M}}$. By $\langle f_1, \dots, f_r \rangle$ we will denote the ideal contained in $\mathbb{R}[p_1, \dots, p_{\dim M}]$ generated by the set of polynomials $f_1, \dots, f_r$.

Given an ideal $\langle f_1, \dots, f_r\rangle$ and an arbitrary polynomial $g \in \mathbb{R}[p_1, \dots, p_{\dim M}]$, we want to decide whether $g \in \langle f_1, \dots, f_r\rangle$. This requires the notion of Gröbner bases and the use of the division algorithm, which we now review.

DEFINITION. *A monomial in $\mathbb{R}[p_1, \dots, p_{\dim M}]$ is an element of the form $p_1^{\alpha_1} \cdots p_{\dim M}^{\alpha_{\dim M}}$. We say that $>$ is a monomial order in $\mathbb{R}[p_1, \dots, p_{\dim M}]$ if $>$ satisfies*

(1) if $x^\alpha > x^\beta$, then $x^\gamma x^\alpha > x^\gamma x^\beta$ for all multi-indices α, β, γ and

(2) a well-ordering property: an arbitrary set of monomials $\{x^\alpha\}_{\alpha \in A}$ has a least element.

In particular, condition (2) ensures that any decreasing sequence of monomials $x^{\alpha(1)} > x^{\alpha(2)} > \cdots$ eventually terminates.

Example. A monomial order is provided by the so-called lexicographic order. This is defined such that $x^\alpha > x^\beta$ if the first non-zero entry of $(\alpha_1 - \beta_1, \dots, \alpha_{\dim M} - \beta_{\dim M})$ is positive.

DEFINITION. *Given a monomial order $>$ on $\mathbb{R}[p_1, \dots, p_{\dim M}]$ and one of its elements $f = \sum_\alpha c_\alpha x^\alpha$, the leading monomial $LM(f)$ of f is defined as the largest monomial x^α such that $c_\alpha \neq 0$. The leading term $LT(f)$ of f is then defined as $LT(f) = c_\alpha x^\alpha$, where $LM(f) = x^\alpha$.*

DEFINITION. *Fix a monomial order $>$ and let $I \subset \mathbb{R}[p_1, \dots, p_{\dim M}]$ be an ideal. A Gröbner basis for I is a collection of non-zero polynomials $\{f_1, \dots, f_r\} \subset I$ such that $LT(f_1), \dots, LT(f_r)$ generate $LT(I)$, where $LT(I) := \langle LT(g) | \, g \in I\rangle$.*

With these definitions at hand, we can now present the division algorithm which will then lead to proposition 3.4.1.

DIVISION ALGORITHM: Fix a monomial order $>$ in the ideal $\mathbb{R}[p_1, \dots, p_{\dim M}]$ and non-zero polynomials $\{f_1, \dots, f_r\}$ generating the ideal $I = \langle f_1, \dots, f_r\rangle$. Given an arbitrary polynomial $g \in \mathbb{R}[p_1, \dots, p_{\dim M}]$, the division algorithm proceeds as follows. Put $g_0 = g$ and construct the polynomials

$$g_{i+1} = g_i - f_{j_i} \frac{\mathrm{LT}(g_i)}{\mathrm{LT}(f_{j_i})} \, ,$$

where f_{j_i} is one element of the set of polynomials $\{f_1, \dots, f_r\}$ such that $\mathrm{LM}(f_{j_i})$ divides $\mathrm{LM}(g_i)$. If there is no polynomial in $\{f_1, \dots, f_r\}$ such that $\mathrm{LM}(f_{j_i})$ divides $\mathrm{LM}(g_i)$, then the procedure terminates.

We then have the following

PROPOSITION 3.4.1. (proposition 2.13 of [30]) *Let $I \subset \mathbb{R}[p_1, \dots, p_{\dim M}]$ be an ideal and $f_1, \dots, f_r$ a Gröbner basis for I. The division algorithm terminates in a finite number of steps, with either $g_i = 0$ or $LT(g_i)$ not divisible by any of the leading terms $LT(f_j)$. In the first case, we can represent g as $g = \sum_{s=1}^r h_s f_s$ with $h_s \in \mathbb{R}[p_1, \dots, p_{\dim M}]$ and hence $g \in I$. In the second case, $g = \sum_{s=1}^r h_s f_s + g_i$ and hence $g_i \notin I$.*

With this proposition, we can decide whether a polynomial g is an element of a given ideal I. Indeed, in order to decide whether $g \in I$, it is enough to find a Gröbner basis for I and then apply the division algorithm. As a corollary of this proposition, it also follows that if I is an ideal and $f_1, \dots, f_r$ a Gröbner basis for I, then $I = \langle f_1, \dots, f_r\rangle$. Existence of Gröbner bases is ensured by the Hilbert's basis theorem (theorem 2.21 of [30]), which asserts that every ideal $I \subset \mathbb{R}[p_1, \dots, p_{\dim M}]$ admits a Gröbner basis and, as a consequence, every polynomial ideal is finitely generated.

Buchberger's algorithm provides an explicit algorithm for the construction of Gröbner bases. Before presenting this algorithm, we define the least common multiple LCM of monomials x^α and x^β as

$$\mathrm{LCM}(x^\alpha, x^\beta) = x_1^{max(\alpha_1,\beta_1)} \cdots x_n^{max(\alpha_n,\beta_n)}.$$

We then have

Buchberger's algorithm. Fix a monomial order $>$ on $\mathbb{R}[p_1, \dots, p_{\dim M}]$. A Gröbner basis for the ideal $\langle f_1, \dots, f_r\rangle$ is then obtained by iterating the following procedure: For all i, j apply the division algorithm to the polynomials

$$S(f_i, f_j) = \frac{\mathrm{LCM}(\mathrm{LM}(f_i), \mathrm{LM}(f_j))}{\mathrm{LT}(f_i)} f_i - \frac{\mathrm{LCM}(\mathrm{LM}(f_i), \mathrm{LM}(f_j))}{\mathrm{LT}(f_j)} f_j$$

in order to get the expressions

$$S(f_i, f_j) = \sum_{l=1}^{r} h(ij)_l f_l + r(ij)\,,$$

where each $\mathrm{LM}(r(ij))$ is not divisible by any of the $\mathrm{LM}(f_l)$. If all remainders $r(ij) = 0$ then $f_1, \dots, f_r$ are already a Gröbner basis. Otherwise we let $f_{r+1}, \dots, f_{r+s}$ denote the non-zero remainders $r(ij)$ and adjoin these to get a new set of generators.

A key result for us is the following theorem on elimination theory.

Theorem 3.4.2. (Elimination theorem in [30]). *Let $J \subset \mathbb{R}[p_1, \dots, p_{\dim M}, v^1, \dots, v^{\dim M}]$ be an ideal and $>$ a monomial order in $\mathbb{R}[p_1, \dots, p_{\dim M}, v^1, \dots, v^{\dim M}]$ such that*[3] *if $LM(g) \in \mathbb{R}[v^1, \dots, v^{\dim M}]$, then $g \in \mathbb{R}[v^1, \dots, v^{\dim M}]$. Let $\{f_1, \dots, f_r\}$ be a Gröbner basis for J with respect to $>$. Then $J \cap \mathbb{R}[v^1, \dots, v^{\dim M}]$ is generated by the elements of the Gröbner basis contained in $\mathbb{R}[v^1, \dots, v^{\dim M}]$. More precisely,*

$$J \cap \mathbb{R}[v^1, \dots, v^{\dim M}] = \left\langle f_j \mid f_j \in \mathbb{R}[v^1, \dots, v^{\dim M}] \right\rangle \subset \mathbb{R}[v^1, \dots, v^{\dim M}]\,.$$

We now recall that given a real vector space V, the projective space $\mathbb{P}V$ associated with V, is defined as $\mathbb{P}V = (V \setminus \{0\})/\sim$, where $\sim$ denotes the equivalence relation such that $u \sim v$ if there is a $\lambda \in \mathbb{R}$ with $\lambda \neq 0$ such that $u = \lambda v$ for $u, v \in V \setminus \{0\}$. The notation $[v]$ is used to denote the resulting equivalence class containing the element v.

Turning now to our specific problem, given a homogeneous polynomial $P_x : T_x^*M \to \mathbb{R}$, we define the hypersurface or projective variety defined by P_x as

$$S(P_x) = \{[q] \in \mathbb{P}(T_x^*M) \mid P(q) = 0\}\,.$$

[3] Such a monomial order is called an elimination order, see definition 4.6 of [30].

More generally, given a homogeneous ideal $J \subset \mathbb{R}[T_x^*M]$ (an ideal generated only by homogeneous polynomials), we define

$$S(J) = \{[q] \in \mathbb{P}(T_x^*M) \,|\, P(q) = 0 \text{ for each homogeneous } P \in J\}$$

as the projective variety defined by J. On the other way around, given $S \subset \mathbb{P}(T_x^*M)$, the homogeneous ideal vanishing along S is defined as

$$J(S) = \langle P \in \mathbb{R}[T_x^*M] \text{ homogeneous } \,|\, P(q) = 0 \text{ for all } q \in S\rangle \ .$$

Given a hypersurface $S \subset \mathbb{P}(T_x^*M)$ generated by a homogeneous polynomial $P : T_x^*M \to \mathbb{R}$, the dual variety $\tilde{S}$ is defined as the closure[4] of the locus of all hyperplanes tangent to S at smooth points $q \in S$. We say that $q \in S$ is a smooth point of S if $D_qP(q) \neq 0$. We then have the following crucial theorem describing the image of the Gauss map found in the previous section.

THEOREM 3.4.3. (Proposition 11.10 of [30]). *Let S be a hypersurface with $J(S) = \langle P_x\rangle$. Then the dual variety $\tilde{S}$ is the image in the projective tangent space $\mathbb{P}(T_xM)$ of the so defined Gauss map*

$$S(P_x) \to \mathbb{P}(T_xM)\,, \qquad [q] \mapsto [DP_x(q)]\,.$$

Furthermore, the equations determining the image of the Gauss map are explicitly given by

$$\Big\langle P_x(q), \quad X^1 - DP_x(q)^1, \quad \dots, \quad X^{\dim M} - DP_x(q)^{\dim M}\Big\rangle \cap \mathbb{R}[X^1, \dots, X^{\dim M}]\,. \tag{3.28}$$

But this theorem automatically holds if P_x reduced and hyperbolic, as follows from proposition 3.3.1, because then $J(S) = \langle P_x\rangle$ is satisfied. This already guarantees the existence of the dual polynomial $P_x^\#$ because it can then be constructed as follows: from theorem 3.4.2, by using Buchberger's algorithm and Gröbner basis, we can compute the elimination ideal in equation (3.28). The so calculated elimination ideal, however, may turn out to be generated by several real homogeneous polynomials. However, making use of the fact that we are dealing with real polynomials, it is easy to construct the dual $P^\#$ as a sum of appropriate even powers of the generating polynomials, which obviously vanishes where and only where all generators vanish. It should be said that while for most polynomials of interest, a direct calculation of dual polynomials using elimination theory exhausts the capability of current computer algebra systems, in many cases one is nevertheless able to guess the dual polynomial by physical reasoning (as we will illustrate for the cases of metric and area metric geometry in chapter 6). Once such an educated guess has been obtained, one may directly use the defining equation (3.17) to verify that one has found the dual polynomial. In any case, since its existence is guaranteed, we will simply assume in the following that a dual $P^\#$ has been found by some method.

3.5. Time-orientability

In section 3.3, we saw that the dual polynomial defined by the tangent bundle function $P^\#$ in each tangent space plays an essential rôle.

[4]The closure $\bar{S}$ of a set S is the smallest closed set containing S, in this case with respect to the Zariski topology. In the Zariski topology, the open sets are of the form $U = \{v \in V \,|\, P_j(v) \neq 0,\, j = 1, \dots, r,\, P_j \in \mathbb{R}[V]\}$.

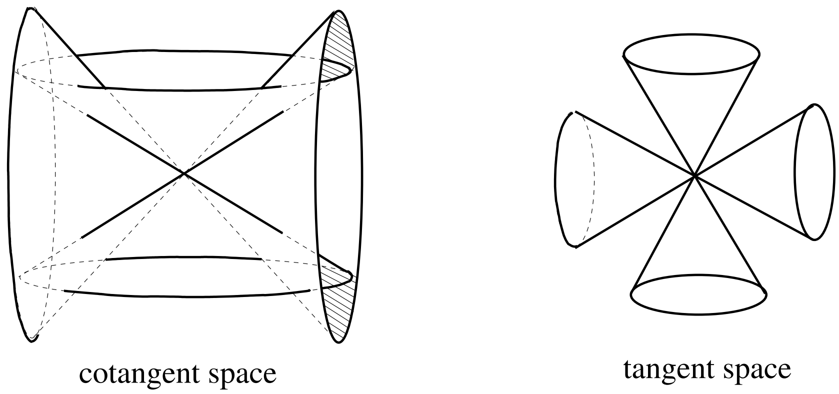

FIGURE 3.2. Example of a hyperbolic polynomial with non-hyperbolic dual polynomial; shown are the respective vanishing sets.

The condition for matter field dynamics on (M, G) to be interpretable is translated to the condition that the geometry (M, G) be time-orientable, since time-orientability allows for a good notion of observers. We will now see that in order to have time-orientability, one needs to restrict attention to spacetime geometries in which both P and its dual $P^{\#}$ are hyperbolic. Thus, when P and $P^{\#}$ are indeed hyperbolic, we will simply say that P is hyperbolic (hyperbolicity of P) and time-orientable (hyperbolicity of $P^{\#}$). More precisely, this corresponds, from the definition of a hyperbolic polynomial, to the requirement that there is a vector field T such that for every vector field R there are only real functions μ on M such that

$$P^{\#}(x, R(x) + \mu(x)T(x)) = 0$$

everywhere. A vector field T satisfying this property will be called a time-orientation on M. A time-orientation is thus equivalent to a choice of a hyperbolicity cone $C^{\#}$ of $P^{\#}$ (which is defined precisely as the hyperbolicity cones of P in section 2.2) and we stipulate that it contain the tangent vectors to admissible observers at a spacetime point (we will prove a non-trivial consistency result concerning the stability of the so defined observers in chapter 5). Having chosen the observer cone $C^{\#}$ in the tangent bundle, however, one can immediately see that those momenta p at a point x whose energy is positive from every observer's point of view now constitute a convex cone

$$(C^{\#}_x)^{\perp} = \{p \in T^*_xM \,|\, p(v) > 0 \text{ for all } v \in C^{\#}\}\,. \tag{3.29}$$

For the simple Lorentzian metric case, this is just (the closure of) what has been chosen as the forward cone, while in general the situation is more complicated. If the polynomial P is of the product form (3.5), we find that the positive energy cone is simply the sum of the positive energy cones coming from the duals of the factors of P [32], i.e.

$$(C^{\#})^{\perp} = (C^{\#}_1)^{\perp} + \cdots + (C^{\#}_l)^{\perp}\,, \tag{3.30}$$

where the sum of two convex sets is just the set of all sums of any two elements of the two sets.

That predictivity (mathematically: hyperbolicity of P) does not imply interpretability (mathematically: hyperbolicity of $P^{\#}$) is illustrated by the counterexample in figure 3.2. Thus the condition of P and $P^{\#}$ being hyperbolic indeed presents a rather stringent condition on the geometry underlying a classical spacetime. This will be illustrated in chapter 6, first for the instructive case of metric geometry, where the hyperbolicity and time-orientability conditions each amount to the requirement that the metric be Lorentzian, and second for the case of area metric geometry, where a similar exclusion of algebraic classes follows from the hyperbolicity and time-orientability condition for P in conjunction with the quantizability condition. A similar study one may—and indeed must—conduct for any other candidate for a spacetime geometry.

In summary, we have learnt in this chapter that P_x defines the massless dispersion relation as the solvability condition for the eikonal equation, which can always be solved for hyperbolic polynomials P_x. Moreover, in order to obtain the action for massless particles, it is necessary to invert the Gauss map defined by $[DP_x]$. The inverse of the Gauss map is given by $[DP_x^{\#}]$, where $P_x^{\#}$ is the dual polynomial to P_x, and whose existence is guaranteed for a reduced and hyperbolic polynomial P_x. The dual polynomial $P_x^{\#}$ also encodes the geometry seen by massless particles in tangent space. Finally, requiring $P^{\#}$ to be hyperbolic one can choose a vector field that is hyperbolic with respect to $P^{\#}$ everywhere on the manifold, which vector field then corresponds to a time-orientation on the manifold. Thus if matter field dynamics on (M, G) is to be predictive and interpretable, we found that P must be hyperbolic and time-orientable, meaning the hyperbolicity of both P and $P^{\#}$.

CHAPTER 4

Condition III: Energy-distinguishability

In this chapter, we introduce the energy-distinguishing condition as the third and last condition that a pair (M, G) must satisfy in order to be able to carry (in addition to predictive and interpretable) quantizable matter field dynamics, and therefore to provide a classical spacetime structure. This condition will enable us to study the behaviour of massive point particles and thus develop the necessary dual theory relating velocity vectors with massive particle momenta.

4.1. Energy-distinguishing condition

As seen at the end of the previous section, having chosen the observer cone $C^{\#}$ in the tangent bundle by way of a time-orientation, the convex set $(C_x^{\#})^{\perp}$ constitutes the set of those momenta p at a point x whose energy is positive from every observer's point of view. The quantizability requirement on matter field dynamics now translates into the requirement that the geometry (M, G) be energy-distinguishing, which refers then to the requirement that any massless momentum q have either positive or negative energy, independent of which observer measures the energy. More precisely, we require the set N of massless non-zero covector fields to disjunctively decompose into positive and negative energy parts

$$N = N^{+} \,\dot{\cup}\, N^{-}\,, \tag{4.1}$$

where N^{+} is defined as the intersection of N with the positive energy cone $(C^{\#})^{\perp}$, and N^{-} as the intersection with the negative energy cone $(-C^{\#})^{\perp}$. *We will refer to such hyperbolic and time-orientable cotangent bundle functions P as energy-distinguishing.* Figure 4.1 shows the vanishing sets of a hyperbolic, time-orientable and energy-distinguishing polynomial P_x and its dual $P_x^{\#}$.

For geometries that are hyperbolic, time-orientable and energy-distinguishing, we have the following important

Proposition 4.1.1. *For hyperbolic, time-orientable and energy-distinguishing geometries, the set of massless momenta N_x cannot contain any null planes in spacetime dimensions $d \geq 3$, which in turn implies that the degree of P cannot be odd*

Proof. First, we prove that the hyperbolicity and time-orientability of P_x implies that

$$\text{closure}(C_x^{\#\perp}) \cap -\text{closure}(C_x^{\#\perp}) = \{0\}\,. \tag{4.2}$$

Let k_0 be such that $k_0 \in \text{closure}(C_x^{\#\perp})$ and $k_0 \in -\text{closure}(C_x^{\#\perp})$. It follows from the definition of the dual cone that the following inequalities are true for all $x \in C_x^{\#} : x.k_0 \geq 0$ and $x.k_0 \leq 0$. If this would be true, the hyperbolicity cone $C_x^{\#}$ had to be a plane or a subset of a plane. That

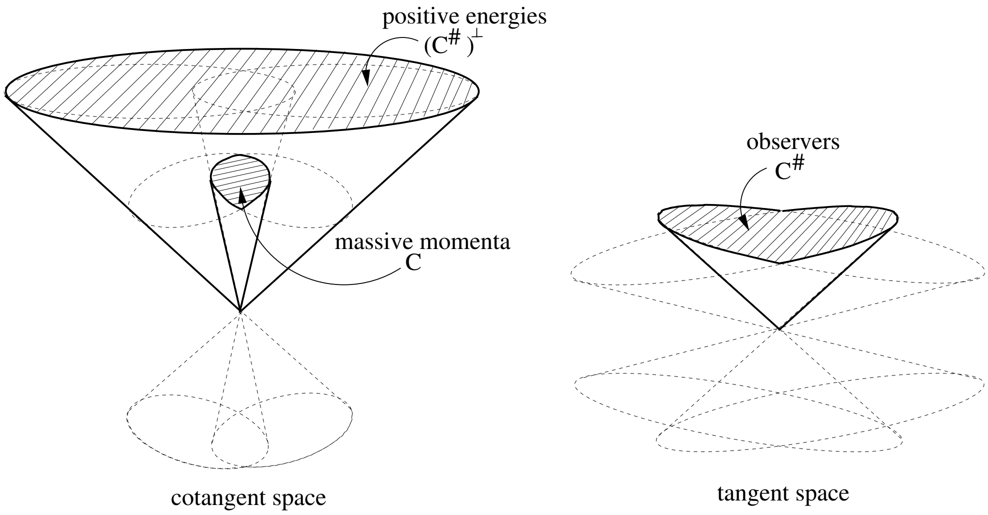

FIGURE 4.1. A hyperbolic, time-orientable and energy-distinguishing polynomial P.

would contradict the property of $C_x^{\#}$ to be open. Second, suppose that the zero set N_x contains a plane. From closure$(C_x^{\#\perp}) \cap -$closure$(C_x^{\#\perp}) = \{0\}$ it follows that $C_x^{\#\perp} \setminus \{0\}$ is a proper subset of a halfspace. A proper subset of a halfspace cannot contain any complete plane through the origin. Hence the existence of a null plane of P_x would obstruct the energy-distinguishing property. Third, this fact immediately restricts us to cotangent bundle functions P of even degree. For suppose $\deg P$ was odd. Then on the one hand, we would have an odd number of null sheets. On the other hand, the homogeneity of P implies that null sheets in a cotangent space come in pairs, of which one partner is the point reflection of the other. Together this implies that we would have at least one null hyperplane.

At this point, we have finally arrived at the insight that *an admissible physical geometry* (M, G) *underlying a classical spacetime must be such that the cotangent bundle function* P *(purely constructed from the geometry* G*) is a hyperbolic, time-orientable and energy-distinguishing reduced hyperbolic homogeneous polynomial in each fibre.* These are now all the conditions on P, and therefore on the geometry, we identify in this work. The following sections in this chapter serve to show that the theory also extends to massive particles.

4.2. The massive dispersion relation

For a hyperbolic, time-orientable and energy-distinguishing cotangent bundle function P, there is always a hyperbolicity cone of P_x in each cotangent space that is of positive energy with respect to a chosen time orientation $C^{\#}$. For let $\tilde{C}_x$ be some hyperbolicity cone of P_x, whose boundary $\partial\tilde{C}_x$ we know to be a connected set of null covectors. Now on the one hand, the complete zero set of P_x is contained in $(\tilde{C}_x^{\#})^\perp \cup -(\tilde{C}_x^{\#})^\perp$ due to the energy-distinguishing property. On the other hand, we have that (4.2) holds. Hence either $\tilde{C}_x$ or $-\tilde{C}_x$ is of positive energy.

The covector fields in the thus selected positive energy cone C play two related rôles. The first rôle, from theorem 2.1.1, was that a hypersurface can only be an initial data surface if its normal covector field lies in C. Now in order to identify the second rôle of the cone C in relation to massive matter, first observe that within the hyperbolicity cone C, the sign of P cannot change (see theorem 2.2.1), which is why we could arrange for P to be positive on C in section 2.2 and everywhere afterwards. Since it is particularly important again in what follows, we re-emphasize that this choice has been made. But then we have for any momentum q in C_x at a spacetime point x that

$$P_x(q) = m^{\deg P} \tag{4.3}$$

for some positive real number $m > 0$, which we call the mass associated with the momentum q. It must be emphasized that the definition of mass associated to a momentum, as provided by (4.3), hinges on the choice of a particular volume density ρ in (2.2). Physically this is understood from the need to convert mass densities in field theory into point masses in particle theory, which conversion requires exactly a definition of volume. But then (4.3) represents a massive dispersion relation whose mass shells foliate the interior of C_x, see figure 4.2. An immediate physical consequence of the convexity of the cone C_x is that even for modified dispersion relations, a decay of a positive energy massless particle into positive energy massive particles is kinematically forbidden.

At this point we derive a further important consequence of the hyperbolicity, time-orientability and energy-distinguishing properties in the following

PROPOSITION 4.2.1. *If the geometry of spacetime is hyperbolic (P_x hyperbolic), time-orientable ($P_x^{\#}$ hyperbolic) and energy distinguishing, then P_x is complete.*

Proof. Recall from section 2.2 that a hyperbolic polynomial P_x is called complete if the lineality space

$$L(P) = \{a \in T_x^*M \,|\, \text{for all } y \in T_x^*M \text{ and } \lambda \in \mathbb{R} : P(y + \lambda a) = P(y)\} \tag{4.4}$$

only contains the zero covector and that completeness is equivalent to

$$\text{closure}(C(P_x, h)) \;\cap\; \text{closure}(C(P_x, -h)) = \{0\}\,. \tag{4.5}$$

Using now the argument given at the start of this section, we know that

$$\text{closure}(C_x^{\#\perp}) \cap -\text{closure}(C_x^{\#\perp}) \supseteq \text{closure}(C_x) \cap -\text{closure}(C_x)\,. \tag{4.6}$$

Thus if the right hand side differs from $\{0\}$ (meaning that P is incomplete), the left hand side will contain non-zero covectors, too (showing that P is not energy-distinguishing). Because of the inclusion, this only holds in this direction. We thus conclude that the energy-distinguishing property already implies completeness.

There are three principal reasons why it is so important that completeness holds. First, completeness will play a crucial rôle in ensuring, as we will see in the next section, that there is a well-defined duality theory associating massive covectors with their vector counterparts.

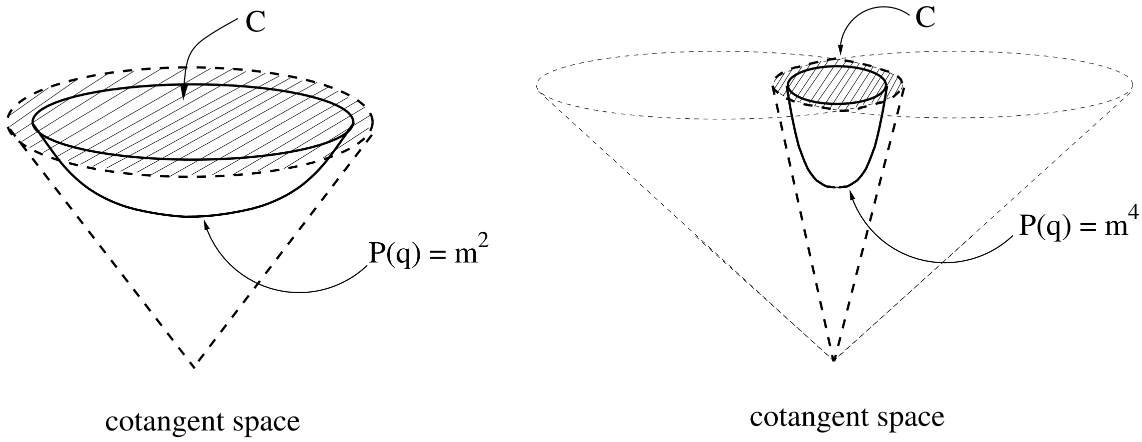

FIGURE 4.2. Mass shells defined by a hyperbolic, time-orientable and energy-distinguishing cotangent bundle functions P. On the left, the familiar second degree Lorentzian case; on the right, a fourth degree case defined by a product of two Lorentzian metrics.

Thus remarkably, the hyperbolicity, time-orientability and the energy-distinguishing properties, originally conceived in the context of massless dispersion relations, also take care of this in the massive case, via completeness. Second, if P_x is complete then theorem 2.2.3 holds, so that taking P_x to be positive everywhere on C_x, we have the reverse triangle inequality

$$P_x^{1/\deg P}(k_1 + k_2) \geq P_x^{1/\deg P}(k_1) + P_x^{1/\deg P}(k_2) \tag{4.7}$$

for all k_1 and k_2 in the same hyperbolicity cone C_x. Equality holds if and only if k_1 and k_2 are all proportional. Physically, this is necessary for the massive dispersion relation to make sense, since the reverse triangle inequality generalizes a familiar result from Lorentzian geometry to any viable dispersion relation in our sense, namely that the decay of a massive particle generically gives rise to a mass defect. Third, recalling from section 2.2 that defining the tensor

$$P_x(k_1, \dots, k_{\deg P}) = \frac{1}{(\deg P)!} \prod_{J=1}^{\deg P} \left(\sum_{i=1}^{\dim V} (k_J)_i \frac{\partial}{\partial k_i} \right) P_x(k) \tag{4.8}$$

as the totally symmetric polarization of the polynomial P_x, we also know from theorem 2.2.3 that the reverse Cauchy-Schwarz inequality

$$P_x(k_1, \dots, k_{\deg P}) \geq P_x(k_1)^{1/\deg P} \cdots P_x(k_{\deg P})^{1/\deg P} \tag{4.9}$$

holds for all $k_1, \dots, k_{\deg P}$ in the same hyperbolicity cone C_x. Similar to the reverse triangle inequality above, equality holds for the reverse Cauchy-Schwarz inequalities if and only if all arguments k_i are proportional to each other.

4.3. Massive duality theory: the Legendre map and action for massive particles

We wish to associate vector duals with massive momenta (having done so for massless momenta in chapter 3), and to this end we employ the Helmholtz action

$$I[x, q, \lambda] = \int d\tau \left[q_a \dot{x}^a - \lambda m \ln P(x, \frac{q}{m}) \right] , \tag{4.10}$$

which describes free particles of mass m, since the massive dispersion relation $P(x,q) = m^{\deg P}$ is enforced through variation with respect to λ. The particular form of the Lagrange multiplier term here has been chosen for the technical reason of having available the theory of Legendre duals on the open convex cones C_x, see [32]. More precisely, the so-called barrier function,

$$f_x : C_x \to \mathbb{R}\,, \qquad f_x(q) = -\frac{1}{\deg P} \ln P_x(q)\,, \tag{4.11}$$

which we employed in the massive particle action above, is firstly guaranteed to be strictly convex, i.e., for each $\lambda \in (0,1)$ we have $f_x((1-\lambda)v + \lambda w) < (1-\lambda)f_x(v) + \lambda f_x(w)$ for all v, w in the hyperbolicity cone C_x, due to the completeness of P [23], which in turn is guaranteed by the energy-distinguishing property, as we saw in the previous chapter; secondly, near the boundary of the convex set, it behaves such that for all $q \in C_x$ and $b \in \partial C_x$

$$\lim_{\lambda \to 0^+} (D_{q-b} f_x)(b + \lambda(q-b)) = 0\,, \tag{4.12}$$

which property is known as essential smoothness in convex analysis. The important point is that strict convexity and essential smoothness together ensure that the barrier function f_x induces an invertible Legendre map

$$L_x : C_x \to L_x(C_x)\,, \qquad q \mapsto -(Df_x)(q)\,, \tag{4.13}$$

where[1]

$$L_x(C_x) = \text{interior}\{v \in T_xM \mid q(v) \geq 0 \text{ for all } q \in C_x\}\,, \tag{4.14}$$

and a Legendre dual function

$$f_x^L : L_x(C_x) \to \mathbb{R}, \quad f_x^L(v) = -L_x^{-1}(v)v - f_x(L_x^{-1}(v)) \tag{4.15}$$

which can be shown, ultimately by virtue of the above conditions, to be an again strictly convex and essentially smooth function on the open convex set $L_x(C_x)$. Note that the two minus signs in (4.15) are correct, and due to our sign conventions. In fact, the inverse Legendre map is the Legendre map of the Legendre dual function f^L:

$$-Df_x^L = L_x^{-1}(v) + DL_x^{-1}(v)v + DL_x^{-1}(v)Df_x(L_x^{-1}(v)) = L_x^{-1}(v)\,. \tag{4.16}$$

In other words, the Legendre dual of the Legendre dual $(L_x(C_x), f_x^L)$ of (C_x, f_x) is again (C_x, f_x), see theorem 26.5 of [32].

The existence of this Legendre theory now enables us to eliminate the q and λ degrees of freedom, in order to obtain an equivalent particle action $I[x]$ in terms of the particle trajectory x. In the process, we will identify the definition of proper time that renders the law of free particle motion simple. Variation of the action (4.10) with respect to q yields $\dot{x} = (\lambda \deg P) L_x(q/m)$, which we know may be inverted to yield

$$q = mL_x^{-1}(\dot{x}/(\lambda \deg P))\,. \tag{4.17}$$

It is now obvious why it was convenient to encode the dispersion relation by a Lagrange multiplier term involving the barrier function (4.11); while many other ways to enforce the very

[1] Note that the here defined set $L_x(C_x)$ is equivalent to the set $C^\perp$, where the duality operation $\perp$ was defined in equation (3.29).

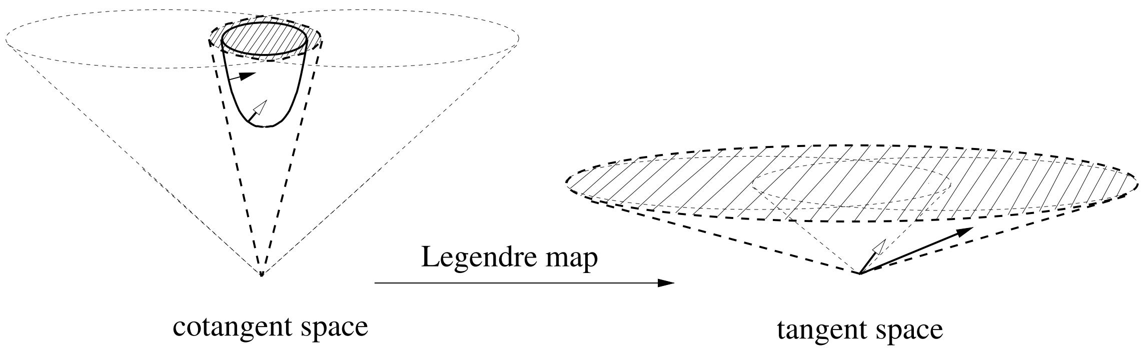

FIGURE 4.3. Mass shell and Legendre map of massive momenta to tangent space.

same dispersion relation of course do exist, the latter allows to make use of the above theory of Legendre transformations in a straightforward manner. Using the thus obtained relation and the definitions of the barrier function and the Legendre dual to eliminate q, one obtains the equivalent action

$$I[x,\lambda] = -m\deg P\int d\tau\,\lambda f^L(\dot{x}/(\lambda\deg P)) = -m\deg P\int d\tau\,\left[\lambda f^L_x(\dot{x}) + \lambda\ln(\lambda\deg P)\right]\,, \quad (4.18)$$

where for the second equality we used the easily verified scaling property $f^L(\alpha\dot{x}) = f^L(\dot{x}) - \ln\alpha$. From variation of the action (4.18) with respect to λ we then learn that

$$f^L(\dot{x}) + \ln(\lambda\deg P) + 1 = 0\,. \quad (4.19)$$

Using this twice, we have $\lambda f^L_x(\dot{x}) + \lambda\ln(\lambda\deg P) = -\lambda = -\exp(-f^L_x(\dot{x}) - 1)/\deg P$. Noting that because of $\dot{x}\in L_x(C_x)$ we also have $L^{-1}(x,\dot{x})(\dot{x}) = 1$ and thus $f^L_x(\dot{x}) = -1 - f_x(L^{-1}(\dot{x}))$, and defining the tangent bundle function

$$P^*_x : L_x(C_x)\to\mathbb{R}\,,\qquad P^*_x(v) = P_x(L^{-1}_x(v))^{-1}\,, \quad (4.20)$$

we eliminate λ in (4.18) and finally arrive at the equivalent action

$$I[x] = m\int d\tau P^*(x,\dot{x})^{1/\deg P} \quad (4.21)$$

for a free point particle of positive mass m. While the tangent bundle function P^* is generically non-polynomial, it is elementary to see that it is homogeneous of degree $\deg P$, and for later reference we also display the useful relation

$$L^{-1}_x(v) = \frac{1}{\deg P}\frac{DP^*_x(x,v)}{P^*_x(x,v)}\,. \quad (4.22)$$

The action (4.21) is reparametrization invariant, as it should be. However, parametrizations for which $P(x, L^{-1}(x,\dot{x})) = 1$ along the curve are distinguished since they yield the simple relation

$$q = mL^{-1}_x(\dot{x}) \quad (4.23)$$

between the free massive particle velocity $\dot{x}$ and the particle momentum q everywhere along the trajectory x. As usual, we choose such clocks and call the time they show proper time. Thus we have established the physical meaning of the Legendre map, and may thus justifiedly call the open convex cone $L_x(C_x)$ the cone of massive particle velocities, and the function P^* the massive dual of P, which indeed encodes the tangent bundle geometry seen by massive particles.

4.4. Lorentzian Finsler geometry and freely falling observers

It is often useful to go further and to consider freely falling non-rotating observer frames[2]. This is needed, for instance, if one wishes to determine the electric and magnetic field strengths seen by such an observer for a given electromagnetic field strength two-form F. But the definition of non-rotating frames requires to establish a meaningful parallel transport, and we will now see how the latter arises from our general constructions, which will lead to their the identification of the physical meaningful extension of Finsler geometry to the Lorentzian case. Since we saw in chapter 4 that observers are necessarily massive, their free motion is governed by an action functional

$$S[x] = \int d\tau P^*(x, \dot{x})^{1/\deg P}\,, \tag{4.24}$$

which we know to represent the trajectories of point particles of non-zero mass. Using the reparametrization invariance to set $P^*(x, \dot{x}) = 1$ along the curve, it is straightforward to derive the equations of motion

$$\ddot{x}^a + \Gamma^a(x, \dot{x}) = 0 \tag{4.25}$$

with the geodesic spray coefficients [33, 34]

$$\Gamma^a(x, v) = \frac{1}{2} g^{am}_{(x,v)} \left(\frac{\partial g_{(x,v)\,mc}}{\partial x^b} + \frac{\partial g_{(x,v)\,bm}}{\partial x^c} - \frac{\partial g_{(x,v)\,bc}}{\partial x^m} \right) v^b v^c\,. \tag{4.26}$$

These in turn are constructed from the tangent space Finsler metrics g_{e_0} [33, 34] on tangent space defined by

$$g_{(x,e_0)}(u, v) = \frac{1}{2} \left. \frac{\partial^2 P^*(x, e_0 + su + tv)^{2/\deg P}}{\partial s \partial t} \right|_{s=t=0}\,, \tag{4.27}$$

whose inverses appearing in the expression (4.26) are guaranteed to exist from the completeness of the cotangent bundle function P. Indeed, for $e_0 = L(\epsilon^0)$ with $\epsilon^0 \in C$, an explicit expression for the metric (4.27) in terms of f^L is given by

$$g_{(x,e_0)\,ab} = P_x^{*\,2/degP}(e_0) \left(-(DDf^L_x(e_0))_{ab} + 2{L_x^{-1}}_a(e_0){L_x^{-1}}_b(e_0) \right), \tag{4.28}$$

and for its inverse in terms of f by

$$g^{ab}_{(x,\epsilon^0)} = {P_x}^{2/degP}(\epsilon^0) \left(-(DDf_x(\epsilon^0))^{ab} + 2{L_x}^a(\epsilon^0){L_x}^b(\epsilon^0) \right), \tag{4.29}$$

where $(DDf_x(\epsilon^0))^{ab}(DDf^L_x(L(\epsilon^0)))_{bc} = \delta^a_c$, see [35]. Remarkably, the Finsler metric (4.28) is automatically Lorentzian as we see in the following

PROPOSITION 4.4.1. *The Finsler metric (4.29) and therefore its inverse (4.28) are of Lorentzian signature.*

[2]The precise definition of observer frames is provided in the next chapter.

Proof. Consider a cotangent frame ϵ^a with $\epsilon^\alpha(L(\epsilon^0)) = 0$ for all $\alpha = 1, \dots, \dim M - 1$, then from expression (4.29) it follows that

$$g^{ab}_{(x,\epsilon^0)}\epsilon^0_a\epsilon^0_b = {P_x}^{2/degP}(\epsilon^0) > 0, \tag{4.30}$$

$$g^{ab}_{(x,\epsilon^0)}\epsilon^0_a\epsilon^\alpha_b = 0. \tag{4.31}$$

But since any covector $\vec{p}$ on the spatial hyperplane defined by $L_x(\epsilon^0)$ can be written as $\vec{p} = p_\alpha\epsilon^\alpha$, we have

$$g^{ab}_{(x,\epsilon^0)}p_\alpha\epsilon^\alpha_a p_\beta\epsilon^\beta_b = -P_x^{2/degP}(\epsilon^0)(DDf_x(\epsilon^0))^{ab}p_\alpha\epsilon^\alpha_a p_\beta\epsilon^\beta_b < 0, \tag{4.32}$$

where the last inequality follows from the positive definiteness of the Hessian of f (see theorem 4.2 and remark 4.3 of [23]). Thus we conclude that the metric (4.29) and hence its inverse (4.28) are Lorentzian.

The metric (4.28) and its inverse (4.29) will be seen to provide a normalization for local frames which is preserved along free observer worldlines. The form of equation (4.25) indeed suggests to identify a parallel transport on the manifold M which, on the one hand, allows to recast the geodesic equation in the form of an autoparallel equation, and on the other hand, provides us with the means to define parallel transport also for purely spatial vectors. To this end, it is known to be convenient to define the derivative operators

$$\delta_i = \frac{\partial}{\partial x^i} - \Gamma^j{}_i(x,v)\frac{\partial}{\partial v^j}\,, \qquad \text{where} \quad \Gamma^i{}_j(x,v) := \frac{\partial\Gamma^i(x,v)}{\partial v^j}\,, \tag{4.33}$$

since now one can define, in full formal analogy to the Levi-Civita connection in metric geometry, the Chern-Rund connection coefficients

$$\Gamma^i{}_{jk}(u,v) = \frac{1}{2}g^{is}_{(x,v)}\left(\delta_j g_{(x,v)\,sk} + \delta_k g_{(x,v)\,js} - \delta_s g_{(x,v)\,ik}\right)\,. \tag{4.34}$$

These transform, due to the use of the δ_i operators, precisely as a linear connection would under a change of coordinates $x = x(\tilde{x})$. It is then straightforward to see that for any vector $w \in L(C)$ and vector field u on M, one may define a new vector field with components

$$(\nabla_w u)^i = w^a\partial_a u^i + \Gamma(x,w)^i{}_{jk}w^j u^k\,. \tag{4.35}$$

Clearly, ∇_w acts as a derivation on vector fields, namely $\nabla_w(u+v) = \nabla_w u + \nabla_w v$ and $\nabla_w(fu) = (wf)u + f\nabla_w u$ for any function f and vector fields u, v. Thus ∇_w may be consistently extended to act on arbitrary tensor fields S, T on M by imposing the Leibniz rule

$$\nabla_w(S \otimes T) = (\nabla_w S) \otimes T + S \otimes (\nabla_w T) \tag{4.36}$$

for arbitrary tensor fields T and S. The derivation ∇_w is not linear in its directional argument w, though, and thus amounts to what is often called a non-linear connection in the literature. Nevertheless, the non-linear covariant derivative ∇ achieves the desired reformulation of the geodesic equation (4.25) as the autoparallel equation

$$\nabla_{\dot{x}}\,\dot{x} = 0\,. \tag{4.37}$$

The non-linear connection ∇ provides sufficient structure for the discussion of freely falling non-rotating frames. The key technical observation is that for a frame field $e_0, \ldots, e_{d-1}$ that is parallely transported along the first frame vector e_0,

$$\nabla_{e_0} e_a = 0\,, \tag{4.38}$$

we have the conservation equation

$$\nabla_{e_0} (g_{e_0}(e_a, e_b)) = 0\,. \tag{4.39}$$

This means in particular that any normalization imposed on spacetime frames by virtue of the metric (4.28) is preserved along the worldline of a freely falling observer. In turn, (4.38) establishes a consistent notion of freely falling and non-rotating observer frames, and thus inertial laboratories.

In summary, we have introduced in this section the energy-distinguishing condition as the last condition on the geometry in order to provide a consistent classical spacetime structure able to carry matter field dynamics that are predictive, interpretable and quantizable. The energy-distinguishing condition also allowed to extend the theory to describe the behaviour of massive particles. Moreover, comparing the results of this chapter with those of the previous chapter, we see that there is a fundamental difference between the ways in which null covectors on the one hand, and massive covectors on the other hand, are mapped to the respective velocities on tangent space. In the null case, the Gauss maps $[DP_x]$ and $[DP^\#_x]$ associate massless particle momenta with the respective null velocities, up to an undetermined real factor. In the massive case, in contrast, the Legendre map L_x and its inverse L_x^{-1} afford the same for massive particle momenta and velocities. As a consequence, the dual geometries seen on the tangent bundle by massless and massive particles differ. For the former, the Gauss dual $P^\#$ is the relevant structure, and for the latter the Legendre dual P^*. We wish to emphasize again that while $P^\#$ is polynomial in its fibre argument, P^* generically is not. Indeed, to explicitly find the inverse Legendre map L^{-1}, and thus P^*, can be very hard in concrete applications, although its existence and uniqueness are guaranteed. Also in this sense, the tangent bundle geometry $(TM, P^\#, P^*)$ is considerably less straightforward than the cotangent bundle geometry (T^*M, P) it dualizes. This explains to some extent the difficulties noticed by Skakala and Visser in [36, 37] to identify a single Finsler-type tangent bundle geometry: generically there simply is no such geometry on *tangent space* that could give rise, dually, to a hyperbolic, time-orientable and energy-distinguishing geometry on cotangent space. The case of a Lorentzian geometry presents one notable exception.

On the positive side, on the cotangent bundle, any hyperbolic, time-orientable and energy-distinguishing reduced homogeneously polynomial geometry, provided by P, provides a perfectly fine spacetime geometry as far as point particle theory is concerned. And if one wishes to consider the coupling of fields, one needs to couple to an underlying tensorial geometry G that gives rise, by the very same field equations, to the cotangent bundle function P at hand, as discussed in chapter 2. Again, the metric case is degenerate, as we will see in chapter 6, since there one does not recognize the difference between the four different rôles played by the metric: the inverse metric plays the rôle of the fundamental spacetime structure to which fields couple, as well as the rôle of defining the (structurally very different) cotangent bundle function P, while the metric plays the rôle of both the dual $P^{\#}$ as well as the tangent bundle function P^*, which define the tangent space geometries seen by massless and massive particles, respectively. All these different structures are, strictly conceptually speaking, of course already at play in the familiar metric case, but display their different nature explicitly only in the general case.

CHAPTER 5

General properties of tensorial spacetimes

In this chapter, we will show general properties of any hyperbolic, time-orientable and energy-distinguishing spacetime (M, G). In particular, we will construct infinitesimal transformations connecting observers and show how to perform the temporal-spatial split of the dispersion relations for massless and massive point particles. Furthermore, we will show that superluminal propagation is generically allowed for any spacetime geometry giving rise to a cotangent bundle function P with $\deg P > 2$.

5.1. Observer frames and observer transformations

As agreed in chapter 3, once a time-orientation vector field T has been chosen, the convex cone $C^{\#}_x$ in tangent space containing the time-orientation T constitutes the cone of tangent vectors to observer worldlines. More precisely, if $\lambda \to x(\lambda)$ is to be an observer worldline with parameter λ, its tangent $\dot{x}(\lambda) = e_0(\lambda)$ at any point x of the worldline must be an element of the observer cone $C^{\#}_x$ with $P^*_x(e_0) = 1$. Using the restriction of the inverse of Legendre map $L_x^{-1} : C^{\#}_x \to L_x^{-1}(C^{\#}_x) \subset C_x$, we may extend $e_0(\lambda)$ to a frame bundle curve $(e_0(\lambda), e_\alpha(\lambda))$ with purely spatial frame vectors $e_\alpha(\lambda)$, i.e.,

$$L_x^{-1}\left(e_0(\lambda)\right) e_\alpha(\lambda) = 0\,. \tag{5.1}$$

The dual basis $\{\epsilon^0(\lambda), \epsilon^\alpha(\lambda)\}$ then, by definition, satisfies (for all λ)

$$\epsilon^0(e_0) = 1 \quad \text{and} \quad \epsilon^\alpha(e_0) = 0 \quad \text{for all } \alpha\,, \tag{5.2}$$

so that we find that $\epsilon^0 = L_x^{-1}(e_0)$. These equations are equivalent to

$$P_x(\epsilon^0, \dots, \epsilon^0) = 1 \tag{5.3}$$

$$P_x(\epsilon^0, \dots, \epsilon^0, \epsilon^\alpha) = 0\,. \tag{5.4}$$

Again, the first of these equations above simply expresses that the observer e_0 carries a clock that shows proper time, and the second one that the purely spatial sections seen by e_0 are those annihilated by $L_x^{-1}(e_0)$, see figure 5.1. Note that there is no distinguished way to further orthonormalize the purely spatial frame vectors amongst each other. Depending on various possible measurement prescriptions corresponding to spatial distance measurements such additional constraints on frames might of course be useful, but we will do have to do without such additional conditions here.

We now study infinitesimal transformations from one observer co-frame to another one. So we consider $\epsilon'^0 \in C_x$ satisfying the shell condition $P_x(\epsilon'_0) = m^{\deg P}$ such that $\epsilon'^0 = \epsilon^0 + \delta\epsilon^0$, where

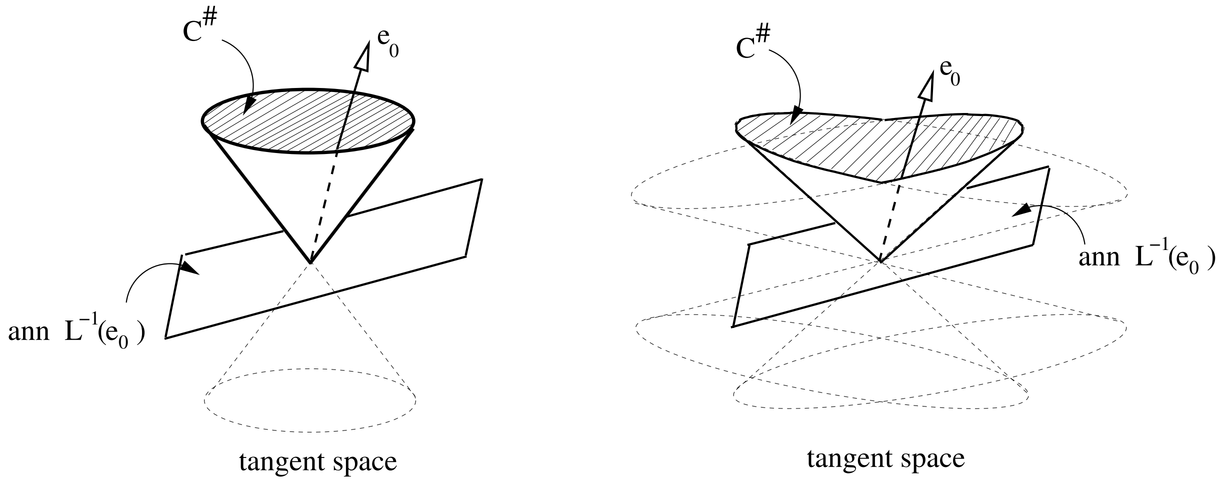

FIGURE 5.1. Purely spatial directions with respect to e_0 are those annihilated by $L_x^{-1}(e_0)$. For the Lorentzian metric case on the left, this coincides with the space of vectors g-orthogonal to e_0.

$\delta\epsilon^0$ is an infinitesimal variation of ϵ^0 leaving it on the same shell. The condition that ϵ'_0 as well as ϵ_0 are members of the same shell can also be encoded in the equation

$$f_x(\epsilon^0) = f_x(\epsilon^0 + \delta\epsilon^0)\,,$$

with f_x the hyperbolic barrier function defined in equation (4.11), because the massive dispersion relation is expressed in terms of f_x. Expanding the right hand side to first order in a neighbourhood of ϵ_0, we find of course the condition

$$L_x^a(\epsilon_0)\delta\epsilon_a^0 = 0\,, \tag{5.5}$$

where $L_x(\epsilon_0)$ is the action of the Legendre map L_x on ϵ_0. This equation is satisfied whenever

$$\delta\epsilon_a^0 = \omega_{ab}L_x^b(\epsilon^0) \quad \text{for any antisymmetric } \omega_{ab}\,, \tag{5.6}$$

so that there are $\dim M\ (\dim M - 1)/2$ parameters ω_{ab} controlling the infinitesimal transformation. These transformations are generically non-linear and they achieve to connect any two near-by covectors on a mass-shell. The infinitesimal transformation (5.6) also determines the transformation of an observer vector e_0. Indeed, if the condition $(\epsilon^0 + \delta\epsilon^0)(e_0 + \delta e_0) = 1$ is to be satisfied, it is easy to show that equation

$$\delta e_0^a = \omega_{cl} P_x^{0\cdots 0al}\, e_0^c\,, \tag{5.7}$$

with $P_x^{0\cdots 0al} = P_x(\epsilon^0, \cdots, \epsilon^0, \epsilon^a, \epsilon^l)$, must hold. However, we would also like to obtain conclusions about the transformation properties of a complete observer frame $\{e_0, e_\alpha\}$. So let us consider the infinitesimal variation $e'_\alpha = e_\alpha + \delta e_\alpha$ of the spatial vectors of an observer frame, which, by definition, must satisfy

$$(\epsilon^0 + \delta\epsilon^0)(e_\alpha + \delta e_\alpha) = 0\,.$$

This equation, by using (5.6), yields the infinitesimal transformation rule on purely spatial vectors

$$\delta e_\alpha^m = \omega_{cl} P_x^{0\cdots 0ml}\, e_\alpha^c\,. \tag{5.8}$$

Equations (5.8) and (5.7) can now be cast together in the single expression

$$e'^m_s = \left[\delta^m_k + \omega_{cl} P_x^{0\cdots 0m[l}\delta^{c]}_k\right] e^k_s \,. \tag{5.9}$$

These are now the generically non-linear infinitesimal transformations connecting two near-by observer frames. In particular, note that for $\deg P = 2$ (the metric case), the quantities $P_x^{m[l}\delta^{c]}_k$ in equation (5.9) correspond to the generators of the Lorentz group.

5.2. Temporal-spatial split of modified dispersion relations

Converting the covariant dispersion relations for massive ($m > 0$) and massless ($m = 0$) point particles,

$$P_x(p) - m^{\deg P} = 0$$

into non-covariant dispersion relations is conceptually and mathematically straightforward with the machinery we have already developed. This is so because from section 5.1, we can take an observer frame $\{e_0, e_\alpha\}$ where e_0 is an element of the observer cone $C^\#_x$ such that any spacetime momentum p can be uniquely decomposed as

$$p = E\,L^{-1}(e_0) + \vec{p} = E\,L^{-1}(e_0) + p_\alpha \epsilon^\alpha \,, \tag{5.10}$$

namely into an energy E and a purely spatial momentum $\vec{p}$ satisfying $\vec{p}(e_0) = 0$. Employing such a particular observer-dependent split, one may solve the covariant dispersion relation

$$P(x, E\,L^{-1}(e_0) + \vec{p}) = m^{\deg P} \tag{5.11}$$

for the energy E in terms of the spatial momentum $\vec{p}$, and thus obtain an observer-dependent, non-covariant dispersion relation $E = E(\vec{p})$. For the massless case $m = 0$, this equation has $\deg P$ many real physical solutions. For the massive case, the dispersion relation was introduced in the Helmholtz action in terms of the hyperbolic barrier function as $f_x(q/m) = 0$, which is defined only for q lying in one of the hyperbolicity cones of P. Thus, although equation (5.11) has $\deg P$ complex solutions for $m > 0$, only the real solutions lying in C_x and $-C_x$ are physical and only they must be considered. In any case, by considering only the physical solutions, one may then expand E in terms of the spatial momentum as

$$E(\vec{p}) = \sum_{i=0}^{\infty} c^{\alpha_1\cdots\alpha_i} p_{\alpha_1}\cdots p_{\alpha_i} \,.$$

Note that the expression $E(\vec{p})$ depends in two ways on the cotangent bundle function P: Indirectly through the temporal-spatial split (5.10) imposed by it and directly through the dispersion relation (5.11). This non-covariant version can be useful since it more directly relates to measurable quantities. However, due to Galois theory, we know that the energy will not even be an analytic expression in terms of the spatial momentum unless $\deg P \leq 4$, and not polynomial in any case. The crucial properties of the polynomial P_x being hyperbolic, time-orientable and energy-distinguishing are even more hidden in the non-covariant formulation. This is of course the key reason for having dealt exclusively with a strictly covariant treatment of dispersion relations for all formal developments throughout this work.

The conversion of a non-covariant dispersion relation into a covariant one will thus be prohibitively difficult in most cases. This is essentially due to the fact that given a relation $E = E(\vec{p})$, the reconstruction of a spacetime momentum p from E and $\vec{p}$, and indeed their physical meaning, is not directly possible without the cotangent bundle function P. We feel that this is often not considered where modified dispersion relations are proposed. Sometimes recourse to an 'anyway' underlying spacetime metric is made, but it is hard to see how this would be consistent with the stipulation of a modified dispersion relation, due to the above double rôle played by the cotangent bundle function P.

5.3. 'Superluminal' propagation of matter and vacuum Cherenkov process

In this section, we will see that massive matter can causally propagate faster than some massless particles. For simplicity, we will speak of this phenomenon as 'superluminal' massive motion. However, we will see that although particles can be superluminal, they ultimately tend to infraluminal propagation, since they can radiate massless particles by means of a Cherenkov radiation process, but only until they reach infraluminal speed. We will also study the mentioned Cherenkov process.
Superluminal propagation of massive particles on generic hyperbolic, time-orientable and energy-distinguishing spacetimes is indeed allowed since the inclusion $L_x(C_x) \supseteq C_x^{\#}$ is generically proper (i.e. equality does not hold), so that there are massive particle velocities (namely those in $L_x(C_x) \setminus \text{closure}(C_x^{\#})$) higher than some massless velocities (namely those on the boundary of $C_x^{\#}$). Since the above statements, in slightly refined form, will be of central importance in the proof of proposition 5.3.1, we will formulate them by way of the following two lemmas:

Third Lemma. For any reduced hyperbolic homogeneous cotangent bundle function P we have[1] $L_x(C_x) = \text{interior}(C_x^{\perp})$.

Proof. Since by assumption P_x is reduced, hyperbolic and homogeneous, we get from the First and the Second Lemma at the end of section 3.1 the statement: for all $p \in T_x^*M \setminus \text{closure}(C_x)$ there exists an $r \in \partial C_x$ such that $p.DP_x(r) < 0$. Since $p.DP_x(q)$ is a continuous function of q, we conclude that for all $p \in T_x^*M \setminus \text{closure}(C_x)$ there exists an $q \in C_x$ such that $p.DP_x(q) < 0$. That implies that the set $L_x(C_x)^{\perp}$ is a subset of closure$(C_x) \setminus \{0\}$. Since $L_x(C_x)$ is convex, we get $L_x(C_x) \supseteq (\text{closure}(C_x) \setminus \{0\})^{\perp} = \text{interior}(C_x^{\perp})$. Furthermore, we know that $L_x(C_x) \subseteq C_x^{\perp}$. Since $L_x(C_x)$ is open it follows that $L_x(C_x) = \text{interior}(C_x^{\perp})$.

Fourth Lemma. For any hyperbolic, time-orientable and energy-distinguishing cotangent bundle function P, we have $C_x^{\#} \subseteq \text{interior}(C_x^{\perp})$.

Proof. From section 4.2 we know that there exists a hyperbolicity cone C_x of P_x that lies completely in $(C_x^{\#})^{\perp}$. From $(C_x^{\#})^{\perp} \supseteq C_x$ and the fact that $C_x^{\#}$ is open, we conclude that

[1]The duality operation $\perp$ was defined in equation (3.29).

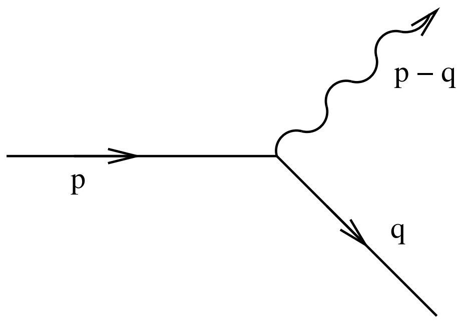

FIGURE 5.2. Vertex for the vacuum Cherenkov process. A particle of momentum p and mass m radiates off a particle of the same mass m and momentum q and a massless particle of momentum $p - q$.

$C_x^\# \subseteq \text{interior}(C_x^\perp)$.

Hence, since the slowest light is precisely the one on the boundary of $C_x^\#$, and $C_x^\# \subseteq L_x(C_x)$, it follows that if there is a particle worldline whose tangent $\dot{x}$ at the point x lies outside $C_x^\#$, but of course inside $L_x(C_x)$, this particle indeed propagates faster than the slowest light. Specializing to the familiar case of Lorentzian spacetime, one of course obtains that $L_x^{-1}(C_x^\#) = C_x$; in other words, faster-than-light propagation is simply not allowed.

We now consider a process where a positive energy massive particle of momentum p radiates off a positive energy massless particle of momentum k and a positive energy massive particle of momentum q in vacuo (see figure 5.2) such that

$$P_x(p) = P_x(q) = m^{\deg P} .$$

This process presents a vacuum Cherenkov radiation and we will prove that such a process is forbidden if and only if the ingoing momentum p lies in the stability cone

$$L_x^{-1}(C_x^\#) \,, \qquad (5.12)$$

which in turn always lies entirely within the cone C_x of massive momenta with positive energy. For the proof of these assertions, see further below; for an illustration, see figure 5.3. Specializing again to the familiar case of Lorentzian spacetime, we already know that $L_x^{-1}(C_x^\#) = C_x$ so that there is no Cherenkov radiation in vacuo.

Now we turn to the proof of the assertion that the stability cone (5.12) contains precisely the momenta of those massive particles that cannot radiate off a massless particle in vacuo. To this end, we will need to employ the first and second lemma proven in chapter 3 as well as the two lemmas proven in this section.

PROPOSITION 5.3.1. *The Cherenkov process as described above is forbidden if and only if the ingoing momentum p lies in the stability cone $L_x^{-1}(C_x^\#)$.*

Proof. First of all we get from the Third and Fourth Lemma that every observer corresponds to a massive momentum, $C_x^\# \subseteq L_x(C_x) = \text{interior}(C_x^\perp)$, so that $L_x^{-1}(C_x^\#)$ is well defined and

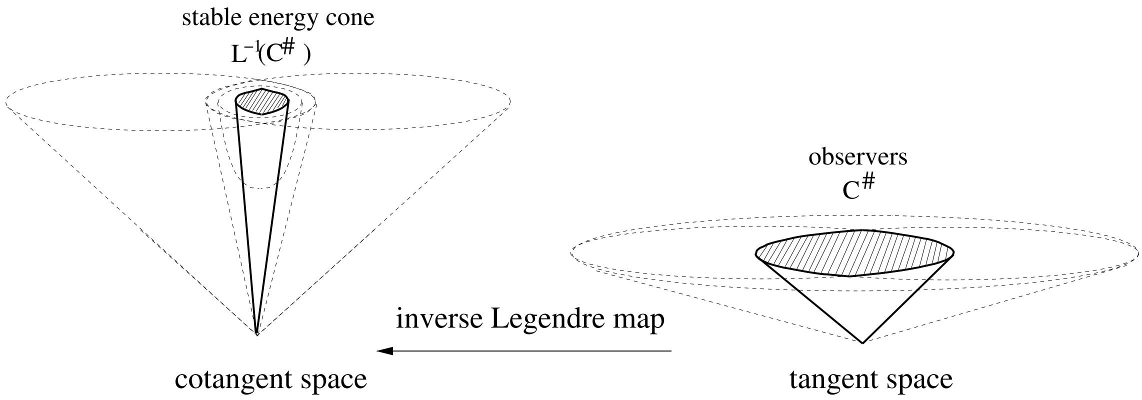

FIGURE 5.3. Stability cone: if and only if an observer can ride on a particle, the particle cannot lose energy by a vacuum Cerenkov process

always lies within C_x. It is now easy to see that a massive particle of mass m and positive energy momentum p may only radiate off a positive energy massless particle if there exists a positive energy massless momentum $r \in N_x^+$ such that $r(L_x(p)) > 0$. For consider the function

$$u(\lambda) := -\ln P_x\left(\frac{p - \lambda r}{m}\right). \tag{5.13}$$

Since for any positive λ, the covector $-\lambda r \in -(C_x^\#)^\perp$ lies in some half-space of the cotangent bundle, while $p \in C_x \subset (C_x^\#)^\perp$ lies in the corresponding other half, we conclude that for some $\lambda_0 > 1$ the line $p - \lambda r$ will necessarily intersect the boundary of C_x, so that $\lim_{\lambda \to \lambda_0} u(\lambda_0) = +\infty$. Further, from theorem 4.2 and remark 4.3 of [23], we know that for a complete hyperbolic P_x the Hessian of the barrier function $-\ln P_x$ is positive definite. Hence, we find that $u''(\lambda) > 0$ everywhere on its domain. Now first assume that the massive particle of momentum p decays into a massive particle of the same mass and of momentum $p - r$ and a massless particle of momentum r, thus respecting energy-momentum conservation. Then we have from the equality of masses for the ingoing and outgoing massive particles that $u(0) = u(1) = 0$. But because $u''(\lambda) > 0$, the only way for the analytic function u to take the same finite values at $\lambda = 0$ and $\lambda = 1$ while tending to $+\infty$ at some $\lambda_0 > 1$ is to have $0 > u'(0) = -r(L_x(p))$. Conversely, assume that $r(L_x(p)) > 0$ for some $r \in N_x^+$. Then $u'(0) < 0$ and we conclude by the mean value theorem that there must be a (because of $u''(\lambda) > 0$ unique) λ_1 with $0 < \lambda_1 < \lambda_0$ such that $u(\lambda_1) = 0$, i.e., there is an outgoing particle of the same mass such that the process occurs. In summary, a massive particle of momentum p can radiate off a positive energy massless particle if and only if there exists an $r \in N_x^+$ such that $r(L_x(p)) > 0$. Now on the one hand, we have that $p \notin L_x^{-1}(C_x^\#)$ if $r(L_x(p)) < 0$ for some $r \in N_x^+$. For then r lies certainly in $(C_x^\#)^\perp$, and thus $r(L_x(p)) > 0$ for all $p \in L_x(C_x^\#)$. On the other hand, if $p \notin L_x^{-1}(C_x^\#)$, we have $r(L_x(p)) < 0$ for some $r \in N_x^+$. This one sees essentially from the fact that $C_x^\#$ is a hyperbolicity cone of $P_x^\#$, since then for every $p \notin L_x^{-1}(C_x^\#)$ there exists some v on the boundary of $C_x^\#$ such that $DP_x^\#(v)(L_x(p)) < 0$, as is shown in the Second Lemma in chapter 3. Clearly, the image $DP_x^\#(v)$ of v under the Gauss map $DP_x^\#$ is then a massless covector, and it remains to be shown that it lies inside the positive energy cone $C_x^\#$. Since in an energy-distinguishing spacetime, a null covector is either of positive

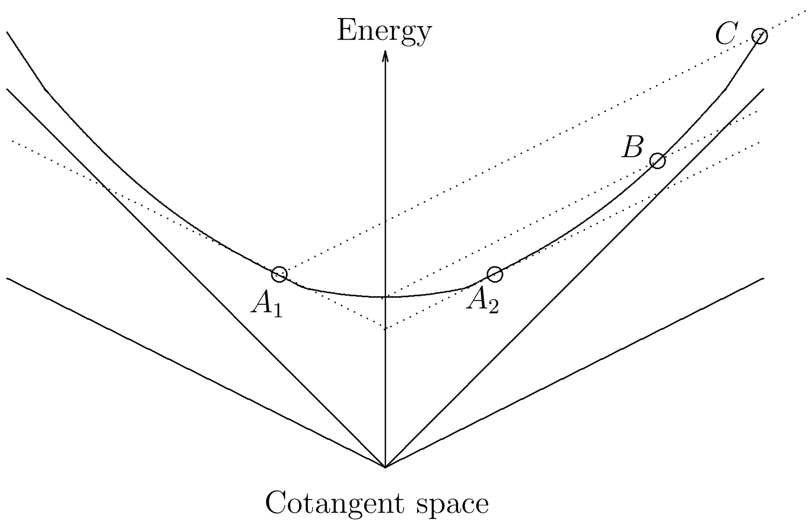

FIGURE 5.4. Slow Massless momenta (outer), fast massless momenta (inner cone), mass-shell and points defining decay regions with different properties.

or of negative energy, it suffices to find a single $y \in C_x^\#$ with $y(DP_x^\#(v)) > 0$ in order to show that $DP_x^\#(v)$ lies indeed in the positive energy cone $(C_x^\#)^\perp$. But this is easily established from the convexity of $C_x^\#$. For then we certainly find some $y \in C_x^\#$ such that $y + v \in C_x^\#$. But then $y(DP_x^\#(v)) = dP_x^\#(v + sy)/ds|_{s=0} > 0$. In summary, $p \in L_x^{-1}(C_x^\#)$ if and only if there exists an $r \in N_x^+$ with $r(L_x(p)) < 0$.

Now we illustrate how this mechanism works for the 1+1-dimensional case and dim $P = 4$, see also [14]. Extension to the physically relevant $3 + 1$-dimensional case presents only calculational, but no conceptual challenges. Consider a fourth degree hyperbolic, time-orientable and energy-distinguishing polynomial P_x in 1+1 dimensions such as the one whose massless momenta are indicated by the diagonal straight lines in figure 5.4, and which defines a mass-shell $P_x(q) - m^4 = 0$. In cotangent space, the inner cone corresponds to fast massless particles and the outer one to slow massless particles because the Gauss map sends 'inner' massless cones in cotangent space to 'outer' cones in tangent space, and vice versa. The points A_1 and A_2 are the points where the outer cone intersects the mass-shell at exactly one point. These points are important because the mass-shell region between these points defines the stable momenta of mass m and, therefore, massive covectors in this region cannot radiate off massless particles. Momenta outside the mass-shell region defined by A_1 and A_2 can radiate off a massless particle travelling at the speed of slow massless particles (defined by the outer cone).

The point C in the figure is constructed by the intersection of the slow massless cone centered at A_1 with the mass-shell. Momenta on the mass-shell lying in the region between A_2 and C can radiate at most one massless particle because the outgoing massive particle momentum will then lie between A_1 and A_2, which is the region of no decay.

The point B is constructed as the intersection of the slow massless cone centered at the massive particle momentum of lowest energy with the mass-shell. The significance of B is that if

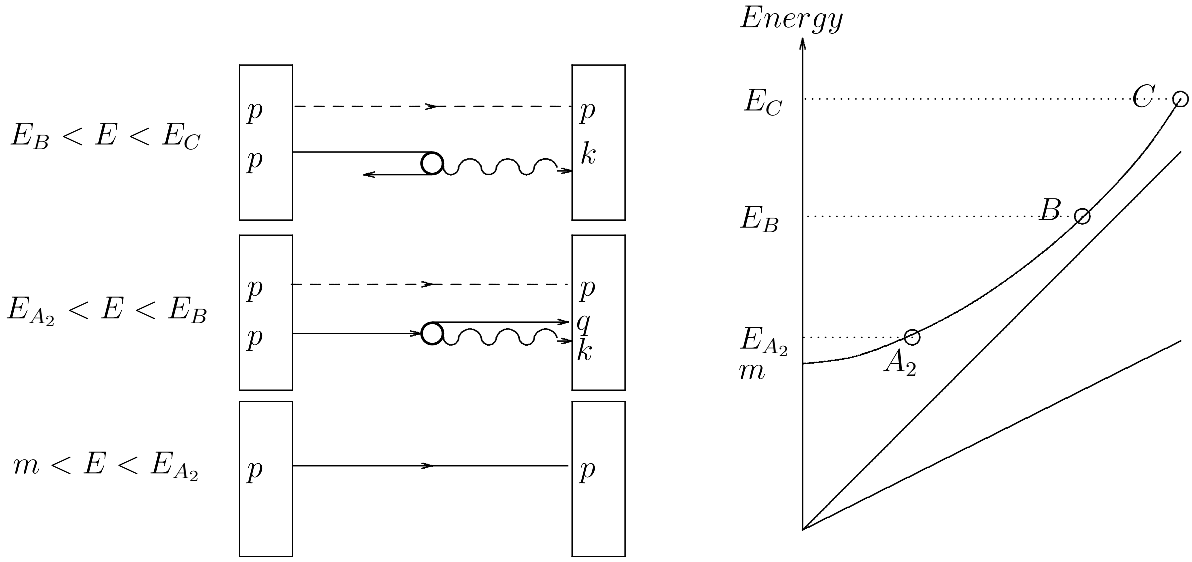

FIGURE 5.5. Detection of superluminal and infraluminal particles according to their energy.

a massive particle in the region between A_2 and B radiates off a massless particle, the outgoing massive momentum will still propagate in the same direction of the ingoing momenta. On the other hand, if a massive particle in the region between B and C radiates off a massless particle, the outgoing massive momentum will propagate in the opposite direction of the ingoing momentum.

Finally, massive particles beyond the region between A_2 and C can radiate off several massless particles until the outgoing massive momentum reaches the stability region between A_1 and A_2.

We thus have the following pattern of detection of superluminal and infraluminal particles, as shown in figure 5.5. Consider that the ingoing massive particle has momentum p. If it has energy in the range $m < E < E_{A_2}$, the particle cannot radiate because its momentum stays in the stability region. Hence, the particle is detected with the same initial momentum p. If the particle has energy in the region $E_{A_2} < E < E_B$, where particles are superluminal, the particle will be either detected with superlumial velocity and the same initial momentum p or the particle will decay in a massless particle and it will continue propagating in the initial direction. Finally, if the particle has energy in the region $E_B < E < E_C$, where particles are superluminal too, then the particle will be either detected with superlumial velocity and the same initial momentum p or the particle will decay in a massless particle and it will now propagate in a direction opposite to the initial one.

For the calculation of the decay rates of this Cherenkov processes, one needs to develop the quantum theory for particles satisfying *hyperbolic, time-orientable and energy-distinguishing* dispersion relations with $\deg P > 2$. The essential steps towards the development of such quantum theory are presented in chapters 7 and 8.

CHAPTER 6

Concrete tensorial spacetime geometries

In this chapter, we illustrate how the cotangent bundle function P is extracted from concrete field theories on a given tensorial geometry and how the conditions of P being hyperbolic, time-orientable and energy-distinguishing are used in order to restrict the geometry. We also show how these conditions can be used in order to test the viability of modified dispersion relations.

6.1. Lorentzian geometry

Purely based on our previous constructions, we study in this section how Lorentzian geometry is distinguished as the physical geometry on a metric manifold. In particular, we show how our constructions reproduce all kinematical results known from general relativity, as presented in chapter 1. So we consider a four dimensional smooth manifold M equipped with a smooth metric tensor field g (with inverse g^{-1}) of arbitrary signature encoding the geometry of the manifold M. We then want to find which metric tensors g can provide a consistent spacetime structure in the sense of giving rise to a *hyperbolic, time-orientable* and energy-distinguishing cotangent bundle function, as discussed in the previous chapters.

The field equations and the principal symbol. The first step is to consider particular matter field dynamics coupling to the geometry. For this purpose, we consider a gauge field A coupling to the metric g via the Maxwell action

$$S[A,g] = -\frac{1}{4}\int d^4x\,\sqrt{|\det(g)|}\,g^{am}g^{bn}F_{mn}F_{ab}\,, \tag{6.1}$$

where $F = dA$ is defined as the field strength and g^{ab} are the components of the inverse metric g^{-1} in a given frame. The manifold M is assumed to be orientable with canonical volume form $(\omega_g)_{abcd} =|\det(g)|^{1/2}\,\epsilon_{abcd}$. Variation of the above action with respect to the electromagnetic gauge field A together with the definition of the field strength result in the equations of motion

$$dF = 0 \quad \text{and} \quad dH = 0\,,$$

where H is defined as the field induction and is related to the field strength by

$$H_{ab} = -\frac{1}{2}\sqrt{|\det(g)|}\epsilon_{abmn}g^{mp}g^{nq}F_{pq}\,.$$

The field equations above are written in components as

$$\left(\sqrt{|\det(g)|}\right)^{-1}\epsilon^{abcd}\partial_b F_{cd} \;=\; 0\,, \tag{6.2}$$

$$\left(\sqrt{|\det(g)|}\right)^{-1}\partial_b\left(\sqrt{|\det(g)|}g^{ac}g^{bd}F_{cd}\right) \;=\; 0\,. \tag{6.3}$$

We now introduce coordinates $x^a = (t, x^\alpha)$ such that $t = 0$ provides an initial data surface Σ. Furthermore, we define the electric and the magnetic fields, E_α and B^α respectively, as

$$E_\alpha = F(\partial_t, \partial_\alpha) \quad \text{and} \quad B^\alpha = \left(\sqrt{|\det(g)|}\right)^{-1} \epsilon^{0\alpha\beta\gamma} F(\partial_\beta, \partial_\gamma)\,. \tag{6.4}$$

At this point we realize that equations (6.2) and (6.3) constitute a system of eight partial differential equations for only six field variables (E_α, B^α). However, one can check that the zero component of both equations does not contain time derivatives, such that they are not dynamical equations but constraint equations. The evolution equations are thus only provided by the spatial components of equations (6.2) and (6.3), which we can write as the first order system

$$\left(A^{bM}{}_N\, \partial_b + B^M{}_N\right) u^N = 0\,, \tag{6.5}$$

where $u^N = (E_\alpha, B^\alpha)$ and the 6×6-matrices A^b are explicitly given by

$$A^{0M}{}_N = \begin{bmatrix} g^{00}g^{\mu\nu} - g^{0\mu}g^{0\nu} & 0 \\ 0 & \delta^\mu_\nu \end{bmatrix}, \tag{6.6}$$

and

$$A^{\alpha M}{}_N = \begin{bmatrix} -2(g^{0(\mu}g^{\nu)\alpha} - g^{0\alpha}g^{\mu\nu}) & -\frac{1}{2}\sqrt{|\det g|}\epsilon_{0\nu\gamma\delta}(g^{\gamma\mu}g^{\delta\alpha} - g^{\gamma\alpha}g^{\delta\mu}) \\ (\sqrt{|\det g|})^{-1}\epsilon^{0\mu\nu\alpha} & 0 \end{bmatrix}. \tag{6.7}$$

The matrices $B^M{}_N$ depend on the metric tensor g and the volume form ω_g, but they will not be relevant for us, since they only define lower order coefficients that do not contribute to the principal polynomial constructed from equation (2.1) (see [26] for the exact dependence).

Choosing the field strengths as dynamical variables is one of the possibilities to deal with the gauge symmetry of the action (6.1) under the transformation $A \to A + d\theta$ (with θ an arbitrary differentiable function) as mentioned in section 2.1. The system (6.5) is now just a particular case of (2.1). Thus we know from chapter 2 that the principal symbol P_g of this system (from which we will extract the cotangent bundle function providing the geometry of the cotangent bundle) is proportional to the density $\tilde{P}_g(x,q) = \det(A^a q_a)$, which is found to be

$$\tilde{P}_g(x,q) = \det(A^a q_a) = (q^0)^2 (g^{ab}_x q_a q_b)^2\,. \tag{6.8}$$

We can now apply our three fundamental conditions.

Condition I: Predictivity (= hyperbolicity of P_g). From chapter 2, we know that the requirement of predictivity is translated into the condition that the found cotangent bundle density $\tilde{P}_g$ provide a hyperbolic polynomial $\tilde{P}_{g\,x}$ in each cotangent space. This is satisfied if only if each of the factors in $\tilde{P}_g(x,q)$, namely $P_1(q) = q^0$ and $P_2(q) = g^{ab}_x q_a q_b$, are hyperbolic and the corresponding hyperbolicity cones of each factor have non-zero intersection with each other (see equation (2.9)). The factor $P_1(q) = q_0$ is clearly hyperbolic with respect to any covector h with $h_0 \neq 0$ (in our chosen basis). This is so because (following the definition of hyperbolicity in equation (2.4)) the equation $P_1(q + \lambda h) = q_0 + \lambda h_0 = 0$ has only the single real root $\lambda = -q_0/h_0$. About the factor $P_2(q) = g^{ab}_x q_a q_b$ we have the following

PROPOSITION 6.1.1. *The metric polynomial $P_2(q) = g_x^{ab} q_a q_b$ is hyperbolic if and only if the associated metric g is of Lorentzian signature.*

Proof. The proof that the polynomial $P_2(q) = g_x^{-1}(q,q)$ is hyperbolic if the metric is Lorentzian was already given as an example of a hyperbolic polynomial in section 2.2. So here we only prove that if P_2 is hyperbolic, the metric is Lorentzian. For this purpose we consider P_2 being hyperbolic with respect to some covector h such that $P_2(h) > 0$. Hence, the equation $P_2(q+\lambda h) = \lambda^2 g_x^{-1}(h,h) + 2\lambda g_x^{-1}(h,q) + g_x^{-1}(q,q) = 0$ has only real roots. But then the discriminant of this equation is positive, i.e.,

$$(g_x^{-1}(h,q))^2 - g_x^{-1}(h,h) g_x^{-1}(q,q) > 0 \,.$$

We now take a covector basis $\{\epsilon^0, \epsilon^\alpha\}$ with $\epsilon^0 = h$ and such that $g_x^{-1}(\epsilon^0, \epsilon^\alpha) = 0$. Thus $g_x^{-1}(\epsilon^0, \epsilon^0) > 0$ and the above discriminant condition is written as $q_\alpha q_\beta g_x^{-1}(\epsilon^\alpha, \epsilon^\beta) < 0$ for all q_α, q_β, which already proves that g_x^{-1} must be of Lorentzian signature.

Hence, we conclude that condition I, predictivity, already restricts metric manifolds to those of Lorentzian type for Maxwell theory to be predictive.

Description of the hyperbolicity cones. We now use theorem 2.2.2 (Routh-Hurwitz theorem) in order to characterize the hyperbolicity cones of $\tilde{P}_g(x,q) = (q^0)^2 (g_x^{ab} q_a q_b)^2$. For the factor $P_1(q)$ we consider a hyperbolic covector h with $P_1(h) > 0$ so that one trivially obtains the hyperbolicity cone

$$C(P_1, h) = \{q \in T_x M \mid q_0 > 0\} \,,$$

which is a half space. For the second factor $P_2(q) = g_x^{ab} q_a q_b$ with metric of signature $(1,-1,-1,-1)$, we consider a hyperbolic covector h such that $g_x^{-1}(h,h) > 0$. Its corresponding hyperbolicity cone is therefore

$$C(P_2, h) = \{q \in T_x M \mid g^{-1}(h,q) > 0 \text{ and } g^{-1}(q,q) > 0\} \,.$$

Hence, the hyperbolicity cone $C(\tilde{P}, h)$ of $\tilde{P}$ is obtained as the intersection of both cones as

$$C(\tilde{P}, h) = C(P_1, h) \cap C(P_2, h) = C(P_2, h) \,.$$

The hyperbolic covectors are therefore only covectors hyperbolic with respect to the metric polynomial.

The geometric optical limit. Following the developments of section 3.1, a condition to construct an asymptotic solution of the system (6.5) to any order is to solve the eikonal equation

$$(\partial_0 S)^2 (g^{ab}(x)\, \partial_a S \partial_b S)^2 = 0 \,.$$

This equation is solved if the eikonal function S satisfies either $\partial_0 S = 0$ or $g_x^{ab} \partial_a S \partial_b S = 0$. But we notice that the solution of $\partial_0 S = 0$ is non-physical because it corresponds to a non-propagating solution. This is the case because in the derivation of the eikonal equation, we assumed that the solution has a wave-like factor e^{iS}, so that the solution of $\partial_0 S = 0$ corresponds to a phase without time dependence and, hence, it does not propagate. This solution is therefore non-physical and must be excluded from the description. We thus recognize

$$g^{ab}(x)\partial_a S \partial_b S = 0$$

as the physical eikonal equation whose solutions represent truly field propagation. This equation already has the form of a reduced polynomial. Hence, we finally identify our covariant cotangent bundle function

$$P_{x\,g}(q) = g_x^{ab} q_a q_b$$

providing the geometry of the cotangent bundle, which is as usual encoded in the Lorentzian metric g. Recall that the condition $P_{x\,g}(q) = g_x^{ab} q_a q_b = 0$ defines the massless dispersion relation with the set of massless momenta N_x at a given point x of M given by

$$N_x = \{q \in T_x^* M \mid g_x^{ab} q_a q_b = 0\}\,,$$

which in this case coincides with N_x^{smooth}, and which we recognize as the standard massless momenta set in Lorentzian spacetime.

Condition II: Interpretability (= time-orientability of $P_{x\,g}$). In order to satisfy the interpretability condition, we need the cotangent bundle function $P_{x\,g}$ to be hyperbolic and time-orientable. So we need to compute the dual polynomial of $P_{g\,x}$. In this case it is easy to guess a dual, namely $P_g^\#(x,v) = g_x(v,v)$. Indeed, $P_{g\,x}^\#(DP_{g\,x}(q)) = 4g_x(g_x^{-1}(q,\cdot), g_x^{-1}(q,\cdot)) = 4g_x^{-1}(q,q) = 4P_{g\,x}(q)$, so that the equation (3.17) defining the dual polynomial is satisfied. We thus take $P_{g\,x}^\#(x,v) = g_x(v,v)$ as the dual polynomial of $P_{g\,x}(q)$, which is hyperbolic because g is of Lorentzian signature (since g^{-1} is of Lorentzian signature). Thus the hyperbolicity and time-orientability conditions are satisfied. But then picking a hyperbolic vector field e_0 with respect to $P_g^\#$ corresponds to picking a time-orientation on M. The hyperbolicity cone $C^\# = C(P_g^\#, e_0)$ is then defined as the cone of future observers. Moreover, the massless point particle action governing the kinematics of light rays is, according to equation (3.27),

$$S[x,\mu] = \int d\tau\, \mu\, g_x(\dot{x},\dot{x})\,,$$

which reproduces the well-known massless point particle action known from general relativity (see chapter 1).

Condition III: Quantizability (= energy-distinguishability of $P_{x\,g}$). In this case, the hyperbolicity and time-orientability of P_g each separately already imply the energy-distinguishing property. For from the explicit definition of $C^\#$ we know that at every point x and for every vector $X \in C_x^\#$ the covector $g_x(X,\cdot) \in (C_x^\#)^\perp$. Arranging for $P_{g\,x}^\#(C_x^\#) > 0$ and knowing that P_g is hyperbolic, it is easy to show that $g_x(\omega, v) > 0$ for every vector $\omega \in \partial C_x^\#$ and $v \in C_x^\#$, which shows that for every vector $\omega \in \partial C_x^\#$ the covector $g_x(\omega,\cdot) \in (C_x^\#)^\perp$ and $g_x(-\omega,\cdot) \in -(C_x^\#)^\perp$. More precisely $g_x(\partial C_x^\#,\cdot) \in (C_x^\#)^\perp$ and $g_x(-\partial C_x^\#,\cdot) \in -(C_x^\#)^\perp$. But $g_x(\partial C_x^\#,\cdot) \in (C_x^\#)^\perp$ is the image of the dual Gauss map induced from ${P_{g\,x}}^\#$ when applied to $\partial C_x^\#$. Thus, we conclude that

Lorentzian metric geometry is a hyperbolic, time-orientable and energy-distinguishing geometry, which therefore can be used, as we know, to provide a classical spacetime structure.

The massive dispersion relation and observer frames. From chapter 4, we now identify the massive dispersion relation as

$$g_x^{-1}(q,q) - m^2 = 0\,.$$

The hyperbolic barrier function is thus given by $f(q) = -1/2\log g_x^{-1}(q,q)$ which defines the Legendre map $L(q) = g_x^{-1}(q,\cdot)/g_x^{-1}(q,q)$. The inverse of this Legendre map is easy to guess, namely $L(v) = g_x(v,\cdot)/g_x(v,v)$. We thus identify the tangent bundle function

$$P^*(x,v) = g_x(v,v)$$

as the geometry seen by massive particles in tangent space on a Lorentzian spacetime. This gives rise, according to equation (4.21), to the well-known action

$$S[x] = \int d\tau\, m\, \sqrt{g_x(\dot{x},\dot{x})}$$

for massive particles on Lorentzian spacetime. Following now the definitions of section 5.1, observer frames $\{e_0, e_\alpha\}$ must satisfy

$$g_x(e_0,e_0) = 1 \quad \text{and} \quad g_x(e_0,e_\alpha) = 0\,.$$

For the metric case, one can additionally normalize the spatial vectors e_α as $g_x(e_\alpha, e_\beta) = -\delta_{\alpha\beta}$. Thus, one has $g_x(e_a, e_b) = \eta_{ab}$ where η is the Minkowski metric. This is of course the prescription for defining $g-$orthonormal observer frames in a Lorentzian spacetime .

Remarks. In the metric case, as we have seen, one does not recognize the difference between the four different rôles played by the metric: the inverse metric plays the rôle of the fundamental spacetime structure G to which fields couple, as well as the rôle of defining the (structurally very different) cotangent bundle function P, while the metric plays the rôle of both the dual $P^\#$ as well as the tangent bundle function P^*, which define the tangent space geometries seen by massless and massive particles, respectively. All these different structures are, strictly conceptually speaking, of course already at play in the familiar metric case, but display their different nature explicitly only in the framework of the general theory presented in this work. Moreover, we have seen that all well-known kinematical results from Lorentzian spacetime were reproduced by purely following the general constructions of this work. This is simply what Einstein told us: time-orientable Lorentzian manifolds provide a viable classical spacetime structure.

6.2. Area metric geometry

We now consider the non-trivial case of area metric spacetimes where the conceptually different structures G, P, $P^\#$ and P^* manifest itself in the actual expressions for these quantities, other than in the familiar metric case discussed above.

We first provide a short review of [18] about area metric manifolds and the classification of area metrics. The reader is referred to that work and to [38, 39] for further details on area metric manifolds.

An are metric is a fourth-rank covariant tensor field G, whose components $G_{abcd} = G(e_a, e_b, e_c, e_d)$ in any given basis $\{e_a\}$ of T_xM features the symmetry conditions

$$G_{abcd} = G_{badc} = -G_{bacd}\,,$$

and an invertibility condition in the sense that there is an inverse area metric with components $(G^{-1})^{abmn}$ such that

$$(G^{-1})^{abmn}G_{mncd} = 4\,\delta^{[a}_{c}\delta^{b]}_{d}\,.$$

Due to the symmetries of an area metric, the indices of G may be combined into antisymmetric Petrov pairs $[ab]$ such that G can be represented by a symmetric square matrix of dimension $D = \dim M(\dim M - 1)/2$. For instance, in four dimensions, which is the case we will consider here, we have index pairs $[01], [02], [03], [23], [31]$ and $[12]$ with the corresponding Petrov indices $A = 1, \ldots, 6$. We can then arrange the independent components of an area metric G in four dimensions as the 6×6 Petrov matrix

$$Petrov(G) = \begin{bmatrix} G_{0101} & G_{0102} & G_{0103} & G_{0123} & G_{0131} & G_{0112} \\ & G_{0202} & G_{0203} & G_{0223} & G_{0231} & G_{0212} \\ \ddots & & G_{0303} & G_{0323} & G_{0331} & G_{0312} \\ & \ddots & & G_{2323} & G_{2331} & G_{2312} \\ & & \ddots & & G_{3131} & G_{3112} \\ & & & \ddots & & G_{1212} \end{bmatrix}. \tag{6.9}$$

An area metric manifold (M, G) carries a canonical volume form ω_G, defined by

$$(\omega_G)_{a_1 \cdots a_{\dim M}} = f(G)\,\epsilon_{a_1 \cdots a_{\dim M}}\,, \tag{6.10}$$

where $f(G) = |\det(\mathrm{Petrov}(G))|^{1/6}$ and ϵ is the Levi-Civita tensor density normalized such that $\epsilon_{01 \cdots M-1} = 1$.
A generic area metric contains more algebraic degrees of freedom than a metric, starting from dimension four. This can be seen by counting the independent components of the symmetric $D \times D$ Petrov matrix representing the area metric, which amounts to $D(D + 1)/2$ independent real numbers. The invertibility requirement does not further reduce this number since it is an open condition. Thus, for instance, area metrics in dimensions 2, 3, 4 and 5 have 1, 6, 21 and 55 independent components, respectively.

In metric geometry, Sylvester's theorem allows to get good technical control over all metrics, classifying them by their signature and giving normal forms that can be obtained by $GL(4)$ transformations. Four-dimensional area metrics were classified in [18] into $GL(4)$-equivalent algebraic classes. More precisely, the area metrics G and H belong to the same class if there is a $GL(4)$ transformation t such that

$$G^{abcd} = t^a{}_m t^b{}_n t^c{}_p t^d{}_q H^{mnpq} \,.$$

The obtained classes are labelled by Segré types according to the eigenvalues of the endomorphism (in the space of two forms)

$$J_{cd}{}^{ab} = G^{abmn}(\omega_g)_{mncd}$$

as follows: the Segré type of J encodes the size of the Jordan blocks of J, and whether the corresponding eigenvalues of the corresponding blocks are real or complex. Thus a Segré type is represented as $[A\bar{A}\ldots BCD]$, with A, B, C, D positive integers. An integer labelled by A and followed by $\bar{A}$ means that the endomorphism J contains a Jordan block of size A with a complex eigenvalue of J and a Jordan block of the same size with the complex conjugate eigenvalue of J. Otherwise the endomorphism contains a real Jordan block of sizes B, C and D. For instance, the metaclass $[1\bar{1}\,1\bar{1}\,1\bar{1}]$ means that the corresponding endomorphism J has a six distinct complex eigenvalues where three of them are simply the complex conjugates of the other three. The resulting classification is provided in theorem 4.3 of [18]. The result is that area metrics in four dimensions are classified into

- three metaclasses where the Jordan blocks of the corresponding endomorphism J only have complex eigenvalues $\sigma_i \pm i\tau_i$, namely

metaclass I $[1\bar{1}\ 1\bar{1}\ 1\bar{1}]$

$$\begin{bmatrix} -\tau_1 & 0 & 0 & \sigma_1 & 0 & 0 \\ 0 & -\tau_3 & 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & -\tau_2 & 0 & 0 & \sigma_2 \\ \sigma_1 & 0 & 0 & \tau_2 & 0 & 0 \\ 0 & \sigma_3 & 0 & 0 & \tau_3 & 0 \\ 0 & 0 & \sigma_2 & 0 & 0 & \tau_1 \end{bmatrix}, \qquad (6.11)$$

metaclass II $[2\bar{2}\ 1\bar{1}]$

$$\begin{bmatrix} 0 & 0 & 0 & \sigma_1 & -\tau_1 & 0 \\ 0 & 0 & 0 & \tau_1 & \sigma_1 & 0 \\ 0 & 0 & -\tau_2 & 0 & 0 & \sigma_2 \\ \sigma_1 & \tau_1 & 0 & \tau_2 & 0 & 0 \\ -\tau_1 & \sigma_1 & 0 & 0 & 0 & 1 \\ 0 & 0 & \sigma_2 & 0 & 1 & 0 \end{bmatrix},$$

metaclass III $[3\bar{3}]$

$$\begin{bmatrix} 0 & 0 & 0 & \sigma_1 & -\tau_1 & 0 \\ 0 & 0 & 0 & \tau_1 & \sigma_1 & 0 \\ 0 & 0 & -\tau_1 & 1 & 0 & \sigma_1 \\ \sigma_1 & \tau_1 & 1 & \tau_1 & 1 & 0 \\ -\tau_1 & \sigma_1 & 0 & 1 & 0 & 0 \\ 0 & 0 & \sigma_1 & 0 & 0 & 0 \end{bmatrix},$$

- four metaclasses with real Jordan blocks in J of at most size one

metaclass IV $[1\bar{1}\ 1\bar{1}\ 11]$

$$\begin{bmatrix} -\tau_1 & 0 & 0 & \sigma_1 & 0 & 0 \\ 0 & -\tau_2 & 0 & 0 & \sigma_2 & 0 \\ 0 & 0 & \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \sigma_1 & 0 & \lambda_2 & \lambda_1 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & \tau_2 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & \tau_1 \end{bmatrix},$$

metaclass V $[2\bar{2}\ 11]$

$$\begin{bmatrix} 0 & 0 & 0 & \sigma_1 & -\tau_1 & 0 \\ 0 & 0 & 0 & \tau_1 & \sigma_1 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & \lambda_2 \\ \sigma_1 & \tau_1 & 0 & \lambda_1 & 0 & 0 \\ -\tau_1 & \sigma_1 & 0 & 0 & 0 & 1 \\ 0 & 0 & \lambda_2 & 0 & 1 & 0 \end{bmatrix},$$

metaclass VI $[1\bar{1}\ 11\ 11]$ metaclass VII $[11\ 11\ 11]$

$$\begin{bmatrix} -\tau_1 & 0 & 0 & \sigma_1 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & \lambda_2 \\ \sigma_1 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & \tau_1 \end{bmatrix}, \qquad \begin{bmatrix} \lambda_5 & 0 & 0 & \lambda_6 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & \lambda_2 \\ \lambda_6 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & \lambda_4 & 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & \lambda_6 & 0 & 0 & \lambda_5 \end{bmatrix},$$

- 16 metaclasses with at least one real Jordan block in J of size greater or equal two.

The explicit expression for the last 16 metaclasses is not included here because, as we will see in lemma 6.2.1, they cannot provide a hyperbolic, time-orientable and energy-distinguishing cotangent bundle function P for area metric Maxwell electrodynamics.

Remark. Four-dimensional area metrics that are induced by a Lorentzian metric automatically lie in the first metaclass $[1\bar{1}1\bar{1}1\bar{1}]$. This corresponds to the values $\tau_1 = \tau_2 = \tau_3 = 1$ and $\lambda_1 = \lambda_2 = \lambda_3 = 0$ in equation (6.11). Moreover, the continuous dependence of the eigenvalues of an endomorphism on the components of a representing matrix implies that any area metric in the neighbourhood of such a metric-induced area metric is equally of class $[1\bar{1}1\bar{1}1\bar{1}]$. Thus area metrics of immediate phenomenological relevance are clearly those of class I.

Equipped with this knowledge on area metric manifolds and the classification of area metrics, we can proceed with our study, precisely as we did for metric manifolds, of which area metric manifolds can provide a classical spacetime structure.

The field equations and the principal symbol. The first step is to consider a matter field coupling to the geometry, in this case the area metric G. For this purpose we consider a one-form gauge field A coupling to the area metric G according to the action

$$S[A,G] = -\frac{1}{8}\int dx^4 f(G)\,\Big[F_{ab}F_{cd}\,G^{abcd}\Big]\,, \tag{6.12}$$

where $F = dA$ is defined as the field strength and G^{abcd} are the components of the inverse area metric G^{-1} in a given coordinate system, and f is the density used to define the volume form in (6.10). In fact, this is the most general action for a one-form gauge potential that results in a linear constitutive law [40, 41]. Variation of the above action together with the definition of the field strengths results in the equations of motion

$$dF = 0 \quad \text{and} \quad dH = 0\,,$$

where H is defined as the field induction and is related to the field strength by

$$H_{ab} = -\frac{1}{4} f(G)\,\epsilon_{abmn} G^{mnpq} F_{pq}\,.$$

These field equations are written in components as

$$f^{-1}(G)\epsilon^{abcd}\partial_b F_{cd} = 0 \tag{6.13}$$

$$f^{-1}(G)\partial_b\Big(f(G)g^{abcd}F_{cd}\Big) = 0\,. \tag{6.14}$$

We now introduce coordinates $x^a = (t, x^\alpha)$ such that $t = 0$ provides an initial data surface Σ, and define the electric and the magnetic fields, E_α and B^α respectively, as

$$E_\alpha = F(\partial_t, \partial_\alpha) \quad \text{and} \quad B^\alpha = f^{-1}(G)\epsilon^{0\alpha\beta\gamma}F(\partial_\beta, \partial_\gamma)\,. \tag{6.15}$$

Equations (6.13) and (6.14) constitute a system of eight partial differential equations for only six field variables (E_α, B^α). But one can check, as in the metric case, that the zero component of both equations does not contain time derivatives and they are therefore constraint equations. The evolution equations are therefore only provided by the spatial components of (6.13) and (6.14), which we can write precisely as the first order system (6.5) with $u^N = (E_\alpha, B^\alpha)$ and the 6×6 dimensional matrices A^b now given by

$$A^{0M}{}_N = \begin{bmatrix} G^{0\mu0\nu} & 0 \\ 0 & \delta^\mu_\nu \end{bmatrix}, \tag{6.16}$$

and

$$A^{\alpha M}{}_N = \begin{bmatrix} -2\,G^{0(\mu\nu)\alpha} & -\frac{1}{2}f(G)\epsilon_{0\nu\gamma\delta}\,G^{\gamma\delta\mu\alpha} \\ f(G)^{-1}\epsilon^{0\mu\nu\alpha} & 0 \end{bmatrix}. \tag{6.17}$$

The matrices $B^M{}_N$ depend on the area metric tensor G and the volume form ω_G, but sence they present only lower order coefficients, they are not be relevant for us .

By choosing again the field strengths as dynamical variables one does away with the gauge symmetry of the action. As we know, the system (6.5) is a particular case of (2.1), so that from chapter 2 the principal symbol $\tilde{P}_G$ of this system is proportional to the determinant of the matrix $A^a q_a$, which in this case is

$$A^a q_a = \begin{bmatrix} G^{0\mu0\nu}p_0 - 2G^{0(\mu\nu)\alpha}p_\alpha & -\frac{1}{2}\epsilon_{0\nu\gamma\delta}G^{\gamma\delta\mu\alpha}p_\alpha \\ \epsilon^{\mu\nu\alpha}p_\alpha & \delta^\nu_\mu p_0 \end{bmatrix}. \tag{6.18}$$

Before computing the determinant, we decompose the area metric G as

$$\text{Petrov}(G)^{[ab][cd]} = \begin{bmatrix} M & K \\ K^T & N \end{bmatrix}, \tag{6.19}$$

with antisymmetric index pairs $[01], [02], [03], [23], [31], [12]$. The matrices M, K, N are 3×3 matrices related to the area metric through

$$\begin{aligned} M^{\alpha\beta} &= G^{0\alpha0\beta}\,, \\ K^\alpha_\beta &= \frac{1}{2}\epsilon_{0\beta\mu\nu}G^{0\alpha\mu\nu}\,, \\ N_{\alpha\beta} &= \frac{1}{4}\epsilon_{0\alpha\mu\nu}\epsilon_{0\beta\sigma\delta}G^{\mu\nu\sigma\delta}, \end{aligned} \tag{6.20}$$

Using now that for any $n\times n$ matrices A, B, C, D

$$\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD - BC) \quad \text{if} \quad CD = DC, \tag{6.21}$$

we can write the determinant of (6.18) as the determinant of a 3×3 matrix as

$$f^2(G)\det(A^a q_a) = \det\Big(G^{0\mu0\nu}q_0^2 - 2G^{0(\mu\nu)\alpha}q_\alpha q_0 + G^{\mu\alpha\nu\gamma}q_\alpha q_\gamma\Big) = -q_0^2 P_{G\,x}(q)\,, \tag{6.22}$$

where $P_{G\,x}(q)$ on the right hand side of the equation above is known as the Fresnel polynomial and is cast into the covariant expression

$$P_{G\,x}(q) = -\frac{1}{24}(\omega_{G_x})_{mnpq}(\omega_{G_x})_{rstu}G_x^{mnr(a}G_x^{b|ps|c}G_x^{d)qtu}q_aq_bq_cq_d\,. \tag{6.23}$$

Notice that the factor $f^2(G)$ in front of the determinant in equation (6.22) was precisely introduced for $P_{G\,x}$ to be covariant. Writing now $P_{G\,x}$ in terms of the constitutive matrices M, K, N we find

$$\frac{P_{G\,x}(p_0,\vec{p})}{f^2(G)} = a\,p_0^4 + b(\vec{p})\,p_0^3 + c(\vec{p})\,p_0^2 + d(\vec{p})\,p_0 + e(\vec{p}) \tag{6.24}$$

with coefficients

$$\begin{aligned}
a &= -\det(M^{\alpha\beta})\,,\\
b(\vec{p}) &= \epsilon_{0\alpha\beta\gamma}\,G_M^{\alpha\gamma\sigma\mu}\,K^\beta_\sigma p_\mu\,,\\
c(\vec{p}) &= -\left(N_{\rho\sigma}\,G_M^{\kappa\rho\mu\sigma} + \epsilon_{0\alpha\beta\gamma}\epsilon^{0\tau\mu\rho}K^\beta_\tau K^\alpha_\rho M^{\gamma\kappa} + 2K^\epsilon_\alpha K^{[\mu}_\epsilon M^{\alpha]\kappa} + 2K^\kappa_\alpha K^{[\alpha}_\epsilon M^{\mu]\epsilon}\right)p_\kappa p_\mu\,,\\
d(\vec{p}) &= -2\epsilon^{\rho\epsilon\nu}\left(2N_{\epsilon\alpha}K^{[\mu}_\rho M^{\alpha]\sigma} - K^\mu_\rho K^\sigma_\alpha K^\alpha_\epsilon\right)p_\sigma p_\mu p_\nu\,,\\
e(\vec{p}) &= -N_{\beta\kappa}\epsilon^{\beta\mu\rho}\epsilon^{\kappa\nu\gamma}\left(\frac{1}{2}M^{\lambda\upsilon}N_{\rho\gamma} - K^\lambda_\rho K^\upsilon_\gamma\right)p_\lambda p_\mu p_\upsilon p_\nu,
\end{aligned}$$

where $G_M^{\alpha\mu\beta\nu} = M^{\alpha\beta}M^{\mu\nu} - M^{\alpha\nu}M^{\mu\beta}$.

Summarizing, we found that the tangent bundle function providing the geometry of the cotangent bundle on an area metric manifold must be extracted from

$$\tilde{P}_G(x,q) = (q^0)^2 P_{G\,x}(q)\,, \tag{6.25}$$

where $P_{G\,x}$ is the Fresnel polynomial given in expression (6.23). Before we proceed in testing our three fundamental conditions on an area metric manifold, we provide a short aside on the application of area metric geometry to materials, to which we will refer later.

Application to materials (review of [18] *and see also* [40, 41]*).* For Maxwell electrodynamics in linear optical materials, the constitutive relations between the electromagnetic field strength vectors E, B and the electromagnetic induction vector densities D, H are provided by the equations

$$\begin{aligned}
D &= \varepsilon E + \alpha B\\
H &= \beta E + \mu^{-1}B\,.
\end{aligned} \tag{6.26}$$

The matrices ε and μ encode the permittivity and permeability of the material, while the matrices α and β produce magnetoelectric effects. Using the frame $\{e_0, e_\alpha\}$ of the observer measuring these constitutive relations, we now define (in this observer frame) the two form F_{ab}, the bi-vector density $H^{[ab]}$ and the tensor density $\chi^{[ab][cd]}$ such that

$$Petrov(F) = \begin{bmatrix} E\\ B\end{bmatrix}\,,\qquad Petrov(H) = \begin{bmatrix} D\\ H\end{bmatrix}\,,\qquad \text{and}\qquad Petrov(\chi) = \begin{bmatrix} \varepsilon & \alpha\\ \beta & \mu^{-1}\end{bmatrix}\,. \tag{6.27}$$

With these definitions, the constitutive relations (6.26) can be encoded in the single equation

$$H^{[ab]} = \frac{1}{2}\chi^{abcd}F_{cd}\,.$$

The tensor density $\chi^{[ab][cd]}$ has 35 independent components in general. However, if in the constitutive relations we have $\beta = \alpha^T$, the tensor density $\chi^{[ab][cd]}$ has precisely the symmetries of an area metric and the equations of motion for the electric and magnetic fields are obtained from the action $\int d^4x\,\mathcal{L}$ with Lagrangian density

$$\mathcal{L} = -\frac{1}{8}\chi^{abcd}F_{ab}F_{cd}\,. \tag{6.28}$$

The case where χ has the symmetries of an area metric therefore corresponds to non-dissipative linear electrodynamics. This is so because the equations of motion are then derived from an action principle, which leads to an equation for energy-momentum conservation. Moreover, for many materials one finds $\alpha = \beta = 0$.

Finally, comparing the Lagrangian (6.28) with that of electrodynamics on area metric backgrounds (6.12) one identifies

$$\chi^{abcd} = |\det(Petrov(G))|^{1/6}G^{abcd}\,.$$

We thus conclude that all non-dissipative linear optical media are described in terms of area metric backgrounds, and the dynamics for the electric and magnetic fields are those of area metric electrodynamics.

So we now return to our study of which area metric manifolds can provide a consistent classical spacetime structure by investigating our three fundamental conditions on the cotangent bundle function P.

Condition I: Predictivity (= hyperbolicity of P_G). To ensure that this condition is satisfied, we would have to identify the area metrics for which the polynomial $\tilde{P}_G$ in equation (6.25) is hyperbolic. We know that the factor q_0 is hyperbolic, so we only have to check for the hyperbolicity of the Fresnel polynomial P_G. Strictly speaking, we would have to check for which of the 23 area metric metaclasses there exists a hyperbolic covector with respect to P_G. But due to the complexity of the problem (because the metaclasses contain arbitrary scalars which could turn the roots of equation $P_G = 0$ from real to complex), we will have to wait until the application of conditions II and III in order to restrict the metaclasses of area metrics to those that could provide a classical spacetime structure. See [42] for the identification of some conditions for Maxwell's equations with a general linear local constitutive law to be hyperbolic.

The geometric optical limit. In this case, the eikonal equation is given by

$$(\partial_0 S)^2 P_G(x, \partial S) = 0$$

with real eikonal function S. We know that by solving this equation we can construct an asymptotic solution of the system to any desired order. This eikonal equation is clearly solved if $(\partial_0 S)^2 = 0$ or $P_G(x, \partial S) = 0$. But for the same arguments we used in the discussion of the

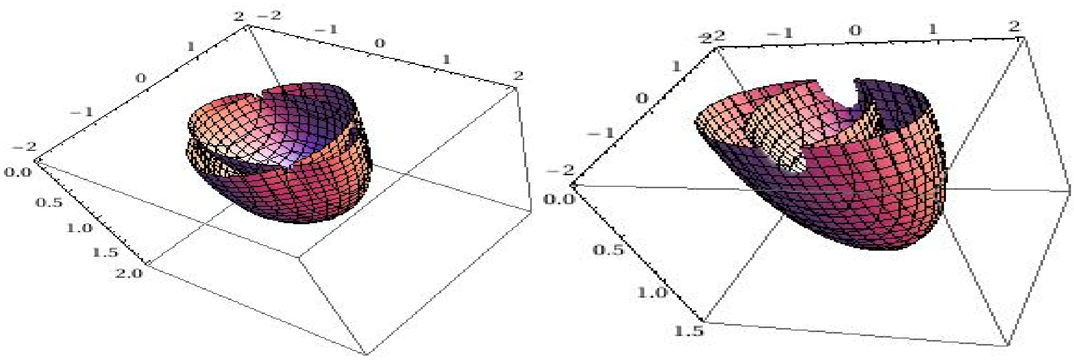

FIGURE 6.1. This picture provides the vanishing set of the Fresnel polynomial arising from class I area metrics for two cases with $q_0 = 1$ in the corresponding polynomial $P(q)$: (i) class I with parameter values $\sigma_1 = \sigma_2 = \sigma_3 = 0$ and $\tau_1 = 3/2$, $\tau_2 = 1\,, \tau_3 = 2$, which corresponds to an irreducible polynomial, and (ii) class I with parameter values $\sigma_1 = \sigma_2 = \sigma_3 = 0$ and $\tau_1 = \tau_3 = 2\,, \tau_2 = 1$, which corresponds to a reducible bi-quadratic polynomial. In both cases, there are intersection points of the corresponding vanishing sets, which need to be removed from N to obtain N^{smooth}.

eikonal equation in the metric case, the solution of the equation $(\partial_0 S)^2 = 0$ does not propagate and must therefore be excluded. We thus recognize

$$P_G(x, \partial S) = 0$$

as the physical eikonal equation whose solutions represent truly field propagation. This equation has no repeated factors (unless the area metric G is induced from a metric g by virtue of $G_{abcd} = g_{ac}g_{bd} - g_{ad}g_{bc}$) and hence is already reduced. We thus identify the covariant cotangent bundle function providing the geometry of the cotangent bundle as $P_{x\,G}(q)$. The condition $P_{x\,G}(q) = 0$ defines the massless dispersion relation with the set of massless momenta N_x at a given point x of M given by

$$N_x = \{q \in T^*_x M \mid P_{G\,x}(q) = 0\}\,,$$

which in contrast to the metric case can differ from N^{smooth} (see figure 6.1).

Condition II: Interpretability (= time-orientability of $P_{G\,x}$). In order to satisfy the interpretability condition, we need the cotangent bundle function $P_G(x, q)$ to be hyperbolic and time-orientable. So we need to compute the dual polynomial of P_G. But this cotangent bundle function P_G is clearly more complicated than the metric one. At first sight it could seem that there is simply no way to avoid the use of elimination theory. However, already in four dimensions, elimination theory is prohibitively difficult for current computer algebra programs, even if full use is made of our knowledge of the normal forms of area metrics. So while in principle Buchberger's algorithm applies, practically one is better off obtaining an educated guess for what the dual polynomial might be, and then verifying that guess employing equation (3.17). Thanks to the invertibility properties of area metrics, an educated guess for the dual of P_G can be done, namely the tangent bundle function

$$\mathcal{P}_{G\,x}(v) = -\frac{1}{24}(\omega^{-1}_{G_x})^{mnpq}(\omega^{-1}_{G_x})^{rstu}G_{x\,mnr(a}G_{x\,b|ps|c}G_{x\,d)qtu}v^a v^b v^c v^d\,, \tag{6.29}$$

which has practically the same structure as P_G but with the inverse area metric G^{-1} replaced by G, and contracted with $\left(\omega_G^{-1}\right)^{mnpq} = f^{-1}(G)\epsilon^{mnpq}$. Indeed, using the algebraic classification of area metrics, it is then a simple exercise for *Mathematica* to verify that for metaclasses I–XI and XIII–XIX, the cotangent bundle function $\mathcal{P}_G$ defined in (6.29) indeed satisfies at every point the defining property of the dual polynomial (3.17). However, the polynomials P arising from the area metric metaclasses VIII–XXIII are not of interest because, as we will see in what follows, they can never give rise to physical spacetimes. Anticipating that result, we recognize $\mathcal{P}_G$ as the dual polynomial $P^\#_G$ to P_G for all viable area metric spacetime geometries (classes I-VII), which as we found provides the geometry on the tangent bundle seen by massless particles.

By taking now $q = q_a\epsilon^a \in T^*_xM$ and $v = v^a e_a \in T_xM$, where $v^a = (t,x,y,z)$ and $\{\epsilon^a\}$ is the frame in which the found metaclasses for area metrics are expressed, we provide the explicit expression for the polynomials P_G and their duals $P^\#$ for metaclasses I–VII.

Class I

$$\begin{aligned}
P(q) \;=\;& \frac{1}{\left((\tau_1^2+\sigma_1^2)(\tau_2^2+\sigma_2^2)(\tau_3^2+\sigma_3^2)\right)^{1/3}}\big[\tau_1\tau_2\tau_3(q_0^4+q_1^4+q_2^4+q_3^4)\\
&+\tau_1(\tau_2^2+\tau_3^2+(\sigma_3-\sigma_2)^2)(q_2^2q_3^2-q_0^2q_1^2)+\tau_2(\tau_1^2+\tau_3^2+(\sigma_3-\sigma_1)^2)(q_1^2q_3^2-q_0^2q_2^2)\\
&+\tau_3(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2)(q_1^2q_2^2-q_0^2q_3^2)\\
&+2q_0q_1q_2q_3\left(\tau_2^2(-\sigma_3+\sigma_1)+\tau_1^2(\sigma_3-\sigma_1)-(\tau_3^2+(\sigma_3-\sigma_2)(\sigma_3-\sigma_1))(\sigma_1-\sigma_2)\right)\big]
\end{aligned}$$

$$\begin{aligned}
P^\#(v) \;=\;& \frac{1}{\left((\tau_1^2+\sigma_1^2)(\tau_2^2+\sigma_2^2)(\tau_3^2+\sigma_3^2)\right)^{1/3}}\big[\tau_1\tau_2\tau_3(t^4+x^4+y^4+z^4)\\
&+\tau_1(\tau_2^2+\tau_3^2+(\sigma_3-\sigma_2)^2)(y^2z^2-t^2x^2)+\tau_2(\tau_1^2+\tau_3^2+(\sigma_3-\sigma_1)^2)(x^2z^2-t^2y^2)\\
&+\tau_3(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2)(x^2y^2-t^2z^2)\\
&-2txyz\left(\tau_2^2(-\sigma_3+\sigma_1)+\tau_1^2(\sigma_3-\sigma_1)-(\tau_3^2+(\sigma_3-\sigma_2)(\sigma_3-\sigma_1))(\sigma_1-\sigma_2)\right)\big]
\end{aligned}$$

Class II

$$\begin{aligned}
P(q) \;=\;& \frac{1}{\left((\tau_1^2+\sigma_1^2)^2(\tau_2^2+\sigma_2^2)\right)^{1/3}}\big[-q_3^4\tau_2+q_1^4\tau_1^2\tau_2+2q_1^2q_2^2\tau_1^2\tau_2+q_2^4\tau_1^2\tau_2+4t^2q_3^2\tau_1^2\tau_2\\
&+2q_1q_2q_3^2\left(\tau_1^2-\tau_2^2-(\sigma_1-\sigma_2)^2\right)-2q_0q_1^2q_3\tau_1\left(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2\right)\\
&-2tq_2^2q_3\tau_1\left(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2\right)+2q_1^2q_3^2\tau_1(\sigma_1-\sigma_2)+2q_2^2q_3^2\tau_1(-\sigma_1+\sigma_2)\big]
\end{aligned}$$

$$\begin{aligned}
P^\#(v) \;=\;& \frac{1}{\left((\tau_1^2+\sigma_1^2)^2(\tau_2^2+\sigma_2^2)\right)^{1/3}}\big[-t^4\tau_2+\left(x^2+y^2\right)^2\tau_1^2\tau_2\\
&-2t\left(x^2+y^2\right)z\tau_1\left(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2\right)\\
&+2t^2\left(xy\left(\tau_1^2-\tau_2^2-(\sigma_1-\sigma_2)^2\right)+x^2\tau_1(\sigma_1-\sigma_2)+\tau_1\left(2z^2\tau_1\tau_2+y^2(-\sigma_1+\sigma_2)\right)\right)\big]
\end{aligned}$$

Class III

$$P(q) = \frac{1}{{\tau_1}^2+{\sigma_1}^2}\left[2q_2{q_3}^3 - {q_1}^4{\tau_1}^3 - {q_2}^4{\tau_1}^3 + 2q_0{q_3}^2\tau_1\left(q_3 - 2q_0{\tau_1}^2\right) + {q_1}^2q_3\tau_1\left(3q_3+4q_0{\tau_1}^2\right)\right.$$
$$\left.+{q_2}^2\left(3{q_3}^2\tau_1 - 2{q_1}^2{\tau_1}^3 + 4q_0q_3{\tau_1}^3\right)\right]$$

$$P^{\#}(v) = -\frac{5}{\left({\tau_1}^2+{\sigma_1}^2\right)^2}\left[\left(x^2+y^2\right)^2{\tau_1}^3 - 4t\left(x^2+y^2\right)z{\tau_1}^3 + 2t^3(y-z\tau_1)\right.$$
$$\left.+t^2\left(-3x^2\tau_1 - 3y^2\tau_1 + 4z^2{\tau_1}^3\right)\right]$$

Class IV

$$P(q) = -\frac{1}{\left(\left(\tau_1^2+\sigma_1^2\right)\left(\tau_2^2+\sigma_2^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{1/3}}\left[\tau_1\tau_2\lambda_1\left(q_0^4-q_1^4-q_2^4+q_3^4\right)\right.$$
$$-\lambda_1\left(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2\right)\left(q_1^2q_2^2+q_0^2q_3^2\right) + \tau_2\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)\left(q_0^2q_2^2-q_1^2q_3^2\right)$$
$$+\tau_1\left(\tau_2^2-\lambda_1^2+(\sigma_2-\lambda_2)^2\right)\left(q_0^2q_1^2-q_2^2q_3^2\right)$$
$$+2q_0q_1q_2q_3\left(\lambda_1^2(-\sigma_1+\sigma_2)+\left(\tau_1^2+(\sigma_1-\sigma_2)(\sigma_1-\lambda_2)\right)(\sigma_2-\lambda_2)+\tau_2^2(-\sigma_1+\lambda_2)\right)\Big)$$

$$P^{\#}(v) = \frac{1}{\left(\left(\tau_1^2+\sigma_1^2\right)\left(\tau_2^2+\sigma_2^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{2/3}}\left[\tau_1\tau_2\lambda_1\left(t^4-x^4-y^4+z^4\right)\right.$$
$$-\lambda_1\left(\tau_1^2+\tau_2^2+(\sigma_1-\sigma_2)^2\right)\left(x^2y^2+t^2z^2\right) - \tau_2\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)\left(t^2y^2-x^2z^2\right)$$
$$-\tau_1\left(\tau_2^2-\lambda_1^2+(\sigma_2-\lambda_2)^2\right)\left(t^2x^2-y^2z^2\right)$$
$$+2txyz\left(\lambda_1^2(-\sigma_1+\sigma_2)+\left(\tau_1^2+(\sigma_1-\sigma_2)(\sigma_1-\lambda_2)\right)(\sigma_2-\lambda_2)+\tau_2^2(-\sigma_1+\lambda_2)\right)\Big)$$

Class V

$$P(q) = -\frac{1}{\left((\tau_1^2+\sigma_1^2)^2(\lambda_1^2-\lambda_2^2)\right)^{1/3}}\left[-q_0^4\lambda_1 - \left(q_1^2+q_2^2\right)^2\tau_1^2\lambda_1\right.$$
$$-2q_0\left(q_1^2+q_2^2\right)q_3\tau_1\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)$$
$$\left.+2q_0^2\left(q_1q_2\left(\tau_1^2+\lambda_1^2-(\sigma_1-\lambda_2)^2\right)+q_1^2\tau_1(\sigma_1-\lambda_2)+\tau_1\left(2q_3^2\tau_1\lambda_1+q_2^2(-\sigma_1+\lambda_2)\right)\right)\right]$$

$$P^{\#}(v) = -\frac{1}{\left((\tau_1^2+\sigma_1^2)^2(\lambda_1^2-\lambda_2^2)\right)^{2/3}}\left[-t^4\lambda_1 - \left(x^2+y^2\right)^2\tau_1^2\lambda_1\right.$$
$$-2t\left(x^2+y^2\right)z\tau_1\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)$$
$$\left.+2t^2\left(xy\left(\tau_1^2+\lambda_1^2-(\sigma_1-\lambda_2)^2\right)+x^2\tau_1(\sigma_1-\lambda_2)+\tau_1\left(2z^2\tau_1\lambda_1+y^2(-\sigma_1+\lambda_2)\right)\right)\right]$$

Class VI

$$
\begin{aligned}
P(q) \;=\; & -\frac{1}{\left(\left(\tau_1^2+\sigma_1^2\right)\left(\lambda_4^2-\lambda_3^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{1/3}}\left[\tau_1\lambda_1\lambda_3\left(q_0^4+q_1^4-q_2^4-q_3^4\right)\right.\\
&+\lambda_1\left(\tau_1^2+\lambda_4^2-2\lambda_4\sigma_1+\sigma_1^2-\lambda_3^2\right)\left(q_1^2q_2^2+q_0^2q_3^2\right)\\
&+\lambda_3\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)\left(q_0^2q_2^2+q_1^2q_3^2\right)\\
&+\left(\lambda_1^2-\lambda_4^2+\lambda_3^2+2\lambda_4\lambda_2-\lambda_2^2\right)\left(q_0^2q_1^2\tau_1-q_2^2q_3^2\right)\\
&\left.-2q_0q_1q_2q_3\left(\lambda_1^2(\lambda_4-\sigma_1)+\tau_1^2(\lambda_4-\lambda_2)-(\sigma_1-\lambda_2)\left(\lambda_4^2-\lambda_3^2+\sigma_1\lambda_2-\lambda_4(\sigma_1+\lambda_2)\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
P^{\#}(v) \;=\; & \frac{1}{\left(\left(\tau_1^2+\sigma_1^2\right)\left(\lambda_4^2-\lambda_3^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{2/3}}\left[\tau_1\lambda_1\lambda_3\left(t^4+x^4-y^4-z^4\right)\right.\\
&-\lambda_1\left(\tau_1^2+\lambda_4^2-2\lambda_4\sigma_1+\sigma_1^2-\lambda_3^2\right)\left(x^2y^2+t^2z^2\right)\\
&-\lambda_3\left(\tau_1^2-\lambda_1^2+(\sigma_1-\lambda_2)^2\right)\left(t^2y^2+x^2z^2\right)\\
&+\tau_1\left(\lambda_1^2-\lambda_4^2+\lambda_3^2+2\lambda_4\lambda_2-\lambda_2^2\right)\left(t^2x^2-y^2z^2\right)\\
&\left.-2txyz\left(\lambda_1^2(-\lambda_4+\sigma_1)+\tau_1^2(-\lambda_4+\lambda_2)+(\sigma_1-\lambda_2)\left(\lambda_4^2-\lambda_3^2+\sigma_1\lambda_2-\lambda_4(\sigma_1+\lambda_2)\right)\right)\right]
\end{aligned}
$$

Class VII

$$
\begin{aligned}
P(q) \;=\; & \frac{1}{\left(\left(\lambda_5^2-\lambda_6^2\right)\left(\lambda_4^2-\lambda_3^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{1/3}}\left[-\lambda_5\lambda_1\lambda_3\left(q_0^4+q_1^4+q_2^4+q_3^4\right)\right.\\
&-\lambda_1\left(\lambda_5^2-\lambda_4^2+2\lambda_4\lambda_6-\lambda_6^2+\lambda_3^2\right)\left(q_1^2q_2^2+q_0^2q_3^2\right)\\
&-\lambda_3\left(\lambda_5^2+\lambda_1^2-(\lambda_6-\lambda_2)^2\right)\left(q_0^2q_2^2+q_1^2q_3^2\right)\\
&+\lambda_5\left(-\lambda_1^2+\lambda_4^2-\lambda_3^2-2\lambda_4\lambda_2+\lambda_2^2\right)\left(q_0^2q_1^2+q_2^2q_3^2\right)\\
&\left.+2q_0q_1q_2q_3\left(\lambda_1^2(-\lambda_4+\lambda_6)+\lambda_5^2(\lambda_4-\lambda_2)+(\lambda_6-\lambda_2)\left(\lambda_4^2-\lambda_3^2+\lambda_6\lambda_2-\lambda_4(\lambda_6+\lambda_2)\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
P^{\#}(v) \;=\; & \frac{1}{\left(\left(\lambda_5^2-\lambda_6^2\right)\left(\lambda_4^2-\lambda_3^2\right)\left(\lambda_1^2-\lambda_2^2\right)\right)^{2/3}}\left[\lambda_5\lambda_1\lambda_3(t^4+x^4+y^4+z^4)\right.\\
&+\lambda_1\left(\lambda_5^2-\lambda_4^2+2\lambda_4\lambda_6-\lambda_6^2+\lambda_3^2\right)(x^2y^2+t^2z^2)\\
&+\lambda_3\left(\lambda_5^2+\lambda_1^2-(\lambda_6-\lambda_2)^2\right)(t^2y^2+x^2z^2)\\
&+\lambda_5\left(\lambda_1^2-\lambda_4^2+\lambda_3^2+2\lambda_4\lambda_2-\lambda_2^2\right)(t^2x^2+y^2z^2)\\
&\left.-2txyz\left(\lambda_1^2(-\lambda_4+\lambda_6)+\lambda_5^2(\lambda_4-\lambda_2)+(\lambda_6-\lambda_2)\left(\lambda_4^2-\lambda_3^2+\lambda_6\lambda_2-\lambda_4(\lambda_6+\lambda_2)\right)\right)\right]
\end{aligned}
$$

Thus having the dual polynomial, we would still have to check *the hyperbolicity and time-orientability conditions. For those classes satisfying the hyperbolicity and time-orientability conditions*, picking a hyperbolic vector e_0 with respect to $P_G^\#$ corresponds to picking a time-orientation for M. The hyperbolicity cone $C^\# = C(P_G^\#, e_0)$ is then defined as the cone of future observers. Moreover, the massless point particle action which defines the trajectory of light rays[1] is

$$S[x] = \int d\tau\, \mu\, P_G^\#(x, \dot{x})\,.$$

Condition III: Quantizability (= energy-distinguishability of P_G). We will now see that the energy-distinguishing condition truly restricts the area metric metaclasses in order to provide a classical spacetime structure. This follows from

LEMMA 6.2.1. (Lemma 5.1 of [18]) *Let (M, G) be a four-dimensional area metric manifold of metaclass $VIII$ to $XXIII$. Then there exists a plane of null covectors.*

But we know from proposition 4.1.1 that the existence of null planes violates the energy-distinguishing condition. Thus, as previously stated, only classes I to VII can provide a classical spacetime structure.

The massive dispersion relation and observer frames. We now identify, from chapter 4, that for this case the massive dispersion relation is

$$P_G(x, q) - m^4 = 0\,.$$

The hyperbolic barrier function is in this case given by $f(q) = -1/\deg P \log P_{G\,x}(q)$ which defines the Legendre map

$$L(q) = \frac{1}{\deg P}\frac{DP_{G\,x}(q)}{P_{G\,x}(q)}\,.$$

We know that the inverse of this Legendre map exists if P is bi-hyperbolic and energy-distinguishing. However, it is prohibitively difficult to obtain an analytical expression for it. Even more, the function P_G^* is not a polynomial. This is so due to theorem 6.3 in [20], which states that if the Legendre dual f_x^L is also a hyperbolic barrier function, then $C(P)$ must be a cone of the Lorentzian type.

Following now the definitions of section (5.1), if $\{\epsilon^0, \epsilon^\alpha\}$ is to be an observer frame, it must satisfy

$$P_{G\,x}(\epsilon^0, \epsilon^0, \epsilon^0, \epsilon^0) = P_{G\,x}^{0000} = 1 \tag{6.30}$$

$$P_{G\,x}(\epsilon^0, \epsilon^0, \epsilon^0, \epsilon^\alpha) = P_{G\,x}^{000\alpha} = 0 \tag{6.31}$$

Thus, in any observer frame, the coefficient $b(\vec{p}) = 0$ in (6.25) must vanish, implying that the matrices K and M must satisfy

$$K_\rho^{[\mu} M^{\nu]\rho} = 0 \tag{6.32}$$

[1]Preliminary results for the propagation of light in area metric electrodynamics have been given [43].

in any observer frame. In this case, equations (6.25) reduce to

$$\begin{aligned} a &= -\det(M^{\alpha\beta})\,, \\ b(\vec{p}) &= 0, \\ c(\vec{p}) &= -\left(N_{\rho\sigma}\,G_M^{\kappa\rho\mu\sigma} + \epsilon_{0\alpha\beta\gamma}\epsilon^{0\tau\mu\rho}K^{\beta}_{\tau}K^{\alpha}_{\rho}M^{\gamma\kappa} + 2K^{\epsilon}_{\alpha}K^{[\mu}_{\epsilon}M^{\alpha]\kappa}\right)p_\kappa p_\mu\,, \\ d(\vec{p}) &= -2\epsilon^{\rho\epsilon\nu}\left(2N_{\epsilon\alpha}K^{[\mu}_{\rho}M^{\alpha]\sigma} - K^{\mu}_{\rho}K^{\sigma}_{\alpha}K^{\alpha}_{\epsilon}\right)p_\sigma p_\mu p_\nu\,, \\ e(\vec{p}) &= -N_{\beta\kappa}\epsilon^{\beta\mu\rho}\epsilon^{\kappa\nu\gamma}\left(\frac{1}{2}M^{\lambda\upsilon}N_{\rho\gamma} - K^{\lambda}_{\rho}K^{\upsilon}_{\gamma}\right)p_\lambda p_\mu p_\upsilon p_\nu\,. \end{aligned} \tag{6.33}$$

Moreover, we notice that the normalization condition in equation (6.30) implies that

$$\det(M^{\alpha\beta}) = \det(G^{0\alpha0\beta}) \neq 0\,, \tag{6.34}$$

such that the matrix $G^{0\alpha0\beta}$ is invertible.

We can now prove the following

PROPOSITION 6.2.2. *Let G be an area metric decomposed as in (6.19). If the area metric G gives rise to hyperbolic, time-orientable and energy-distinguishing polynomial P_G and there exist an observer frame in which $K = \phi\mathbb{I}$, then the area metric is of class I.*

Proof. From the assumptions we know that there is an observer frame such that $G^{0\alpha\beta\gamma} = \phi\epsilon^{0\alpha\beta\gamma}$, or equivalently $K^{\alpha}_{\beta} = \phi\delta^{\alpha}_{\beta}$. In this case, using equations (6.33), the polynomial $P_G(p)$ is further reduced to

$$\begin{aligned} P_G(p_0,\vec{p}) &= -p_0^4\det(M) - p_0^2(N_{\mu\nu}G_M^{\kappa\mu\tau\nu})p_\kappa p_\tau - \frac{1}{2}(M^{\epsilon\gamma}p_\epsilon p_\gamma)(G_M^{\beta\kappa\sigma\tau}N_{\beta\sigma}p_\kappa p_\tau) \\ &= -p_0^4\det(M) - p_0^2(N_{\mu\nu}G_M^{\kappa\mu\tau\nu})p_\kappa p_\tau - \det(N)(M^{\epsilon\gamma}p_\epsilon p_\gamma)(N^{-1})^{\kappa\tau}p_\kappa p_\tau\,. \end{aligned} \tag{6.35}$$

From the energy-distinguishing property of P_G, it then follows that $P(p_0,\vec{p}) = 0$ does not have any solutions $p_0 = 0$ unless $\vec{p} = 0$. But then the matrix $M^{\alpha\beta}$ must be of definite signature. For suppose that this is not the case, then one could find $\vec{p} \neq 0$ such that $M^{\mu\nu}p_\mu p_\nu = 0$. That would imply extra zero solutions for p_0, in contradiction to the energy-distinguishing condition. The same holds for the matrix N. Thus, without loss of generality, we assume that M is negative definite; then using Descarte's rule of signs, hyperbolicity of P implies that N must be positive definite. Now, since M is negative definite there is a matrix T such that $T\,M\,T^T = \mathbb{I}$. We can extend the transformation T to a $GL(4)$ transformation Λ by taking $\Lambda^0_\alpha = \Lambda^\alpha_0 = 0$ and $\Lambda^0_0 = 1$. Applying this Λ transformation to the initial area metric, we find that the transformed area metric G' takes the form

$$\text{Petrov}(G') = \begin{bmatrix} -\mathbb{I} & \phi\det(T)\mathbb{I} \\ \phi\det(T)\mathbb{I} & N' \end{bmatrix}\,, \tag{6.36}$$

where $N' = \tilde{T}N\tilde{T}^T$ and $\tilde{T} = \det(T)T^{-1}$. In particular, N' remains positive definite under this transformation. Computing now the eigenvalues of the endomorphism $J_{cd}{}^{ab} = G^{abmn}(\omega_g)_{mncd}$, one finds the eigenvalues $\lambda_i^{\pm} = \phi\det(T) \pm i\sqrt{\mu_i}$, where μ_i are the positive and hence real eigenvalues of N'. But this corresponds precisely to class I area metrics, which shows the proposition.

This proposition shows that if there exists a frame so that $K = \phi\mathbb{I}$, then the area metric must be of class I in order to provide a classical spacetime structure. But this is important because in many linear materials there is no mixture of the electric and magnetic fields in the field induction, and this precisely corresponds to the case $\phi = 0$ in the matrix K.

Remarks. As we have seen, in the case of area metric spacetimes one clearly recognizes the different rôles played by the different structures G, P, $P^{\#}$ and P^*, although they are all defined in terms of the area metric tensor: The gauge field A couples directly to the inverse area metric; however, the geometry of the tangent bundle is provided by the Fresnel polynomial P_G, which is constructed from the inverse area metric but it is clearly structurally different to it; on the other hand, in the tangent bundle, the geometry seen by massless particles is provided by the dual polynomial $P_G^{\#}$, which is constructed from the area metric, while the geometry seen by massive particles is given by P^*, which cannot even have a polynomial structure. Moreover, we have also seen that the hyperbolicity, time-orientability and energy-distinguishing conditions imply that the possible area metric metaclasses are reduced to those belonging to metaclass I to VII. This restriction is the area metric analogue to the Lorentzian signature condition on metrics.

6.3. Testing modified dispersion relations

The developments of this work are far from academic musings of only remote relevance to physics. Indeed, the identification of the *hyperbolicity, time-orientability and the energy-distinguishing conditions* as inevitable properties of dispersion relations, once known, provide a simple algebraic check on the physical consistency of any given dispersion relation. How powerful these conditions are has already been shown when we derived that only for certain classes of area metric geometries, the general linear electrodynamics formulated on such backgrounds satisfy the physicality conditions.

In this section, we show that it is equally simple to extract from our results that some rather popular modifications of electrodynamics, namely those of Gambini-Pullin and Myers-Pospelov, indeed possess dispersion relations that render the underlying field theory non-predictive. In the case of Myers-Pospelov, hyperbolicity (and thus predictivity) can be restored, but unfortunately only at the expense of destroying the energy-distinguishing conditions (and thus a well-defined notion of positive energy). These theories thus do not have the physical interpretation that would be required in order to render observational investigations of bounds on their parameters meaningful. It is obvious that it is both necessary, and indeed simple, to subject also any other proposal for modified dispersion relations to the same straightforward tests.

Gambini-Pullin field equations. Gambini and Pullin [44] obtained a modified dispersion relation by studying the interaction Hamiltonian for electromagnetic and gravitational fields in a semi-classical approximation motivated by loop quantum gravity. More precisely, they found the following refined equations for the electromagnetic field

$$\begin{aligned} \nabla\times\vec{B}-\partial_t\vec{E}+\alpha\nabla^2(\nabla\times\vec{B}) &= 0 \\ \nabla\times\vec{E}+\partial_t\vec{B}+\alpha\nabla^2(\nabla\times\vec{E}) &= 0, \end{aligned} \tag{6.37}$$

with α being a length scale. In fact, it is easy to see that equations (6.37) are not well-posed. For if one defines $u^A=(\vec{E},\vec{B})$, equations (6.37) become

$$D_{AB}(\partial)u^B=0, \tag{6.38}$$

with $D_{AB}(\partial)$ a matrix-valued differential operator explicitly given by

$$D_{AB}(\partial)=\begin{bmatrix} -\delta_{ik}\partial_t & \epsilon_{ijk}\partial_j+\alpha\,\epsilon_{ijk}\partial_l\partial_l\partial_j \\ \alpha\,\epsilon_{ijk}\partial_j+\epsilon_{ijk}\partial_l\partial_l\partial_j & \delta_{ik}\partial_t \end{bmatrix}, \tag{6.39}$$

where in the above expression ϵ_{ijk} is the standard Levi-Civita symbol and Einstein's summation convention is used. If we now try to compute the principal symbol of this equation as prescribed in chapter 2, we will find that it is identically zero. This is not a problem at all for the case of a linear partial differential equation with constant coefficients, as it is the case here, because we can then use the following

Proposition 6.3.1. (Lemma 3.2 of [45]) *A differential operator $D(\partial)$ whose coefficients are constant square matrices is hyperbolic with respect to a covector q if and only if its determinant* $(\det D)(\partial)$ *has that property.*

Hence, in order to test for hyperbolicity and to compute the principal symbol of the operator (6.38), we only have to compute its determinant and read the principal symbol, as usual, from it. Thus the polynomial $Q(q)=\det(D(iq))$ associated with the differential operator (6.39) is easily found to be

$$Q(q)=\det(D(iq))=-q_0^2\left(q^2+2\alpha(\vec{q}\cdot\vec{q})^2-\alpha^2(\vec{q}\cdot\vec{q})^3\right)^2, \tag{6.40}$$

with $p^2=p_0^2-\vec{p}\cdot\vec{p}$, so that its principal part is given by

$$P(q_0,\vec{q})=\alpha^4q_0^2(\vec{q}\cdot\vec{q})^6\,, \tag{6.41}$$

which cannot be physical. This is so because the factor $\vec{q}\cdot\vec{q}$ is not even hyperbolic, and the factor q_0 is not admissible because it gives rise to non-propagating solutions (as we saw in the discussion for the geometric optical limit in the metric and area metric cases.) Even if one argues that only lower orders of α should be considered, the problem remains. Thus, the Gambini-Pullin field equations are not predictive, and the corresponding dispersion relation is non-physical. It has been argued [46] that inadequate quantum states were considered by Gambini and Pullin in the derivation of equations (6.37). Urrutia et. al. [47] performed a re-examination of Gambini-Pullin calculations, with a more careful motivation for the quantum states considered. However, these result in only slightly different refined equations (neglecting a non-linear term in the magnetic field) for the electromagnetic field, and a very similar analysis as above also shows that again the associated dispersion relation is not hyperbolic.

Myers-Pospelov field equations. Myers and Pospelov studied dimension 5 operators [48] leading to cubic modified dispersion relations. Specifically, they proposed the following modified equations for the electromagnetic field

$$D_a^b(\partial)A_b = 0, \tag{6.42}$$

with $D_b^a(\partial)$ a matrix-valued differential operator explicitly written as

$$D_b^a(\partial) = \Box\delta_b^a + \gamma\,\eta_{cb}\,\epsilon^{cdea}n_d(n\cdot\partial)^2\partial_e, \tag{6.43}$$

where in the above expression, γ is the free parameter of the theory, η is the standard Lorentzian metric $\eta = \mathrm{diag}(1,-1,-1,-1)$, and n is a time-like covector with respect to η, i.e. $\eta(n,n) > 0$, that breaks Lorentz invariance. For the operator (6.43), using again proposition 6.3.1, one finds

$$\det(D(iq)) = (q^2)^2\left[(q^2)^2 - \gamma^2(n\cdot q)^4\left(n^a n^c - \eta^{ac}n^2\right)q_a q_c\right], \tag{6.44}$$

from which we read off the principal part

$$P(q) = \gamma^2(p^2)^2(n\cdot p)^4\left(n^a n^c - \eta^{ac}n^2\right)p_a p_c\,, \tag{6.45}$$

which for $\eta(n,n) > 0$ is not hyperbolic. This is so, because the matrix $n^a n^c - \eta^{ac}n^2$, under the assumption $\eta(n,n) > 0$, is positive semi-definite, which implies that one of the factors of the principal part of $P(p)$, namely $\left(n^a n^c - \eta^{ac}n^2\right)p_a p_c$ is not hyperbolic. Hence, the Myers-Pospelov field equations are non-predictive. Furthermore, even if one were to choose n such that $\eta(n,n) \leq 0$, one would still have a null-plane due to the term $n\cdot p$, which we saw at the end of chapter 4 to obstruct the energy-distinguishing property, and thus to lead to a non-physical dispersion relation.

In conclusion, the field equations found by Gambini, Pullin and Urrutia in the spirit of loop quantum gravity, as well as the field equations found by Myers and Pospelov in the framework of effective field theory do not lead to physical dispersion relations. More precisely, there is no spacetime hypersurface Σ on which initial data for the electromagnetic field could be given so that its values on a later hypersurface would be uniquely prescribed. Hence, phenomenological conclusions, such as the identification of bounds based on these modified dispersion relations, are unfortunately not conclusive.

CHAPTER 7

Free QED on area metric spacetimes

We carefully develop the Hamiltonian formulation of area metric electrodynamics in four dimensions from first principles taking into account the fourth-order polynomial dispersion relation and associated causal structure. Canonical quantization of the resulting constrained system then results in a quantum vacuum sensitive to the electromagnetic constitutive tensor of the classical theory. As an application, we calculate the Casimir effect in a bi-refringent linear optical medium.

7.1. Hamiltonian formulation and gauge fixing for area metric electrodynamics

Consider the action for area metric electrodynamics

$$S[A,G] = -\frac{1}{8}\int d^4x f(G)\,\left[F_{ab}F_{cd}\,G^{abcd}\right]\,, \tag{7.1}$$

whose causal structure is encoded in the cotangent bundle function P_G studied in chapter 6 (see equation (6.23)). We will restrict our investigation to flat area metric spacetimes, which is to say that there are global coordinate systems in which the components of the area metric have constant values. Moreover, we will assume, without loss of generality, that the factor $f(G)$ featuring in the volume form $(\omega_G)_{abcd} = f(G)\epsilon_{abcd}$ for area metric spacetimes (see equation 6.10) satisfies $f(G) = |\det(G)|^{1/6} = 1$, such that the volume form is only given by the totally antisymmetric Levi-Civita symbol ϵ_{abcd} defined by $\epsilon_{0123} = +1$.

We then want to develop the Hamiltonian formulation of the dynamics encoded in the action (7.1). In the coordinate system we have chosen, we obtain the canonical momenta associated with the field variables (A_0, A_i) from the Lagrangian density $\mathcal{L}$ in the action (8.1) as

$$\begin{aligned} \pi^0 = \frac{\delta\mathcal{L}}{\delta(\partial_0 A_0)} &= 0, \\ \pi^\alpha = \frac{\delta\mathcal{L}}{\delta(\partial_0 A_\alpha)} &= -G^{0\alpha0\beta}\partial_0 A_\beta - G^{0\alpha\beta0}\partial_\beta A_0 - G^{0\alpha\beta\gamma}\partial_\beta A_\gamma, \end{aligned} \tag{7.2}$$

where latin spacetime indices range from 0 to 3, while purely spatial greek indices range from 1 to 3. In the language of the theory of constrained systems [49, 50, 51, 52], we thus identify $\phi_1 = \pi^0 \approx 0$ as a primary constraint of the dynamics. We now define the matrix $M_{\alpha\beta}$ such that $M_{\alpha\beta}G^{0\beta0\gamma} = \delta^\gamma_\alpha$, whose existence is guaranteed from equation (6.34). Using then equation (7.2) to express $\partial_0 A_\alpha$ in terms of the canonical momenta π^α, we find the total Hamiltonian density

$$\begin{aligned} \mathcal{H} &= -\frac{1}{2}M_{\alpha\beta}\pi^\alpha\pi^\beta - A_0\partial_\alpha\pi^\alpha - \pi^\alpha M_{\beta\alpha}G^{0\beta\gamma\rho}\partial_\gamma A_\rho \\ &\quad +\frac{1}{2}G^{\alpha\beta\gamma\rho}\partial_\alpha A_\beta\partial_\gamma A_\rho - \frac{1}{2}M_{\alpha\beta}G^{0\alpha\gamma\kappa}G^{0\beta\mu\nu}\partial_\mu A_\nu\partial_\gamma A_\kappa + u_1(x)\pi^0(x). \end{aligned} \tag{7.3}$$

Following the Dirac-Bergmann algorithm [50] for obtaining the Hamiltonian formulation of systems with constraints, we now compute the commutator $\{\pi^0, \mathcal{H}\}$. If this commutator does not automatically vanish, we need to impose $\{\pi^0, \mathcal{H}\} \approx 0$ as a secondary constraint in order to ensure that the primary constraint $\phi_1 \approx 0$ is preserved under time evolution. Indeed, one obtains $\{\pi^0, \mathcal{H}\} = -\partial_\alpha \pi^\alpha$. We thus impose $\phi_2 = \partial_\alpha \pi^\alpha \approx 0$ as a secondary constraint, which must be added to (7.3) with a corresponding Lagrange multiplier. The total Hamiltonian now reads

$$\mathcal{H} = \mathcal{H}_0 + u_1(x)\pi^0(x) + (u_2(x) - A_0)\partial_\alpha \pi^\alpha, \tag{7.4}$$

with

$$\mathcal{H}_0 = -\frac{1}{2}M_{\alpha\beta}\pi^\alpha\pi^\beta - \pi^\alpha M_{\beta\alpha}G^{0\beta\rho\kappa}\partial_\rho A_\kappa + \frac{1}{2}(G^{\mu\nu\gamma\kappa} - M_{\alpha\rho}G^{0\alpha\gamma\kappa}G^{0\rho\mu\nu})\partial_\mu A_\nu \partial_\gamma A_\kappa. \tag{7.5}$$

We then find $\{\phi_2, \mathcal{H}\} = 0$, so that the Dirac-Bergmann algorithm ends here and $\phi_1 \approx 0$ and $\phi_2 \approx 0$ exhaust the constraints. However, $\{\phi_1(t,\vec{x}), \phi_2(t,\vec{y})\} = 0$, so that ϕ_1 and ϕ_2 are first class constraints, implying that the multipliers $u_1(x)$ and $u_2(x)$ are completely undetermined. The infinitesimal gauge transformations induced by (ϕ_1, ϕ_2) on the canonical variables (A_a, π^a) are

$$\begin{aligned} \delta A_a(t,\vec{x}) &= \int d^3y\, \epsilon^I(t,\vec{y})\{A_a(t,\vec{x}), \phi_I(t,\vec{y})\} = \epsilon^1(t,\vec{x})\delta^0_a - \delta^\alpha_a \partial_\alpha \epsilon^2(t,\vec{x})\,, &(7.6)\\ \delta\pi^a(t,\vec{x}) &= \int d^3y\, \epsilon^I(t,\vec{y})\{\pi^a(t,\vec{x}), \phi_I(t,\vec{y})\} = 0, &(7.7) \end{aligned}$$

with $I = 1, 2$ and $\epsilon^I(t,\vec{x})$ being the infinitesimal parameters of the transformations. Knowledge of these generators of gauge transformations allows us to identify classical observables of the theory as those functionals that are invariant under gauge transformations. Equivalently, observables O commute with the constraints $\{O, \phi_I\} \approx 0$. In the present case, it can be checked that the electromagnetic inductions

$$\begin{aligned} D^\alpha &= -G^{0\alpha0\beta}F_{0\beta} - \frac{1}{2}G^{0\alpha\beta\kappa}F_{\beta\kappa} &(7.8)\\ &= -G^{0\alpha0\beta}\partial_0 A_\beta - G^{0\alpha\beta0}\partial_\beta A_0 - G^{0\alpha\beta\kappa}\partial_\beta A_\kappa\,, \\ H_\alpha &= -\frac{1}{2}\epsilon_{0\alpha\beta\kappa}\left[G^{\beta\kappa\mu0}F_{\mu0} + \frac{1}{2}G^{\beta\kappa\mu\nu}F_{\mu\nu}\right] &(7.9)\\ &= -\frac{1}{2}\epsilon_{0\alpha\beta\kappa}\left[G^{\beta\kappa\mu0}(\partial_\mu A_0 - \partial_0 A_\mu) + G^{\beta\kappa\mu\nu}\partial_\mu A_\nu\right]\,, \end{aligned}$$

defined with respect to the chosen foliation of spacetime into spacelike hypersurfaces[1], indeed commute with the constraints, so that they can be used as observables. Thus we are finally able to write the Hamiltonian (7.5) for our system in terms of gauge-invariant observables D^α and H_α as

$$\mathcal{H}_0 = \frac{1}{2}U_{\alpha\beta}D^\alpha D^\beta + \frac{1}{2}V^{\alpha\beta}H_\alpha H_\beta, \tag{7.10}$$

where the matrices U and V are given as

$$\begin{aligned} U_{\alpha\beta} &= -M_{\alpha\beta} + \frac{1}{8}T_{\mu\nu\gamma\kappa}G^{0\sigma\gamma\kappa}G^{0\tau\mu\nu}M_{\sigma(\beta}M_{\alpha)\tau} &(7.11)\\ V^{\alpha\beta} &= -\frac{1}{8}\epsilon^{0\gamma\kappa(\alpha}\epsilon^{|0|\beta)\mu\nu}T_{\mu\nu\gamma\kappa}, &(7.12) \end{aligned}$$

[1]Spacelike hypersurfaces in the general theory developed here are defined in terms the cotangent bundle function P, see chapter 2.

with $T_{\mu\nu\gamma\kappa}$ defined such that

$$(G^{\mu\nu\alpha\beta} - G^{0\rho\mu\nu}G^{0\kappa\alpha\beta}M_{\kappa\rho})T_{\mu\nu\tau u} = -8\,\delta^{\alpha}_{[\tau}\delta^{\beta}_{u]}\,. \tag{7.13}$$

The existence of T is guaranteed due to the invertibility properties of area metrics; indeed, it can be written explicitly in terms of the block matrices in (6.19) constituting the area metric tensor as[2]

$$T_{\alpha\beta\gamma\rho} = -2\epsilon_{0\gamma\rho\mu}\epsilon_{0\alpha\beta\nu}\left((N - K^T MK)^{-1}\right)^{\mu\nu}\,.$$

In order to determine the Dirac brackets [50, 51] associated with our system, one needs to remove the indeterminacy in the Lagrange multipliers by fixing a gauge. This is achieved here by manually imposing two further constraints $\phi_3 \approx 0, \phi_4 \approx 0$ such that $\det\{\phi_I(\vec{x}), \phi_J(\vec{y})\} \neq 0$, with $I, J = 1, \ldots 4$, so that the extended set of constraints ϕ_I is now of second class. In our case, the Euler-Lagrange equations for the gauge field A obtained from the action (8.1) are given by

$$G^{abcd}\,\partial_b\,\partial_d\,A_c = 0, \tag{7.14}$$

which is conveniently split into one temporal equation

$$G^{0\alpha0\beta}\,\partial_\alpha\,\partial_\beta A_0 + \left[G^{0\alpha\beta\gamma}\partial_\alpha\,\partial_\gamma - G^{0\alpha0\beta}\partial_0\,\partial_\alpha\right]A_\beta = 0 \tag{7.15}$$

and three spatial equations

$$\left[G^{0\beta\rho\alpha}\,\partial_\alpha\,\partial_\beta - G^{0\rho0\mu}\,\partial_0\,\partial_\mu\right]A_0 + \left[G^{0\rho0\mu}\partial_0^2 - 2G^{0(\rho\mu)\alpha}\partial_0\,\partial_\alpha + G^{\rho\alpha\mu\sigma}\partial_\alpha\,\partial_\sigma\right]A_\mu = 0\,. \tag{7.16}$$

As the third constraint we impose the Glauber gauge [53]

$$\phi_3 = A_0(\vec{x}) - \int d^3x'\,G(\vec{x},\vec{x}')G^{0\alpha\beta\gamma}\partial'_\alpha\,\partial'_\gamma A_\beta(\vec{x}') \approx 0 \tag{7.17}$$

with $-G^{0\alpha0\beta}\,\partial_\alpha\,\partial_\beta\,G(\vec{x},\vec{x}') = \delta(\vec{x}-\vec{x}')$, or more explicitly,

$$G(\vec{x},\vec{x}') = -\frac{1}{4\pi\sqrt{-M_{\alpha\beta}(x^\alpha - x'^\alpha)(x^\beta - x'^\beta)}}. \tag{7.18}$$

The expression under the square root is non-negative ultimately due to the energy distinguishing property because then the matrix $M_{\alpha\beta}$ can be chosen to be negative definite (see chapter 5). Consistency of the gauge (7.17) with the temporal equation (7.15) requires that the further constraint

$$\phi_4 = G^{0\alpha0\beta}\,\partial_\alpha A_\beta \approx 0 \tag{7.19}$$

holds and one checks that ϕ_3 and ϕ_4 are conserved under time evolution. In summary, the set $\{\phi_I\}$ of all constraints is given by

$$\begin{array}{ll} \phi_1 = \pi^0 \approx 0, & \phi_3 = A_0(\vec{x}) - \int d^3x'\,G(\vec{x},\vec{x}')G^{0\alpha\beta\gamma}\partial'_\alpha\,\partial'_\gamma A_\beta(\vec{x}') \approx 0, \\ \phi_2 = \partial_\alpha\pi^\alpha \approx 0, & \phi_4 = G^{0\alpha0\beta}\,\partial_\alpha\,A_\beta \approx 0, \end{array} \tag{7.20}$$

and satisfies

$$\{\phi_I(t,\vec{x}),\phi_J(t,\vec{y})\} = \int \frac{d^3k}{(2\pi^3)}\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -G^{0\alpha0\beta}\,k_\alpha k_\beta \\ 1 & 0 & 0 & 0 \\ 0 & G^{0\alpha0\beta}k_\alpha k_\beta & 0 & 0 \end{bmatrix} e^{i\vec{k}\cdot(\vec{x}-\vec{y})}. \tag{7.21}$$

[2]This follows from the invertibility of the block $G^{0\alpha0\beta}$ in the area metric.

The matrix above $\{\phi_I(t,\vec{x}),\phi_J(t,\vec{y})\}$ is invertible, so that the constraints ϕ_I are now of second class and the gauge freedom is gone. Its inverse $(\{\phi(\vec{x}),\phi(\vec{y})\}^{-1})^{IJ}$, defined through

$$\int d^3y\,\{\phi_I(\vec{x}),\phi_J(\vec{y})\}\,(\{\phi(\vec{y}),\phi(\vec{z})\}^{-1})^{JM} = \delta_I^M\delta(\vec{x}-\vec{z}), \tag{7.22}$$

is simply given as

$$\{\phi_I(t,\vec{x}),\phi_J(t,\vec{y})\}^{-1} = \int \frac{d^3k}{(2\pi^3)}\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{G^{0\alpha0\beta}k_\alpha k_\beta} \\ -1 & 0 & 0 & 0 \\ 0 & -\frac{1}{G^{0\alpha0\beta}k_\alpha k_\beta} & 0 & 0 \end{bmatrix} e^{i\vec{k}.(\vec{x}-\vec{y})}. \tag{7.23}$$

Equipped with equation (7.23) we can now follow Dirac's procedure and replace the standard Poisson bracket $\{,\}$ by the Dirac bracket $\{,\}_D$, which is defined as

$$\{A(\vec{x}),B(\vec{y})\}_D = \{A(\vec{x}),B(\vec{y})\} - \int d^3z\,d^3w\,\{A(\vec{x}),\phi_I(\vec{z})\}(\{\phi(\vec{z}),\phi(\vec{w})\}^{-1})^{IJ}\{\phi_J(\vec{w}),B(\vec{y})\}. \tag{7.24}$$

Use of the Dirac brackets relieves one from using weak equalities, since by construction insertion of constraints into the Dirac brackets automatically implements these weak equalities. Thus we arrive at the fundamental Dirac brackets of the system

$$\begin{aligned} \{A_a(t,\vec{x}),\pi^b(t,\vec{y})\}_D &= \int \frac{d^3k}{(2\pi)^3}\left[\delta_a^b - \delta_a^0\delta_0^b - \frac{\delta_a^\sigma\delta_\kappa^b\,k_\rho\,k_\sigma\,G^{0\rho0\kappa}}{G^{0\mu0\nu}k_\mu\,k_\nu} - \frac{\delta_a^0\delta_\sigma^b G^{0\rho\sigma\gamma}\,k_\rho\,k_\gamma}{G^{0\mu0\nu}k_\mu\,k_\nu}\right] e^{i\vec{k}.(\vec{x}-\vec{y})}\,, \\ \{A_a(t,\vec{x}),A_b(t,\vec{y})\}_D &= 0\,, \\ \{\pi^a(t,\vec{x}),\pi^b(t,\vec{y})\}_D &= 0\,, \end{aligned} \tag{7.25}$$

which can be used to proceed with the quantization of the system. The dynamics of the system is therefore generated by the Hamilton equations

$$\begin{aligned} \partial_t A_a(t,\vec{x}) &\approx \int d^3y\,\{A_a(t,\vec{x}),\mathcal{H}_0(\vec{y})\}_D\,, \\ \partial_t\pi^a(t,\vec{x}) &\approx \int d^3y\,\{\pi^a(t,\vec{x}),\mathcal{H}_0(\vec{y})\}_D\,, \end{aligned} \tag{7.26}$$

where, due to the use of Dirac brackets, only $\mathcal{H}_0$ is involved.

So far we have made only implicit use of the requirement that the area metric background be hyperbolic, time-orientable and energy-distinguishing, namely in the abstract constructions underlying the definition of spacetime foliations into spacelike leaves. But now we need to explicitly solve the field equations (7.16) with the gauge imposed by (7.17), and this requires to restrict attention to concrete hyperbolic, time-orientable and energy-distinguishing area metric backgrounds. Moreover, for actual calculations it is most convenient to choose a coordinate frame in which the area metric takes a simple normal form. The normal form theory of area metrics in four dimensions was already described in chapter 6 where it was shown that the area metric cannot be hyperbolic, time-orientable and energy-distinguishing unless the endomorphism J on the space of two-forms defined through

$$J_{cd}{}^{ab} = G^{cdmn}\omega_{mnab} \tag{7.27}$$

has a complex eigenvalue structure (Segré type) of the form $[1\bar{1}1\bar{1}1\bar{1}]$, $[2\bar{2}1\bar{1}]$, $[3\bar{3}]$, $[1\bar{1}1\bar{1}11]$, $[2\bar{2}11]$, $[1\bar{1}1111]$ or $[111111]$. We also know that four-dimensional area metrics that are induced by a

Lorentzian metric automatically lie in the first class, and so does any area metric in the neighbourhood of such a metric-induced area metric. Thus area metrics of immediate phenomenological relevance are clearly those of this first class, which by a $GL(4)$ frame transformation can be be brought to the normal form

$$G^{[ab][cd]} = \begin{bmatrix} -\tau_1 & 0 & 0 & \sigma_1 & 0 & 0 \\ 0 & -\tau_2 & 0 & 0 & \sigma_2 & 0 \\ 0 & 0 & -\tau_3 & 0 & 0 & \sigma_3 \\ \sigma_1 & 0 & 0 & \tau_1 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & \tau_2 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & \tau_3 \end{bmatrix} \qquad \begin{array}{l} \text{for real } \sigma_1, \sigma_2, \sigma_3 \text{ and} \\ \text{real positive } \tau_1,\, \tau_2,\, \tau_3 \end{array} \tag{7.28}$$

It is then straightforward to show that if $\sigma = \sigma_1 = \sigma_2 = \sigma_3$, which is equivalent to the condition

$$G^{0abc} = \sigma\, \epsilon^{0abc}\,, \tag{7.29}$$

the polynomial

$$\begin{aligned} \frac{P(q)}{f^2(G)} \quad = \quad & \tau_1\tau_2\tau_3(q_0^4 + q_1^4 + q_2^4 + q_3^4) + \tau_1(\tau_2^2 + \tau_3^2)(q_2^2 q_3^2 - q_0^2 q_1^2) \\ & + \tau_2(\tau_1^2 + \tau_3^2)(q_1^2 q_3^2 - q_0^2 q_2^2) + \tau_3(\tau_1^2 + \tau_2^2)(q_1^2 q_2^2 - q_0^2 q_3^2) \end{aligned} \tag{7.30}$$

associated with an area metric of this class is hyperbolic with respect to $h = L^{-1}(\partial/\partial t)$. This is most efficiently verified in the normal frame by observing that for $h = (1, 0, 0, 0)$, the real symmetric Hankel matrix $H_1(P_{q,h})$ associated with the polynomial $P_{q,h}$ is positive definite for any covector q, which implies that P is hyperbolic [54, 55]. The dual polynomial $P^{\#}$ takes precisely the same shape in the normal form frame employed here, and therefore is also hyperbolic.

From now on in this chapter, we will restrict our investigation precisely to class I area metrics for which (7.29) holds. In this case, the field equations (7.15) and (7.16) significantly simplify and, hence, we will now be able to obtain their solutions, orthogonalize them appropriately, and thus perform the diagonalization of the Hamiltonian.

It is worth noting that the hyperbolic polynomial (7.30) only factorizes if at least two of the scalars τ_1, τ_2, τ_3 coincide, so that area metrics with a bi-metric dispersion relation merely present a subset of measure zero within the set of area metrics neighbouring Lorentzian metrics. Indeed, for the generic case of mutually different scalars, the polynomial P is irreducible. Thus theories trying to account for birefringence in linear electrodynamics by some sort of bi-metric geometry fail to parametrize almost all relevant geometries near Lorentzian metric ones.

7.2. Quantization

In order to diagonalize the Hamiltonian (7.5) for predictive, interpretable and quantizable general linear electrodynamics with a higher-order polynomial dispersion relation given by (7.30), we first need to find the solutions of the classical field equations (7.15) and (7.16). After choosing the gauges (7.17) and (7.19), the first equation is trivially satisfied, and the second one reduced to

$$\Big[G^{0\alpha 0\beta}\partial_0^2 + G^{\alpha\mu\beta\nu}\partial_\mu\,\partial_\nu\Big]\, A_\beta(t, \vec{x}) = 0\,, \tag{7.31}$$

due to (7.29). Clearly, these field equations are completely equivalent to the field equations arising from (7.26). Specifically, we look for plane wave solutions

$$A_\alpha(t,\vec{x}) = \int \frac{d^3p}{(2\pi)^3} e^{-i(\omega t+\vec{p}.\vec{x})} f_\alpha(\vec{p}), \tag{7.32}$$

so that introducing (7.32) into (7.31) we observe that the equation

$$\left[G^{0\alpha 0\beta}(\omega)^2 + G^{\alpha\mu\beta\nu}p_\mu\, p_\nu\right] f_\beta(\vec{p}) = 0 \tag{7.33}$$

must be satisfied if (7.32) is indeed a solution. Equation (7.33) has non-trivial solutions only if

$$\det\left(G^{0\alpha 0\beta}(\omega)^2 + G^{\alpha\mu\beta\nu}p_\mu\, p_\nu\right) = 0\,. \tag{7.34}$$

The non-zero frequencies ω for which this is the case are precisely the solutions of $P(\omega,\vec{p}) = 0$, compare (6.22). From the energy distinguishing condition of an area metric spacetime, it follows that these frequencies are zero only for $\vec{p} = 0$, but always because of the hyperbolicity of P. It is then further immediate from (7.30) that if some (without loss of generality positive) $\omega(\vec{p})$ is a solution for some non-zero given $\vec{p}$ in our normal frame, then so is $-\omega(\vec{p})$, and that $\omega(\vec{p}) = \omega(-\vec{p})$. Thus we have four non-zero energy solutions $\pm\omega^I(\vec{p})$ labelled by $I = 1, 2$, two positive and two negative ones, for each spatial momentum $\vec{p}$. Therefore any solution of the field equations for the real gauge potential A can be expanded as

$$A_\alpha(t,\vec{x}) = \sum_{I=1,2}\int_{N^{\text{smooth}}} \frac{d^3p}{(2\pi)^3}\left(e^{-i(\omega^I(\vec{p})t+\vec{p}.\vec{x})} f^I_\alpha(\vec{p}) + e^{i(\omega^I(\vec{p})t+\vec{p}.\vec{x})} f^{*I}_\alpha(\vec{p})\right)\,, \tag{7.35}$$

where strictly speaking, the integral is to be taken only over spatial momenta $\vec{p}$ for which the roots ω of $P(\omega,\vec{p})$ are non-degenerate, so that the elementary plane wave solutions are linearly independent. However, the set of covectors for which these zeros are degenerate is of measure zero as we showed in the first lemma of chapter 3, so that this restriction of the integral domain can be technically disregarded. It may be worth emphasizing that the standard appearance of this expansion is somewhat deceptive, since the ω^I appearing here are solutions of the fourth degree polynomial (7.34), rather than the second degree standard Lorentzian dispersion relation.

Having obtained a basis of solutions of the classical field equations, we now identify an inner product that is preserved under time evolution and positive definite for positive energy solutions. To this end, consider solutions $A_\alpha(\vec{p})(t,\vec{x})$ and $\tilde{A}_\alpha(\vec{q})(t,\vec{x})$ of the field equation for specific spatial covectors $\vec{p}$ and $\vec{q}$, respectively. Using the field equation (7.31), it can be shown that the continuity equation

$$\partial_0\left[G^{0\alpha 0\beta}\left(A^*_\alpha(\vec{p})\partial_0\tilde{A}_\beta(\vec{q}) - \tilde{A}_\beta(\vec{p})\partial_0 A^*_\alpha(\vec{q})\right)\right] \tag{7.36}$$

$$+\partial_\mu\left[-G^{\alpha(\mu\nu)\beta}\left(A^*_\alpha(\vec{p})\partial_\nu\tilde{A}_\beta(\vec{q}) - \tilde{A}_\alpha(\vec{p})\partial_\nu A^*_\beta(\vec{q})\right)\right] \;=\; 0 \tag{7.37}$$

is satisfied. This implies that we have a conserved charge Q given by

$$Q = \int d^3x\, G^{0\alpha 0\beta}\left(A^*_\alpha(\vec{p})\partial_0\tilde{A}_\beta(\vec{q}) - \tilde{A}_\beta(\vec{p})\partial_0 A^*_\alpha(\vec{q})\right)\,. \tag{7.38}$$

The above defined charge Q can be used to define a scalar product in the space of solutions, which then by definition is conserved under time evolution and is defined as $(A(\vec{p}),\tilde{A}(\vec{q})) = -i\,Q$.

It satisfies the following properties

$$\begin{aligned} (A(\vec{p}), \lambda\tilde{A}(\vec{q})) &= \lambda(A(\vec{p}), \tilde{A}(\vec{q})) \\ (\lambda A(\vec{p}), \tilde{A}(\vec{q})) &= \lambda^*(A(\vec{p}), \tilde{A}(\vec{q})) \\ (A(\vec{p}), \tilde{A}(\vec{q})) &= (\tilde{A}(\vec{q}), A(\vec{p}))^* = -(A^*(\vec{p}), \tilde{A}^*(\vec{q})). \end{aligned} \tag{7.39}$$

Hence, if we define for our different positive frequency solutions

$$F^I_\alpha(\vec{p})(t, \vec{x}) = e^{-i(\omega^I(\vec{p})t + \vec{p}.\vec{x})} f^I_\alpha(\vec{p}), \tag{7.40}$$

we find that $(F^I(\vec{p}), F^{*J}(\vec{q})) = 0$ and

$$(F^I(\vec{p}), F^J(\vec{q})) = -(F^{*I}(\vec{p}), F^{*J}(\vec{q})) = -2\omega^I(\vec{p}) G^{0\alpha 0\beta} f^{I*}_\alpha(\vec{p}) f^I_\beta(\vec{p}) \delta^{IJ} \delta(\vec{p} - \vec{q}). \tag{7.41}$$

In the derivation of the above results we used charge conservation to find that for $I \neq J$

$$G^{0\alpha 0\beta} f^{I*}_\alpha(\vec{p}) f^{J^*}_\beta(-\vec{p}) = G^{0\alpha 0\beta} f^{I*}_\alpha(\vec{p}) f^J_\beta(\vec{p}) = 0. \tag{7.42}$$

Moreover, since $G^{0\alpha 0\beta}$ is negative definite due to (7.28), equation (7.41) shows that the positive energy solutions can be positively normalized, implying in turn that the negative energy solutions are negatively normalized. This indefiniteness of the scalar product which however is under control by virtue of our positive/negative frequency split is responsible for creation and annihilation processes. Renaming, for convenience,

$$f^I_\alpha(\vec{p}) = \frac{a^I_\alpha(\vec{p})}{\sqrt{2\omega^I(\vec{p})}} ,$$

we finally have

$$(F^I(\vec{p}), F^J(\vec{q})) = -(F^{*I}(\vec{p}), F^{*J}(\vec{q})) = -G^{0\alpha 0\beta} a^{I*}_\alpha(\vec{p}) a^I_\beta(\vec{p}) \delta^{IJ} \delta(\vec{p} - \vec{q}), \tag{7.43}$$

and our general solution reads

$$A_\alpha(t, \vec{x}) = \sum_{I=1,2} \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega^I(\vec{p})}} \left(e^{-i(\omega^I(\vec{p})t + \vec{p}.\vec{x})} a^I_\alpha(\vec{p}) + e^{i(\omega^I(\vec{p})t + \vec{p}.\vec{x})} a^{*I}_\alpha(\vec{p}) \right). \tag{7.44}$$

Now that we have the general solution (7.44), we can use it to write the Hamiltonian, evaluated at a solution, in diagonal form,

$$\begin{aligned} H_0 = \int d^3x\, \mathcal{H}_0(\vec{x}) &= -\frac{1}{2} \int d^3x G^{0\alpha 0\beta} \left(\partial_0 A_\alpha \partial_0 A_\beta - A_\alpha \partial_0^2 A_\beta \right) \\ &= \frac{1}{2} \sum_{I,J} \int \frac{d^3p}{(2\pi)^3} \frac{d^3q}{(2\pi)^3} \omega^J(\vec{p}) \left[(F^I(\vec{p}), F^J(\vec{q})) + (F^J(\vec{p}), F^I(\vec{q})) \right] \\ &= -\frac{1}{2} \sum_{I=1,2} \int \frac{d^3p}{(2\pi)^3} \omega^I(\vec{p}) G^{0\alpha 0\beta} \left[a^{*I}_\alpha(\vec{p}) a^I_\beta(\vec{p})) + a^I_\alpha(\vec{p}) a^{*I}_\beta(\vec{p})) \right]. \end{aligned} \tag{7.45}$$

The last expression shows that the classical Hamiltonian is positive because $G^{0\alpha 0\beta}$ is negative definite.

Equipped with the results developed so far, we are now ready to quantize the electromagnetic field. First, notice that if we multiply equation (7.33) by p_α then the amplitude eigenvectors $a^I_\beta(\vec{p})$ satisfy

$$G^{0\alpha 0\beta} p_\alpha a^I_\beta(\vec{p}) = 0, \tag{7.46}$$

such that the constraints $G^{0\alpha 0\beta}\partial_\alpha A_\beta \approx 0$ and $\partial_\alpha \pi^\alpha \approx 0$ are satisfied. Now due to the first lemma in chapter 3, we know that [13] that for almost all spatial momenta $\vec{p}$, the two associated positive energies do not coincide, $\omega^{I=1}(\vec{p}) \neq \omega^{I=2}(\vec{p})$, so that the covectors $a_\beta^{I=1}(\vec{p})$ and $a_\beta^{I=2}(\vec{p})$ are determined up to scale, linearly independent and thus form a basis for the space of all purely spatial covectors v for which $G^{0\alpha 0\beta}p_\alpha v_\beta = 0$, and diagonalize the Hamiltonian (7.45). Thus the only freedom left is a choice of normalization, which we choose such that any solution $a_\beta^I(\vec{p})$ is expressed as $a_\beta^I(\vec{p}) = a^I(\vec{p})\epsilon_\beta^I(\vec{p})$ with the covectors $\epsilon_\beta^I(\vec{p})$ normalized as

$$-G^{0\alpha 0\beta}\epsilon_\alpha^{I^*}(\vec{p})\epsilon_\beta^I(\vec{p}) = 1, \tag{7.47}$$

where there is no summation over I. Furthermore, p_α and any $a_\beta^I(\vec{p})$ are clearly linearly independent, such that the set of covectors

$$\left\{\epsilon_\beta^{I=1}(\vec{p}), \epsilon_\beta^{I=2}(\vec{p}), \frac{\vec{p}}{\sqrt{-G^{0\alpha 0\beta}p_\alpha p_\beta}}\right\}, \tag{7.48}$$

which span a three-dimensional vector space, is orthonormalized with respect to $G^{0\alpha 0\beta}$. Hence, they satisfy the completeness relation

$$-G^{0\rho 0\sigma}\sum_{I=1,2}\epsilon_\sigma^{I*}(\vec{p})\epsilon_\beta^I(\vec{p}) = \delta_\beta^\rho - \frac{p_\mu p_\beta G^{0\mu 0\rho}}{G^{0\alpha 0\beta}p_\alpha p_\beta}. \tag{7.49}$$

Notice that the normalized covectors $\epsilon_\beta^I(\vec{p})$ satisfy the orthogonality identities (7.42). Now the general solution (7.44) takes the form

$$A_\alpha(t,\vec{x}) = \sum_{I=1,2}\int \frac{d^3p}{(2\pi)^3}\frac{1}{\sqrt{2\omega^I(\vec{p})}}\left(e^{-i(\omega^I(\vec{p})t+\vec{p}.\vec{x})}a^I(\vec{p})\epsilon_\alpha^I(\vec{p}) + e^{i(\omega^I(\vec{p})t+\vec{p}.\vec{x})}a^{I*}(\vec{p})\epsilon_\alpha^{*I}(\vec{p})\right),$$

where the independent coefficients $a^I(\vec{p})$ correspond to the amplitudes of the solutions and depend on the initial values that one considers for a specific problem in the classical approach. At the quantum level, these amplitudes are precisely the mathematical objects that should be promoted to operators, such that the corresponding quantum field reads

$$\hat{A}_\alpha(t,\vec{x}) = \sum_{I=1,2}\int \frac{d^3p}{(2\pi)^3}\frac{1}{\sqrt{2\omega^I(\vec{p})}}\left(e^{-i(\omega^I(\vec{p})t+\vec{p}.\vec{x})}\hat{a}^I(\vec{p})\epsilon_\alpha^I(\vec{p}) + e^{i(\omega^I(\vec{p})t+\vec{p}.\vec{x})}\hat{a}^{I\dagger}(\vec{p})\epsilon_\alpha^{*I}(\vec{p})\right).$$

Using this quantum solution and the expressions for the energy and spatial momentum (which can be obtained by calculating the energy-momentum tensor) we find that the quantum Hamiltonian and quantum spatial momentum operators are given by

$$\hat{H}_0 = \frac{1}{2}\sum_{I=1,2}\int \frac{d^3p}{(2\pi)^3}\omega^I(\vec{p})\left[\hat{a}^I(\vec{p})\hat{a}^{I\dagger}(\vec{p})) + \hat{a}^{I\dagger}(\vec{p})\hat{a}^I(\vec{p}))\right], \tag{7.50}$$

$$\hat{P}_i = \frac{1}{2}\sum_{I=1,2}\int \frac{d^3p}{(2\pi)^3}p_i\left[\hat{a}^I(\vec{p})\hat{a}^{I\dagger}(\vec{p})) + \hat{a}^{I\dagger}(\vec{p})\hat{a}^I(\vec{p}))\right]. \tag{7.51}$$

Hence, if we identify the operators $\hat{a}^I(\vec{p})$, $\hat{a}^{I\dagger}(\vec{p})$ with annihilation and creation operators respectively, a condition for the Hamiltonian to be positive definite is that these operators obey the bosonic commutation relations

$$[\hat{a}^I(\vec{p}), \hat{a}^{J\dagger}(\vec{q})] = (2\pi)^3\delta^{IJ}\delta(\vec{p}-\vec{q}), \tag{7.52}$$

$$[\hat{a}^I(\vec{p}), \hat{a}^J(\vec{q})] = [\hat{a}^{I\dagger}(\vec{p}), \hat{a}^{J\dagger}(\vec{q})] = 0.$$

Hence, the quantum Hamiltonian operator can be written as

$$\hat{H}_0 = \sum_{I=1,2} \int \frac{d^3p}{(2\pi)^3} \omega^I(\vec{p}) \hat{a}^{I\dagger}(\vec{p}) \hat{a}^I(\vec{p})) + \sum_{I=1,2} \frac{1}{2} \int d^3p\, \omega^I(\vec{p}) \delta(0)\,, \tag{7.53}$$

from which expression we identify the energy of the electromagnetic vacuum, which was calculated here for plane wave solutions without any boundary conditions, as

$$E_{\text{vac}}(\text{no boundaries}) = \sum_{I=1,2} \frac{1}{2} \int d^3p\, \omega^I(\vec{p}) \delta(0)\,. \tag{7.54}$$

In the next section, we will calculate how this expression changes if one imposes boundary conditions, and thus obtain the associated Casimir effect. Finally, by using the completeness relation (7.49) one confirms that

$$\left[\hat{A}_\alpha(t,\vec{x}), \hat{\pi}^\beta(t,\vec{y})\right] = i \int \frac{d^3p}{(2\pi)^3} \left[\delta^\beta_\alpha - \frac{p_\mu p_\alpha G^{0\mu0\beta}}{G^{0\rho0\sigma} p_\rho p_\sigma}\right] e^{i\vec{p}.(\vec{x}-\vec{y})}, \tag{7.55}$$

which shows the consistency of the quantization procedure with the Dirac brackets (7.25), since the latter reduce to the above form due to (7.29).

7.3. Casimir effect in a birefringent linear optical medium

The Hamiltonian (7.53) shows that the quantization of general linear electrodynamics leads to a modified quantum vacuum compared to standard non-birefringent Maxwell theory, because the $\omega^I(\vec{p})$ are now solutions to a fourth degree polynomial rather than a second degree one. In fact, local physical phenomena which only depend on the quantum vacuum can be used to test and bound the non-metricity of spacetime. In this section we analyse one such phenomenon, namely the Casimir effect; similar studies can be conducted for the Unruh effect and spontaneous emission.

The Casimir effect [56] arises because of the energy cost incurred by imposing boundary conditions on the electromagnetic field strength. Physically, such boundary conditions arise for instance by introducing perfectly conducting metal plates into the spacetime. For two infinitely extended plates parallel to the 1-2-plane, and this is the configuration we will study here for general linear electrodynamics, the electromagnetic field strength must satisfy the boundary conditions

$$F_{01}|_{\text{plates}} = F_{02}|_{\text{plates}} = F_{12}|_{\text{plates}} = 0 \tag{7.56}$$

everywhere on either plate; this follows, by Stokes' theorem and thus independent of the geometric background, from the physical assumption that the plates are ideal conductors inside of which the field strength must vanish.

Now the key point is that having, or not having, boundary conditions for the vacuum amounts to an energy difference, the so-called Casimir energy

$$E_{\text{Casimir}} = E_{\text{vac}}(\text{plate boundaries}) - E_{\text{vac}}(\text{no boundaries})\,. \tag{7.57}$$

But both energies on the right hand side diverge and need to be regularized such that their difference is independent of the regulator. This is most easily achieved by first considering boundary conditions analogous to (7.56), but for all six faces of a finite rectangular box with faces parallel to the coordinate planes, and separated by coordinate distances L_1, L_2, L_3. In a second step we will then push all faces a very large coordinate distance L apart in order to obtain an expression for E_{vac}(no boundaries) regularized by L, and similarly push all but two faces in order to obtain a corresponding regularized expression for E_{vac}(plate boundaries). The difference of these two regulated quantities will indeed turn out to be finite per unit area and be independent of the regulator L.

Now more precisely, a basis of solutions of general linear electrodynamics satisfying the box boundary conditions is labelled by a triple (n_1, n_2, n_3) of non-negative integers and a polarization $I = 1, 2$ and takes the form

$$\begin{aligned} A_x(\vec{x}) &= a_x^I(n_1,n_2,n_3)\cos(n_1\pi\frac{x}{L_1})\sin(n_2\pi\frac{y}{L_2})\sin(n_3\pi\frac{z}{L_3})\,, \\ A_y(\vec{x}) &= a_y^I(n_1,n_2,n_3)\sin(n_1\pi\frac{x}{L_1})\cos(n_2\pi\frac{y}{L_2})\sin(n_3\pi\frac{z}{L_3})\,, \\ A_z(\vec{x}) &= a_z^I(n_1,n_2,n_3)\sin(n_1\pi\frac{x}{L_1})\sin(n_2\pi\frac{y}{L_2})\cos(n_3\pi\frac{z}{L_3})\,, \end{aligned} \tag{7.58}$$

where the $a_m^I(n_1, n_2, n_3)$ are solutions to equation (7.33) for $\omega^I(n_1\pi/L_1, n_2\pi/L_2, n_3\pi/L_3)$, which always exist if the dispersion relation is hyperbolic, time-orientable and energy distinguishing. The vacuum energy in the presence of the box boundary conditions is thus given by the discrete sum

$$E_{vac}(\text{box boundaries}) = \frac{1}{2}\sum_{\vec{n}=0}^{\infty}\sum_{I=1,2}\omega^I(\pi\frac{n_1}{L_1},\pi\frac{n_2}{L_2},\pi\frac{n_3}{L_3}). \tag{7.59}$$

Removing appropriate faces to a coordinate distance L one finds from this, in the very large L limit, the L-regularized expression for the vacuum energy without boundary conditions

$$E^L_{vac}(\text{no boundaries}) = \frac{L^3}{2\pi^3}\sum_{I=1,2}\int_0^\infty d^3p\,\omega^I(\vec{p}), \tag{7.60}$$

and the L-regularized expression for the vacuum energy in the presence of two plates parallel to the 1-2-plane and separated by a coordinate distance d

$$E^L_{\text{vac}}(\text{plate boundaries}) = \frac{L^2}{2\pi^2}\sum_{I=1,2}\sum_{n'}\int_0^\infty dp_x dp_y\,\omega^I\left(p_x^2,p_y^2,(\frac{n\pi}{d})^2\right)\,, \tag{7.61}$$

where the prime in the summation symbol n means that a factor $1/2$ should be inserted if this integer is zero, for then we have just one independent polarization. Hence we find for the physical vacuum Casimir energy $U(d) = (E_{\text{vac}}(\text{plate boundaries}) - E_{\text{vac}}(\text{no boundaries}))/L^2$ per unit area

$$U(d) = \frac{1}{2\pi^2}\sum_{I=1,2}\left[\sum_{n'}\int_0^\infty dp_x dp_y\,\omega^I\left(p_x^2,p_y^2,(\frac{n\pi}{d})^2\right) - \frac{d}{\pi}\int_0^\infty dp_x dp_y dp_z\,\omega^I(p_x^2,p_y^2,p_z^2)\right]. \tag{7.62}$$

In principle, the execution of the above integrals can proceed as in the standard case. However, with the frequencies ω^I now being solutions to a quartic, rather than quadratic, dispersion relation, these integrals are much harder particularly due to the absence of rotational invariance. Fortunately, the fact that contributions from the two different polarizations $I = 1, 2$ are simply added in the above expression allows for an analytic study of the case where the polynomial P is reducible [57]. In terms of the scalars $\tau_1, \tau_2, \tau_3, \sigma$ defining the area metric in a normal form frame, this is the case if and only if two of the scalars τ_1, τ_2, τ_3 coincide, and we may take $\tau_1 = \tau_2$, for instance. Even in this simplest of non-trivial cases, the Casimir energy crucially depends on the birefringence properties of the underlying general linear electrodynamics. More precisely, the polynomial in (7.30) factorizes into two Lorentzian metrics,

$$P(p) = \tau_1(\tau_1 p_0^2 - \tau_1 p_3^2 - \tau_3(p_1^2 + p_2^2))(\tau_3 p_0^2 - \tau_3 p_3^2 - \tau_1(p_1^2 + p_2^2))\,, \tag{7.63}$$

so that we immediately obtain the positive energy solutions

$$\omega^{I=1} = \left[\frac{1}{\tau_1}\left(\tau_1 p_3^2 + \tau_3 p_2^2 + \tau_3 p_2^2\right)\right]^{1/2} \qquad \text{and} \qquad \omega^{I=2} = \left[\frac{1}{\tau_3}\left(\tau_3 p_3^2 + \tau_1 p_1^2 + \tau_1 p_2^2\right)\right]^{1/2}\,, \tag{7.64}$$

turning (7.62) into a sum of integrals as they appear in the standard Casimir problem on a Lorentzian background. Thus from here on the standard calculation of the Casimir effect [58] can be followed for each of these integrals separately, and one finally obtains the Casimir energy (7.62)

$$U(d) = -\frac{1}{2}\left(\frac{\tau_1}{\tau_3} + \frac{\tau_3}{\tau_1}\right)\frac{\pi^2}{720 d^3}\,. \tag{7.65}$$

This energy difference of course results in a Casimir force

$$F(d) = -U'(d) = -\frac{1}{2}\left(\frac{\tau_1}{\tau_3} + \frac{\tau_3}{\tau_1}\right)\frac{\pi^2}{240 d^4} \tag{7.66}$$

between the plates. The standard Casimir force is recovered if and only if $\tau_1 = \tau_2 = \tau_3$, and irrespective of the value of the scalar σ defined below (7.30). This in turn is equivalent to the absence of classical bi-refringence [59]. Note that the amplification of any bi-refringence is limited only by the technological constraint of how small the separation d between the plates can be made in any realistic set-up. In contrast to classical bi-refringence tests, which usually require accumulative effects over large distances (with all the uncertainties present in such non-local measurements), one sees here that the Casmir force allows for a detection of bi-refringence by way of a highly local measurement. Conversely, of course, experimental measurements of the Casimir force agreeing with the standard prediction within the given technological constraints can be used to put stringent bounds on the non-metricity of the spacetime region where the measurement is conducted.

CHAPTER 8

Coupling currents

We have seen that general linear electrodynamics can be canonically quantized on area metric geometry. In this chapter, we now study possible sources of the electromagnetic field on area metric spacetimes. To this end, we first derive a covariant propagator for flat area metric electrodynamics in order to get the general classical solution in presence of charged currents. We then develop the quantum theory of charged massive point particles, which, in particular, will lead us to a particle-antiparticle interpretation. Finally, we study generalizations of the Dirac equation for hyperbolic, time-orientable and energy-distinguishing spacetimes and Dirac-type sources for the electromagnetic field.

8.1. Covariant propagator

We now wish to consider sources for the electromagnetic field on area metric spacetimes. For this purpose, we consider the action

$$S[A,G] = -\int dx^4\, f(G)\, \left[\frac{1}{8}\, F_{ab}F_{cd}\, G^{abcd} + j^a A_a\right]\,, \tag{8.1}$$

where j is a spacetime vector field, which we immediately see to be a conserved current if it is not to brake the gauge symmetry of the source free theory. For, by variation of this action with respect to the gauge field A, we obtain the field equations

$$\partial_d \left[\frac{f(G)}{2}\, G^{abcd} F_{ab}\right] = f(G) j^c, \tag{8.2}$$

which govern the dynamics of the gauge field A. The condition that the action (8.1) be invariant under the gauge transformation $A'_a = A_a + \partial_a\phi$ (with ϕ any differentiable scalar function) implies that the continuity equation

$$\partial_a j^a + \partial_a(\log f(G)) j^a = 0 \tag{8.3}$$

must be satisfied. We now restrict our discussion to flat area metric spacetimes, so that there is a global coordinate system in which the components of the area metric G are constant throughout the manifold M. Thus, in particular, $f(G)$ is constant in this coordinate system. In what follows we will only work in this coordinate system, such that the field equations (8.2) reduce considerably to

$$G^{abcd}\partial_c\partial_a A_b = j^d \quad \text{with the continuity equation } \partial_a j^a = 0\,. \tag{8.4}$$

In order to solve these field equations, one has to invert the differential operator $G^{abcd}\partial_c\partial_a$. However, due to the gauge invariance of the action, this differential operator is not invertible. This is so because the equation $G^{abcd}\partial_c\partial_a\partial_b\theta(x) = 0$ holds for any differentiable function θ, which means that $\partial_b\theta(x)$ is a null eigenvector of the considered differential operator.

All states connected by a gauge transformation represent the same physical state. Thus, in order to describe the system, it is enough to select a representative for each class. This is achieved by fixing the gauge freedom in the action. More precisely, we will use a covariant gauge, which allows to solve the field equations (8.4) for the physical states. First, we recall from chapter 6 that the field equations for the electric and magnetic fields in area metric electrodynamics gave rise to the cotangent bundle function P_G given in equation (6.23), which essentially arose as the principal symbol of the field equations for the related field strenghts. We thus require that the gauge we will impose must give rise to evolution equations for the gauge field A having the same principal symbol P_G. This is achieved, for instance, by choosing the covariant gauge

$$P^{abcd}\partial_a\partial_b\partial_c A_d = 0\,,$$

which can be thought of as the analogue of the Lorentz gauge in Minkowski spacetime. This gauge can always be reached; for suppose that we begin with a gauge field $A_a(x)$ not satisfying this covariant gauge, using then our initial gauge freedom $A'_a(x) = A_a(x) + \partial_a\theta(x)$ we choose $\theta(x)$ to be a solution of the scalar equation $P^{abcd}\partial_a\partial_b\partial_c\partial_d\theta = -P^{abcd}\partial_a\partial_b\partial_c A_d$[1]. This enforces $A'(x)$ to satisfy the chosen covariant gauge.

The gauge is adapted at the level of the action by introducing it in the Lagrangian using the Lagrange multiplier λ as

$$\mathcal{L} = -f(G)\left[\frac{1}{2}G^{abcd}\partial_a A_b\partial_c A_d + \frac{1}{2}\lambda\left(P^{abcd}\partial_a\partial_b\partial_c A_d\right)^2 + A_a j^a\right]\,, \tag{8.5}$$

Thus variation of $\mathcal{L}$ with respect to λ ensures that the covariant gauge $P^{abcd}\partial_a\partial_b\partial_c A_d = 0$ is satisfied. Using now the Euler-Lagrange equations, variation of the Lagrangian with respect to the gauge field A results in

$$D_G^{db}(\partial)A_b = j^d \tag{8.6}$$

where the differential operator $D_G^{db}(\partial)$ is given by

$$D_G^{db}(\partial) = \left(-G^{abcd}\partial_c\partial_a - \lambda P^{dceu}P^{vmnb}\partial_v\partial_m\partial_n\partial_c\partial_e\partial_u\right)\,. \tag{8.7}$$

Calculation of the principal polynomial of this operator as prescribed in chapter 2 gives the result that it is identically zero. But we already know from chapter 6 that in this case proposition 6.3.1 can be used to obtain the principal polynomial. Thus, according to the mentioned proposition, a simple calculation shows that for any area metric

$$\det(D_G(q)) = \lambda(P_G(q))^3\,,$$

so that the principal symbol of the field equations (8.6) is P_G, as desired, and the differential operator $D_G(\partial)$ can therefore be inverted now. It can also be derived from the theory presented in [45] that, for the flat case considered here, the field equations (8.6) are well-posed if and only if P_G is a hyperbolic polynomial.

[1]The solution of this equation exists and is unique since we asummed that P is a hyperbolic polynomial, see [60].

The inverse of the differential operator in (8.6) provides a covariant propagator[2] and, furthermore, the general solution to the field equations. Inverting the differential operator, we find that the inverse $\tilde{\triangle}_{de}$ (in Fourier space) is given as

$$\tilde{\triangle}_{de} \quad = \quad -\frac{1}{2}\frac{\epsilon_{abcd}\epsilon_{rste}M^{bs}M^{ct}v^a v^r}{P(q)^3} - \frac{1}{\lambda}\frac{q_d q_e}{P(q)^2}\,, \tag{8.8}$$

where

$$v^a = P^{abcd}q_b q_c q_d \qquad \text{and} \qquad M^{cf} = G^{cafb}q_a q_b. \tag{8.9}$$

However, due to charge conservation, the second term on the right hand side of (8.8) does not contribute to the propagator. We therefore finally obtain the actual propagator $\triangle_{de}$ as

$$\triangle_{de}(q) \quad = \quad -\frac{1}{2}\frac{\epsilon_{abcd}\epsilon_{rste}M^{bs}M^{ct}v^a v^r}{P(q)^3} \sim O(q^{-2}). \tag{8.10}$$

Note that, the above expression does not depend at all on the Lagrange multiplier λ. In terms of this propagator, the general solution of (8.6) is given as

$$A_a(x) = -\int d^4y \left[\int d^4p\, e^{i\,p.(x-y)}\,\triangle_{ab}(p)\right] j^b(y)\,. \tag{8.11}$$

Providing the explicit expression of the source j^b for the gauge field A and solving this integral, one finds the solution of the field equations (8.6).

For instance, we can consider that the vector current j^a is produced by a point particle of charge e. The charged point particle is assumed to describe a worldline $y(\tau)$ with parameter τ in spacetime, so that the produced vector current is given by

$$j^a(y) = e\frac{dz^a(\tau)}{d\tau}\frac{\delta(y - z(\tau))}{f(G)}\,. \tag{8.12}$$

which has support on the worldline of the particle. By integrating (8.11) assuming this current, one obtains the analogue of the Liénard-Wiechert potentials on area metric spacetimes. We have thus constructed a covariant propagator for flat area metric spacetimes, which, in particular, provides the general solution of the classical field equations (8.2) in the presence of sources.

8.2. Quantum point charges

We may now study how a charged quantum particle couples to a gauge field A on a flat hyperbolic, time-orientable and energy-distinguishing spacetime. We continue working (and we will do it so in the remaining part of this chapter) in the global coordinate system in which the components of P are constant throughout the spacetime manifold.

For this purpose, we recall that the classical massive point particle action is given by

$$S[x] = -m\int_a^b d\tau\, P^*(\dot{x})^{1/\deg P}\,, \tag{8.13}$$

where m is the mass of the particle and P^* is the tangent bundle function providing the geometry seen by massive particles in each tangent space, which is related to the cotangent bundle function P by equation (4.20). Here it is very important to remember that the particle velocity $\dot{x}$ in the action is constrained to be an element of $L(C)\cup L(-C)$, where L is the Legendre map defined in

[2]Another one has been found by Itin [61] by avoiding to chose a gauge, which however cannot be used to perform quantization.

section 4.3. This is so because, in order to obtain the point particle action above, the Legendre map played a key rôle, but is only well-defined for $C \cup -C$.

This free particle action is invariant under transformations $\tau \mapsto \tau'(\tau)$ that are positive, i.e., $d\tau'/d\tau > 0$, and moreover under the inversion $\tau \to -\tau$. This latter time-inversion invariance is gone, however, when the particle is given a charge e and coupled to the electromagnetic gauge field A, according to the action

$$S[x] = -\int_a^b d\tau \left[m\, P^*(\dot{x})^{1/\deg P} - e\,\dot{x}^a A_a(x)\right] \tag{8.14}$$

The first degree of homogeneity of the Lagrangian in the velocities $\dot{x}$ implies, using Euler's theorem[3] (and the notation $D_a := \partial/\partial\dot{x}^a$), that the Hamiltonian

$$H = p_a\dot{x}^a - \mathcal{L} = \dot{x}^a\, D_a\mathcal{L} - \mathcal{L} = \mathcal{L} - \mathcal{L} = 0\,, \tag{8.15}$$

with $a = 0, \ldots, \dim M - 1$, vanishes. Thus the Hamiltonian is identically zero (as it is the case for any reparametrization invariant theory [62]) and therefore standard quantization methods cannot be used because we cannot obtain all canonical momenta in terms of the velocities. This can also be seen because $D_a D_b \mathcal{L}(\dot{x})\dot{x}^b = 0$. In other words, the Hessian of the Lagrangian has an eigenvector with eigenvalue zero. In this case, in order to perform the canonical quantization of the system, Dirac's procedure for constrained systems will be applied in what follows. The reader is referred to [49, 50, 51] for the theory of constrained systems.

We first compute the canonical momenta π of the particle

$$\pi_a = D_a\mathcal{L}(\dot{x}) = -mP^*(\dot{x})^{1/\deg P} L_a^{-1}(\dot{x}) - eA_a\,, \tag{8.16}$$

where L_a^{-1} are the components of the inverse Legendre map given in an observer frame $\epsilon^a = \{\epsilon^0, \epsilon^\alpha\}$. From this equation, we conclude that $\pi + eA = -mP^*(\dot{x})^{1/\deg P} L^{-1}(\dot{x})$ is constrained to be contained in $-\zeta C(P)$ (otherwise the Legendre map is not even defined), where ζ is a parameter given by $\zeta = \mathrm{sign}(\dot{x}^0) = -\mathrm{sign}(\pi_0 + A_0)$ encoding whether $\pi + A$ lies in $C(P)$ or in $-C(P)$. As a consequence of the equation $P^*(\dot{x})P(L^{-1}(\dot{x})) = 1$, which relates P and P^*, and the definition of π, it follows now that

$$P(\pi + eA) - m^{\deg P} = 0\,, \quad \text{with } \pi + eA \in C(P) \cup -C(P)\,, \tag{8.17}$$

which is a primary constraint for the system. We now notice that

(1) if $\pi + eA \in -C(P)$, then $\pi_0 + eA_0 < 0$ and $\dot{x}^0 > 0$. In this case, the constraint (8.17) can be written as

$$\pi_0 + eA_0 - \omega_-(\vec{\pi} + e\vec{A}) = 0\,, \tag{8.18}$$

where $\omega_-(\vec{\pi} + e\vec{A}) < 0$ is the solution of (8.17) lying in $-C(P)$, i.e.,

$$P(\omega_-(\vec{\pi} + e\vec{A}), \vec{\pi} + e\vec{A}) - m^{\deg P} = 0\,.$$

(2) if $\pi + eA \in C(P)$, then $\pi_0 + eA_0 > 0$ and $\dot{x}^0 < 0$. In this case, the constraint (8.17) is written as

$$\pi_0 + eA_0 - \omega_+(\vec{\pi} + e\vec{A}) = 0\,, \tag{8.19}$$

[3] Euler's theorem states that if a function f is homogeneous of degree s, i.e. $f(\tau x) = \tau^s f(x)$, then $D_x f.x = sf$.

where $\omega_+(\vec{\pi}+e\vec{A})$ is the solution of (8.17) lying in $C(P)$.

Although non-covariant, the expressions (8.18) and (8.19) have the advantage that they explicitly show the constraint of $\pi + eA$ to lie in $C(P) \cup -C(P)$.

We can now encode both constraints (8.18) and (8.19) in a single equation by noticing that ω_+ and ω_- are related by $\omega_+(\vec{\pi}+e\vec{A}) = -\omega_-(-\vec{\pi}-e\vec{A})$. This comes from the even degree of homogeneity of the cotangent bundle function P because if $\omega(\vec{p})$ is a solution of $P(\omega(\vec{p}),\vec{p}) - m^{\deg P} = 0$, by changing $\vec{p} \to -\vec{p}$ we obtain $P(\omega(-\vec{p}),-\vec{p}) - m^{\deg P} = P(-\omega(-\vec{p}),\vec{p}) - m^{\deg P}$. Thus we conclude that if $\omega(\vec{p})$ is a solution, $-\omega(-\vec{p})$ is also a solution. Moreover, we know from the energy-distinguishing condition that P must be of even degree. But for a hyperbolic polynomial of even degree $\deg P$ there are only two solutions of $P(p) - m^{\deg P} = 0$ which are real for all $\vec{p}$, so that it follows that $\omega_+(\vec{\pi}+e\vec{A}) = -\omega_-(-\vec{\pi}-e\vec{A})$. Taking now the notation $\omega_+(\vec{\pi}+e\vec{A}) = \omega(\vec{\pi}+e\vec{A}) > 0$ and using the above defined parameter ζ, we finally encode the constraints (8.18) and (8.19) in the single equation

$$\phi_{(1)} = \zeta(\pi_0 + eA_0) + \omega(-\zeta(\vec{\pi}+q\vec{A})) = 0\,. \tag{8.20}$$

Following now Dirac's method [50, 51, 52], the Hamiltonian H of the system is a pure constraint one given by $H = \lambda\,\phi_{(1)}$, where λ is an undetermined Lagrange multiplier. From this expression for the Hamiltonian, it follows that $\dot{\phi}_{(1)} = \{H, \phi_{(1)}\} = 0$, so that there are no further constraints and therefore λ cannot be determined from the requirement $\dot{\phi}_1 = 0$. In Dirac's language, we say that $\phi_{(1)}$ is a first class constraint, so that in order to find the dynamical evolution of the system in phase space it is necessary to fix the gauge freedom that we have in the action. In order to fix the gauge, we use the freedom in the action to fix the scale of τ and eliminate all arbitrariness in the Lagrange multiplier λ. We thus take

$$\phi_{(2)} = x^0 - \zeta\tau \tag{8.21}$$

as a time dependent second constraint fixing the gauge, so that we find that

$$C_{IJ} = \{\phi_I, \phi_J\} = \begin{bmatrix} 0 & -\zeta \\ \zeta & 0 \end{bmatrix}\,, \text{ where } I, J = 1, 2\,,$$

is now invertible, turning the system into a second class one with time dependent constraints. This is the case because now the Hamiltonian is $H = \lambda^I \phi_{(I)}$ (with $I = 1, 2$), such that the requirement that the constraints $\phi_{(I)}$ be preserved in time implies that

$$\dot{\phi}_{(I)} = \{\phi_{(I)}, H\} + \partial_\tau \phi_{(I)} = C_{IJ}\lambda^J + \partial_\tau \phi_{(I)} = 0,$$

and since C_{IJ} is invertible, one obtains

$$\lambda^I = -C^{IJ}\partial_\tau \phi_{(J)}\,,$$

where $C^{IJ}C_{JK} = \delta^I_K$. This equation fixes the Lagrange multipliers λ^I and leaves the parameter ζ undetermined. This way we can simultaneously deal with the both theories labelled by ζ. In fact, one finds $\lambda^{(1)} = \lambda = 1$ and $\lambda^{(2)} = 0$.

The equations of motion are then found to be

$$\dot{x}^0 = \zeta = \pm 1\,, \quad \dot{x}^\alpha = \frac{\partial\omega(-\zeta(\vec{\pi}+e\vec{A}))}{\partial\pi_\alpha}\,, \quad \dot{\pi}_a = -e\zeta\partial_a A_0 - \partial_a\omega(-\zeta(\vec{\pi}+e\vec{A}))\,,$$

together with the constraint equations $\phi_{(I)} = 0$. Moreover, since π_0 and x^0 can be obtained from the constraint equations $\phi_{(I)} = 0$, only $\eta = (\vec{x}, \vec{\pi})$ are independent variables. The equations of motion for the independent variables can thus be obtained from the effective Hamiltonian $H^\zeta_{eff} = e\zeta A_0 + \omega(-\zeta(\vec{\pi}+e\vec{A}))$ as

$$\dot{\eta} = \{\eta, H^\zeta_{eff}\}\,. \tag{8.22}$$

These equations do not involve neither time dependence explicitly (which at the quantum level ensures unitary time evolution [52]) nor constraints.

We are now ready to perform the canonical quantization. We first promote the classical variables $(\vec{x}, \vec{\pi})$ to operators $(\hat{\vec{x}}, \hat{\vec{\pi}})$ acting on a Hilbert space $\mathcal{H}^\zeta$, whose elements are assumed to be L^2-integrable functions ϕ with standard L^2 scalar product

$$(\psi, \phi) = \int d^3x\, \psi^\dagger \phi\,. \tag{8.23}$$

The basic commutators are defined through the Poisson brackets as

$$[\hat{x}^a, \hat{\pi}_b] = i\delta^a_b\,. \tag{8.24}$$

We take the following standard representation for the basic operators acting in $\mathcal{H}^\zeta$

$$\hat{x}^a = x^a, \quad \text{and} \quad \hat{\pi}_a = -i\partial_a\,. \tag{8.25}$$

Clearly, these basic operators are Hermitian with respect to the defined scalar product (8.23) if we define the adjoint of an operator by simply taking its complex conjugate.

The dynamics of the physical states $\Phi^\zeta \in \mathcal{H}^\zeta$ is provided by the Schrödinger equation

$$i\partial_\tau \Phi^\zeta = \hat{H}^\zeta_{eff}\Phi^\zeta = \Big(e\zeta A_0 + \omega(-\zeta(-i\vec{\partial}+e\vec{A}))\Big)\,\Phi^\zeta\,,$$

which, following [63] and using the definition of $\phi_{(2)}$, is given in terms of the physical time x^0 as

$$i\partial_0\Phi^\zeta = \zeta\hat{H}^\zeta_{eff}\Phi^\zeta = \Big(eA_0 + \zeta\omega(-\zeta(-i\vec{\partial}+e\vec{A}))\Big)\,\Phi^\zeta\,. \tag{8.26}$$

The operator $\omega(-\zeta(-i\vec{\partial}+e\vec{A}))$ is defined by its classical Taylor expansion as follows; since at the classical level $\omega(\vec{P})\Big|_{\vec{P}=-\zeta(\vec{\pi}+e\vec{A})}$ comprises the negative and positive real solutions of $P(p) - m^{\deg P} = 0$, it admits a Taylor expansion, in particular, as

$$\omega(\vec{P})\Big|_{\vec{P}=-\zeta(\vec{\pi}+e\vec{A})} = \sum_{n_1=0}^{\infty}\cdots\sum_{n_d=0}^{\infty}\frac{C_{n_1\cdots n_d}}{n_1!\cdots n_d!}(-\zeta)^{n_1+\cdots+n_d}(\pi_1+e\vec{A}_1)^{n_1}\cdots(\pi_d+e\vec{A}_d)^{n_d}\,, \tag{8.27}$$

where $d = \dim M - 1$ and $C_{n_1\cdots n_d}$ are totally symmetric real quantities given by

$$C_{n_1\cdots n_d} = \frac{\partial^{n_1+\cdots+n_d}\omega(\vec{P})}{\partial P_1^{n_1}\cdots\partial P_d^{n_d}}\Bigg|_{\vec{P}=0}\,.$$

Hence, the operator $\omega(-\zeta(-i\vec{\partial}+e\vec{A}))$ is given at the quantum level as

$$\omega(-\zeta(-i\vec{\partial}+e\vec{A})) = \sum_{n_1=0}^{\infty}\cdots\sum_{n_d=0}^{\infty}\frac{C_{n_1\cdots n_d}}{n_1!\cdots n_d!}(-\zeta)^{n_1+\cdots+n_d}(-i\partial_1+e\vec{A}_1)^{n_1}\cdots(-i\partial_d+e\vec{A}_d)^{n_d}\,,$$

which is Hermitian with respect to the defined scalar product if we take its adjoint as the complex conjugate of the operator. This ensures that the complete Hamiltonian in (8.26) is itself Hermitian. But this implies that (Ψ, Φ) is conserved under time evolution if Ψ and Φ are solutions of the Schrödinger equation (8.26) in their respective Hilbert spaces $\mathcal{H}^\zeta$.

Moreover, writing the explicit values of ζ in equation (8.26), we find that the time evolution for the physical states ϕ^+ of $\mathcal{H}^+$ is given by the equation

$$i\,\partial_0\phi^+(x) \quad = \quad \left(eA_0 + \omega(i\vec{\partial} - e\vec{A})\right)\phi^+(x)\,, \tag{8.28}$$

and that the time evolution for the physical states ϕ^- of $\mathcal{H}^-$ is given by the equation

$$i\,\partial_0\phi^-(x) \quad = \quad \left(eA_0 - \omega(-i\vec{\partial} + e\vec{A})\right)\phi^-(x)\,. \tag{8.29}$$

Taking the complex conjugate of the last equation, we have

$$i\,\partial_0(\phi^-)^*(x) = \left(-eA_0 + \omega(i\vec{\partial} - (-e)\vec{A})\right)(\phi^-)^*(x)\,.$$

From this equation, we see that the operation of taking the complex conjugate of a physical state in $\mathcal{H}^-$ describing a particle of charge e produces a physical state of $\mathcal{H}^+$ describing a particle of charge $-e$. In other words, $(\phi^-)^*$ describes an antiparticle in $\mathcal{H}^+$ precisely as in the standard metric description. We thus see that the particle-antiparticle interpretation holds in any hyperbolic, time-orientable and energy-distinguishing spacetime.

If the massive particle is free, that is, when the gauge potential is switched off, it is simple to find a solution of equation (8.28). Indeed, in this case, the equation for ϕ^+ (the solution ϕ^- is obtained by simply taking the complex conjugate of ϕ^+) is reduced to

$$i\,\partial_0\phi^+(x) = \omega(i\vec{\partial})\,\phi^+(x)\,, \tag{8.30}$$

and it is straightforward to check that

$$\phi^+(x) = \int \frac{d\vec{p}}{f(\vec{p})}\,a^+(\vec{p})\,e^{-i\omega(\vec{p})t - i\vec{x}.\vec{p}}\,, \tag{8.31}$$

with $f(\vec{p})$ an arbitrary function of $\vec{p}$, $a^+(\vec{p}) \in \mathbb{C}$ and $\omega(\vec{p})$ a solution of $P(\omega(\vec{p}), \vec{p}) - m^{\deg P} = 0$, is a solution of (8.30). By construction, $\phi^{(\pm)}$ also satisfy

$$\left(P(i\partial_0, i\vec{\partial}) - m^{\deg P}\right)\phi^{(\pm)}(x) = 0\,, \tag{8.32}$$

which can be thought of as a generized Klein-Gordon equation for a $\deg P$ spacetime geometry. We have thus studied the canonical quantization of a charged point particle coupled to the electromagnetic gauge field A on a hyperbolic, time-orientable and energy-distinguishing spacetime, which assumption has been crucial to perform the energy split which ultimately led us to the Schödinger equation (8.26). This also allowed us to show the particle-antiparticle interpretation and to find an explicit solution for the wave function of a free massive particle.

In the next section, we will study how to generate linear field equations whose solutions satisfy (8.32), but can be more special[4].

[4] On Lorentzian spacetimes, solutions of the Dirac equation satisfy the Klein-Gordon equation, but are more special.

8.3. General structure of lower order field equations

In this section, we wish to study lower derivative order incarnations of the field equations (8.32). In particular, their plane wave solutions have to satisfy the dispersion relations encoded in the cotangent bundle function P. We find that one can only obtain field equations of a derivative order that divides the degree of the polynomial P. From this insight, we can then derive the algebraic constraints on the highest derivative coefficients of any lower degree field equation, and further conditions if the field equations are to follow from an action principle.

We first recall the massive dispersion relation

$$P(q) - m^{\deg P} = P^{a_1 a_2 \cdots a_{\deg P}} q_{a_1} \cdots q_{a_{\deg P}} - m^{\deg P} = 0\,, \tag{8.33}$$

where q is a covector field and $P^{a_1 a_2 \dots a_{\deg P}}$ is the polarization tensor of the cotangent bundle function P (see equation (4.8)). We then want to generate linear field equations whose plane wave solutions $e^{-iq.x}$ satisfy the dispersion relation. This is meant in the sense that the covector q in any plane wave solution satisfies $P(q) - m^{\deg P}$=0.

Field equations of maximal derivative order. In order to reproduce the dispersion relation (8.33), these field equations take the form

$$\left[i^{\deg P} P^{a_1 \cdots a_{\deg P}} \partial_{a_1} \cdots \partial_{a_{\deg P}} - m^{\deg P}\right] \phi = 0\,, \tag{8.34}$$

which determines the dynamical evolution of a scalar field $\phi : M \to \mathbb{R}$. This is so because obviously $\phi(q) = e^{-i\,q.x}$ is a solution of this field equation if only if q satisfies the massive dispersion relation. In this case, it is also very easy to find, by using the Euler-Lagrange equations

$$\sum_{l=0}^{\deg P/2} (-1)^l \partial_{a_1} \cdots \partial_{a_l} \left[\frac{\partial^l \mathcal{L}}{\partial\left(\partial_{a_1} \cdots \partial_{a_l} \phi\right)}\right] = 0\,, \tag{8.35}$$

that a possible Lagrangian density giving rise to the above field equations is

$$\mathcal{L} = (-1)^{-\deg P/2}\, i^{\deg P}\, P^{a_1 \cdots a_{\deg P}} \frac{1}{2} (\partial_{a_1} \cdots \partial_{a_{\deg P/2}} \phi)(\partial_{a_{\deg P/2+1}} \cdots \partial_{a_{\deg P}} \phi) - \frac{1}{2} m^{\deg P} \phi^2\,.$$

For the case of a complex scalar field, one can take the real Lagrangian

$$\mathcal{L} = (-1)^{-\deg P/2}\, i^{\deg P}\, P^{a_1 \cdots a_{\deg P}} \frac{1}{2} (\partial_{a_1} \cdots \partial_{a_{\deg P/2}} \phi^*)(\partial_{a_{\deg P/2+1}} \cdots \partial_{a_{\deg P}} \phi) - \frac{1}{2} m^{\deg P} \phi^* \phi\,,$$

so that both, ϕ and ϕ^*, satisfy the scalar field equation (8.34).

If we now consider that ϕ and ϕ' are solutions of equation (8.34), one obtains that the quantity

$$\begin{aligned} J^c(\phi, \phi') \quad = \quad i \sum_{l=0}^{\deg P/2-1} & (-1)^l P^{c a_1 \cdots a_{\deg P-1}} \times \\ & \left[\partial_{a_1} \cdots \partial_{a_l} \phi^* \partial_{a_{l+1}} \cdots \partial_{a_{\deg P-1}} \phi' - \partial_{a_1} \cdots \partial_{a_l} \phi' \partial_{a_{l+1}} \cdots \partial_{a_{\deg P-1}} \phi^*\right] \end{aligned} \tag{8.36}$$

satisfies the continuity equation $\partial_a J^a = 0$. This means that

$$(\phi, \phi') = \int d^{\dim M-1} x\, J^0(\phi, \phi') \tag{8.37}$$

is a conserved quantity and can therefore be used as scalar product on the space of solutions. To see how the normalization proceeds, we consider two arbitrary solutions, $\phi_p = e^{ip.x}$ and $\phi'_{p'} = e^{ip'.x}$, of the field equations (8.34), where p_0 and p'_0 may equally correspond to a real or complex solution of the dispersion relation (8.33). Introducing these solutions into the scalar product (8.37) we find

$$\begin{aligned} (\phi_p, \phi'_{p'}) &= (2\pi)^{\dim M-1}\delta^{\dim M-1}(\vec{p}-\vec{p}')e^{i(p_0-p'_0)}x^0 \times \\ & \sum_{l=0}^{\dim P/2-1} P^{0a_1\cdots a_{\deg P-1}}\left[p^*_{a_1}\cdots p^*_{a_l}p'_{a_{l+1}}\cdots p'_{a_{\deg P-1}} + p'_{a_1}\cdots p'_{a_l}p^*_{a_{l+1}}\cdots p^*_{a_{\deg P-1}}\right]. \end{aligned} \tag{8.38}$$

Moreover, we also obtain that

$$0 = \partial_0(\phi_p, \phi'_{p'}) = i(p_0 - p'^*_0)(\phi_p, \phi'_{p'}). \tag{8.39}$$

Combining both equations, (8.38) and (8.39), we finally find

$$(\phi_p, \phi'_{p'}) = \begin{cases} (2\pi)^{\dim M-1}\delta^{\dim M-1}(\vec{p}-\vec{p}')\nabla^0 P(p) & \text{if} \quad p_0 = p^{*\prime}_0 \\ 0 \quad \text{otherwise} \end{cases} \tag{8.40}$$

We can thus use this inner product in order to normalize the real wave functions $\phi^+(x)$ for a free particle found in equation (8.31) by taking

$$\phi^+_p(x) = \frac{1}{\sqrt{\nabla^0 P(\omega(\vec{p}), \vec{p})}}\, e^{-i\omega(\vec{p})t - i\vec{x}.\vec{p}} \tag{8.41}$$

in the free particle wave function (8.31)

$$\phi^+(x) = \int d\vec{p}\, a^+(\vec{p})\, \phi^+_p(x).$$

General construction of field equations of reduced derivative order. Let us now look for linear field equations satisfying the dispersion relation $P(q) - m^{\deg P} = 0$ but of derivative order less than $\deg P$. For this purpose, we consider the following differential equation

$$(\tilde{P}^{a_1\ldots a_r}\partial_{a_1}\cdots\partial_{a_r} - m^r)\Phi = 0 \tag{8.42}$$

of degree $r < \deg P$. For Φ a scalar quantity, the plane wave solutions of this equation satisfy $\tilde{P}^{a_1\ldots a_r}p_{a_1}\cdots p_{a_r} - m^r$, which are unable to recover the massive dispersion relation. Hence, in order to have matter field equations of reduced order, we have to consider equations of the type (8.42) with $\Phi : M \to V$ taking values in some finite-dimensional complex vector space V, so that we may consider equations of the type

$$\left[i^r\, (\Gamma^M{}_N)^{a_1\cdots a_r}\partial_{a_1}\cdots\partial_{a_r} - m^r \mathbb{I}^M{}_N\right]\Phi^N = 0\,. \tag{8.43}$$

Clearly, plane wave solutions $\Phi^N_q(x) = \tilde{\Phi}^N\, e^{-i\,q.x}$ of this equation with non-trivial $\tilde{\Phi}^N$ satisfy the massive dispersion relation if and only if the equation

$$\det(\Gamma(q) - m^r\mathbb{I}) = c\,(P(q) - m^{\deg P})^\alpha \tag{8.44}$$

is satisfied for some positive integer α. In this equation, $\Gamma(q) = (A_{MN})^{a_1\cdots a_r}q_{a_1}\cdots q_{a_r}$, $\mathbb{I}$ is the identity matrix on V and c is an arbitrary constant proportionality factor. Using now the

Hamilton-Cayley theorem[5], it follows that if the above equation is satisfied, the matrix $\Gamma(q)$ must be such that it satisfies its own characteristic equation. More precisely, the equation

$$\left[\Gamma^{\frac{\deg P}{r}}(q) - P(q)\mathbb{I}\right]^{\alpha} = 0 \tag{8.45}$$

must be satisfied. This equation clearly implies that the matrix $\Gamma^{\frac{\deg P}{r}}(q) - P(q)\mathbb{I}$ must be nilpotent, which condition is found to be equivalent[6] to

$$\Gamma^{\frac{\deg P}{r}}(q) = P(q)\mathbb{I}\,. \tag{8.46}$$

Thus this equation is a necessary condition for equation (8.44) to hold. This equation also gives the restriction that the allowed field equations of reduced order r are those for which $\deg P/r$ is an integer. For instance, for $\deg P = 2$ one can only take $r = 1, 2$, which correspond to the Dirac and Klein-Gordon equations, for $\deg P = 4$ one can take $r = 1, 2, 4$, for $\deg P = 6$ one can take $r = 1, 2, 3, 6$, and so on. The following proposition will now provide sufficient conditions for the matrices $\Gamma(q)$ satisfying equation (8.46) to satisfy equation (8.44).

PROPOSITION 8.3.1. *Let* $\Gamma(q) \in GL(\dim V, \mathbb{C})$ *and* $\deg P/r$ *a positive integer. Then*

$$\det(\Gamma(q) - m^r\,\mathbb{I}) = (-1)^{\dim V}\left[m^{\deg P} - (-1)^{\deg P/r} P(q)\right]^{\frac{r\,\dim V}{\deg P}} \tag{8.47}$$

if and only if

$$\mathrm{Tr}(\Gamma(q)) = 0, \cdots, \mathrm{Tr}(\Gamma^{(\deg P/r)-1}(q)) = 0 \text{ and } \Gamma^{\deg P/r}(q) = P(q)$$

Proof. Assume $\mathrm{Tr}(\Gamma(q)) = 0, \cdots, \mathrm{Tr}(\Gamma^{\deg P/r-1}(q)) = 0$ and $\Gamma^{\deg P/r}(q) = P(q)\mathbb{I}$. By using the identity

$$\det(\mathbb{I} + \Gamma(q)) = \sum_{k=0}^{\infty} \frac{1}{k!}\left[\mathrm{Tr}\left(\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j}\Gamma^j(q)\right)\right]^k, \tag{8.48}$$

it immediately follows that

$$\det(\mathbb{I} + \mu\Gamma(q)) = \left[1 - \mu^{\deg P/r}(-1)^{\deg P/r} P(q)\right]. \tag{8.49}$$

The result follows by setting $\mu = (-m^r)^{-1}$. Conversely, assuming now that equation (8.47) holds and using again identity (8.48), it follows that $\mathrm{Tr}(\Gamma(q)) = 0, \cdots, \mathrm{Tr}(\Gamma^{\deg P/r-1}(q)) = 0$ and that

$$\mathrm{Tr}\left[\left(\frac{\Gamma^{\deg P/r}(q)}{P}\right)^k\right] = \dim V \text{ for all } k = 0, 1, \ldots.$$

But the only matrix satisfying the last equation is the identity matrix, which implies that $\Gamma^{\deg P/r}(q) = P(q)\,\mathbb{I}$, thus proving the proposition.

So this theorem indeed provides necessary and sufficient conditions that the matrices $(A^M{}_N)^{a_1\cdots a_r}$ must satisfy for equations (8.43) to have plane wave solutions satisfying the massive dispersion relation. However, additionally to the generation of equations (8.43), we also want these equations

[5]The Hamilton-Cayley theorem states that every endomorphism on a finite-dimensional vector space satisfies its own characteristic equation.

[6]This is shown by using that a matrix A is nilpotent if and only if $\mathrm{Tr}(A^k) = 0$ for all positive integers k.

to be derived from an action $S[\Phi] = \int d^{\dim M}x\, \mathcal{L}[\Phi]$ with real Lagrangian density $\mathcal{L}$ because then, for instance, one can introduce interaction terms by using the symmetries of the Lagrangian and directly obtain conserved quantities by using Noether's theorem. The following proposition provides sufficient conditions in order to derive the considered equations from an action principle.

PROPOSITION 8.3.2. *The set of equations (8.43) is derived from the scalar action functional*

$$S[\Phi, \bar{\Phi}] = \int d^{\dim M}x\, \bar{\Phi}\, [i^r\, (\Gamma)^{a_1\cdots a_r} \partial_{a_1} \cdots \partial_{a_r} - m^r]\, \Phi\,, \tag{8.50}$$

where $\bar{\Phi}_N = \Phi^\dagger_M \Gamma^M{}_N$ and Γ is a matrix on the representation space V, if the matrix Γ is such that

$$(\Gamma^\dagger)^{-1}(\Gamma^{a_1\cdots a_r})^\dagger \Gamma^\dagger = \Gamma^{a_1\cdots a_r}\,. \tag{8.51}$$

Moreover, the action is real if in addition the matrix Γ is Hermitian, i.e., $\Gamma^\dagger = \Gamma$.

Proof. Variation of the action above with respect to $\bar{\Phi}$ trivially reproduces equations (8.43). But we now have to make sure that variation of the action with respect to Φ also gives rise to the same field equations. But we obtain

$$(\delta_\Phi \mathcal{L})^\dagger = 0 \qquad \Rightarrow \qquad (i)^r (\Gamma^{a_1\cdots a_r})^\dagger \Gamma^\dagger \partial_{a_1} \cdots \partial_{a_r} \Phi - m^r \Gamma^\dagger \Phi = 0\,.$$

Multiplying the last expression by $(\Gamma^\dagger)^{-1}$, we indeed obtain that equation (8.51) must hold. Concerning the reality of the action, using integration by parts and assuming that equation (8.51) is satisfied, we obtain

$$S^\dagger[\Phi, \bar{\Phi}] = \int d^{\dim M}x\, \Phi^\dagger \Gamma^\dagger\, [i^r\, (\Gamma)^{a_1\cdots a_r} \partial_{a_1} \cdots \partial_{a_r} - m^r]\, \Phi\,.$$

Thus, $S^\dagger[\Phi, \bar{\Phi}] = S[\Phi, \bar{\Phi}]$ if $\Gamma = \Gamma^\dagger$, which proves the proposition.

Clearly, the Lagrangian in the action (8.50) is invariant under the $U(1)$ global transformation $\Phi \to e^{i\alpha}\Phi$, where α is the parameter of the transformation. In this case, we can more efficiently find a conserved quantity in the space of solutions (which can therefore be used to normalize them) by applying Noether's theorem to the underlying $U(1)$ symmetry of the Lagrangian. We thus find the current

$$\begin{aligned} J^c(\Phi) &= i\,\alpha \sum_{l=0}^{r-1} (-1)^l \left[\partial_{a_1} \cdots \partial_{a_l} \left(\frac{\partial \mathcal{L}}{\partial(\partial_c \partial_{a_1} \cdots \partial_{a_{r-1}} \Phi)}\right) \partial_{a_{l+1}} \cdots \partial_{a_{r-1}} \Phi\right] \\ &= i^{r+1}\,\alpha \sum_{l=0}^{r-1} (-1)^l \left[\partial_{a_1} \cdots \partial_{a_l} \bar{\Phi}\ (A^{ca_1\cdots a_{r-1}})\, \partial_{a_{l+1}} \cdots \partial_{a_{r-1}} \Phi\right] \end{aligned} \tag{8.52}$$

which satisfies the continuity equation $\partial_c J^c(\Phi) = 0$. This implies that the charge

$$Q(\Phi) = \int d^{\dim M - 1}x\, J^0(\Phi)$$

associated with the $U(1)$ symmetry is a conserved quantity.

We will now focus our investigation on finding a generalized Dirac equation, that is, a first order field equation

$$\left[i\, (\gamma^M{}_N)^a \partial_a - m \mathbb{I}^M{}_N\right] \Phi^N = 0\,, \tag{8.53}$$

which corresponds to the case $r = 1$ in equation (8.43). Thus, from proposition 8.3.1, if the plane wave solutions of this generalized Dirac equation are to satisfy the massive dispersion relation $P(q) - m^{\deg P} = 0$, the matrix $\gamma(q) = \gamma^a q_a$ must be constructed such that it satisfies the main condition

$$(\gamma^a q_a)^{\deg P} = P(q)\mathbb{I}\,,$$

besides the $(\deg P/r) - 1$ suplementary trace conditions

$$\mathrm{Tr}(\gamma(q)) = 0, \cdots, \mathrm{Tr}(\gamma^{(\deg P/r)-1}(q)) = 0\,. \tag{8.54}$$

The main condition can be equivalently written, by using the polarization tensor of P (4.8), as

$$\gamma^{(a_1}\gamma^{a_2}\cdots\gamma^{a_{\deg P})} = P^{a_1 a_2 \cdots a_{\deg P}}\mathbb{I}\,, \tag{8.55}$$

where $\mathbb{I}$ is the unit matrix in the representation space in which the field Φ takes values. Furthermore, this expression can be thought of as the $(\deg P)$–ary generalization[7] of the bin–ary Dirac algebra corresponding to a physical dispersion relation of degree $\deg P$. Once matrices $\gamma^a q_a$ satisfying the main condition (8.55) have been found, one still has to check that the trace conditions (8.54) are satisfied.

In the next section, we present concrete examples for lower order field equations with higher order dispersion relations.

8.4. Examples of second and first order field equations

We now present some examples of how to use the results of the previous section in order to generate field equations of reduced order.

Second order field equation for four dimensional area metric spacetimes. We now consider the subclass of class I of area metrics given in equation (6.11) with $\sigma_1 = \sigma_2 = \sigma_3$, which gives rise to the irreducible quartic massive dispersion relation $P(q) - m^4 = 0$ with $P(q)$ given in equation (7.30). Since $\deg P = 4$, we can consider equations of reduced degree $r = 1, 2$. Here we want to study the case $r = 2$. So we first want to construct ten matrices γ^{ab} labelled by the symmetric pair of indices (ab), such that the plane wave solutions of the linear field equations

$$\left[i^2\gamma^{ab}\partial_a\partial_b - m^2\right]\Phi = 0 \tag{8.56}$$

satisfy the considered massive quartic dispersion relation. According to proposition 8.3.1, we then need that the matrices γ^{ab} satisfy $(\gamma^{ab}q_a q_b)^2 = P(q)\mathbb{I}$, or equivalently

$$\gamma^{(ab}\gamma^{cd)} = P^{abcd}\mathbb{I} \tag{8.57}$$

as the main condition, and

$$\mathrm{Tr}(\gamma^{ab}q_a q_b) = 0 \tag{8.58}$$

[7]The algebra (8.55) for a first order field equation has previously been discussed [24], however, without identification of sufficient conditions for these algebra to reproduce a massive dispersion relation and without the determination of the conditions to derived the first order field equation from an action principle.

as the $(\deg P/r) - 1 = 1$ suplementary condition. It is then a simple exercise for *Mathematica* to show that the matrices (where $f(G)$ is given in equation (6.10))

$$\begin{aligned}
&\gamma^{00} = if(G)(\tau_1\tau_2\tau_3)^{1/2}\Gamma^0 && \gamma^{23} = -1/2f(G)\left(\tau_1(\tau_2^2+\tau_3^2)\right)^{1/2}\Gamma^8\\
&\gamma^{01} = 1/2if(G)\left(\tau_1(\tau_2^2+\tau_3^2)\right)^{1/2}\Gamma^1 && \gamma^{31} = -1/2f(G)\left(\tau_2(\tau_1^2+\tau_3^2)\right)^{1/2}\Gamma^6\\
&\gamma^{02} = 1/2if(G)\left(\tau_2(\tau_1^2+\tau_3^2)\right)^{1/2}\Gamma^2 && \gamma^{12} = -1/2if(G)\left(\tau_3(\tau_1^2+\tau_2^2)\right)^{1/2}\Gamma^5\\
&\gamma^{03} = 1/2if(G)\left(\tau_3(\tau_1^2+\tau_2^2)\right)^{1/2}\Gamma^3 && \gamma^{22} = -f(G)\,(\tau_1\tau_2\tau_3)^{1/2}\,\Gamma^7\\
&\gamma^{11} = -f(G)\,(\tau_1\tau_2\tau_3)^{1/2}\,\Gamma^4 && \gamma^{33} = -f(G)\,(\tau_1\tau_2\tau_3)^{1/2}\,\Gamma^9
\end{aligned}$$

indeed satisfy the condition (8.57) as well as the trace condition (8.58). In this expression, the ten matrices Γ^I provide a 32-dimensional irreducible representation of the Dirac algebra in ten dimensions, i.e., $\Gamma^{(I}\Gamma^{J)} = \eta^{IJ}$ with $\eta^{IJ} = \mathrm{diag}(-1, 1, \dots, 1)$[8]. The matrix Γ, which in this case satisfies the conditions of proposition 8.3.2 in order to derive equations (8.56) from a real action functional, is found to be

$$\Gamma = i2\sqrt{2}\gamma^{00}\gamma^{11}\gamma^{12}\gamma^{13}\gamma^{22}\gamma^{23}\gamma^{33}\,.$$

First order field equation on Lorentzian spacetime. Here we wish to construct a first order field equation with plane wave solutions satisfying the quadratic metric massive dispersion relation $\eta^{ab}q_aq_b - m^2 = 0$ with $\eta = \mathrm{diag}(1, -1, -1, -1)$ being the Minkowski metric in four dimensions. In this case, we have that $\deg P = 2$. Thus, according to the expression (8.55), we need to find matrices $\gamma(q) = \gamma^a q_a$ satisfying

$$\gamma^{(a}\gamma^{b)} = \eta^{ab} \quad \text{and} \quad \mathrm{Tr}(\gamma^a q_a) = 0.$$

This is of course the standard Dirac algebra and the reader already knows the solution; the matrices γ^a are the Dirac gamma matrices (in the Weyl representation)

$$\gamma^a = \begin{bmatrix} 0 & \sigma^a \\ \bar{\sigma}^a & 0 \end{bmatrix}, \tag{8.59}$$

where we have labelled $\sigma^a = (\mathbb{I}_{2\times2}, \sigma^\alpha)$ and $\bar{\sigma}^a = (\mathbb{I}_{2\times2}, -\sigma^\alpha)$, with σ^α the Pauli matrices satisfying

$$\sigma^{(a}\bar{\sigma}^{b)} = \bar{\sigma}^{(a}\sigma^{b)} = \eta^{ab}\mathbb{I}_{2\times2}\,. \tag{8.60}$$

Thus the Dirac matrices γ^a trivially satisfy the conditions of proposition (8.3.1). In this case, it is found that the Γ matrix satisfying the conditions of proposition 8.3.2 is $\Gamma = \gamma^0$, which also satisfies $\gamma^0 = (\gamma^0)^{-1}$. Hence, we recover the metric Dirac equation as obtained from the real Lagrangian

$$\mathcal{L} = i\bar{\Phi}\gamma^a\partial_a\Phi - m\bar{\Phi}\Phi\,.$$

Using now the general current in expression (8.52), we obtain the conserved charge

$$Q(\Phi) = \int d^3x\, q\bar{\Phi}\gamma^0\Phi\,,$$

which is indeed used to normalize the solutions of the metric Dirac equation.

[8]For the explicit construction of the Γ^I matrices see appendix B of [64].

First order field equation on bimetric spacetimes and coupling to bimetric electrodynamics. We now wish to construct a Dirac equation for the considerably more complicated case of a quartic dispersion relation. More precisely, we consider the quartic massive dispersion relation

$$g^{-1}(q,q)h^{-1}(q,q) - m^4 = 0\,,$$

where g^{-1} and h^{-1} are Lorentzian metrics in four dimensions. Thus, in order to find a modified Dirac equation whose plane wave solutions satisfy this bimetric quartic dispersion relation, we have to find (according to the expression (8.55)) matrices Γ^a satisfying the quaternary algebra

$$\Gamma^{(a}\Gamma^b\Gamma^c\Gamma^{d)} = g^{(ab}h^{cd)}\mathbb{I}\,, \tag{8.61}$$

and the suplementary trace conditions

$$\mathrm{Tr}(\Gamma(q)) = \mathrm{Tr}(\Gamma^2(q)) = \mathrm{Tr}(\Gamma^3(q)) = 0\,. \tag{8.62}$$

Since the considered dispersion relation is constituted by the two Lorentzian metrics g^{-1} and h^{-1}, we know that there exist frames e and f such that

$$g^{ab}e^c_a e^d_b = \eta^{cd} \quad \text{and} \quad h^{ab}f^c_a f^d_b = \eta^{cd}\,, \tag{8.63}$$

where η is the Minkowski metric. In terms of these frames, it is a simple exercise to show that the 16×16 matrices

$$\Gamma^a = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & e^a_b\sigma^b & 0 & 0 & 0 & 0 \\ 0 & 0 & f^a_b\sigma^b & 0 & 0 & 0 & 0 & 0 \\ f^a_b\bar\sigma^b & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^a_b\bar\sigma^b & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & e^a_b\sigma^b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & f^a_b\sigma^b \\ 0 & 0 & 0 & 0 & 0 & f^a_b\bar\sigma^b & 0 & 0 \\ 0 & 0 & 0 & 0 & e^a_b\bar\sigma^b & 0 & 0 & 0 \end{array}\right]\,, \tag{8.64}$$

with σ^a and $\bar\sigma^a$ precisely defined as in equation (8.60), indeed satisfy the quartic algebra in equation (8.61) because they satisfy

$$(\Gamma^a q_a)^4 = g^{-1}(q,q)h^{-1}(q,q)\mathbb{I}_{16\times 16}\,, \tag{8.65}$$

and the trace conditions (8.62). Thus the modified bimetric Dirac equation is given by equation (8.43) with the above 16-dimensional Γ–matrices.

The reader can recognize that the two diagonal blocks in (8.64) in fact satisfy equations (8.61) and (8.62) separately. However, we need both blocks because only then one can find a matrix Γ satisfying the conditions of proposition 8.3.2. Indeed, it is easy to check that the matrix

$$\Gamma = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{8.66}$$

satisfies

$$\Gamma^\dagger = \Gamma, \quad \Gamma^{-1} = \Gamma, \quad \text{and} \quad \Gamma\,\Gamma^{a\dagger}\,\Gamma = \Gamma^a\,.$$

Defining then $\bar{\psi} = \psi^\dagger\Gamma$, the bimetric Dirac equation is derived from the real action

$$S[\psi] = \int d^4x\,\bar{\psi}\,[i\Gamma^a\partial_a - m]\,\psi \tag{8.67}$$

according to proposition 8.3.2. Had we taken only one of the blocks in equation (8.64), we would not have been able to write a Lagrangian, although we would have gotten a field equation satisfying the considered bimetric dispersion relation.
Moreover, according to expression (8.52), the charge $Q(\psi) = \int d^3x\,\bar{\psi}\Gamma^a\psi$ is conserved and can therefore be used to induce a scalar product in the space of solutions.

Finally, promoting now the global $U(1)$ symmetry of the action (which was used to derive the conserved charge) to a local one, one finds that the Lagrangian density must be

$$\mathcal{L} = \bar{\psi}\,[i\Gamma^a D_a - m]\,\psi\,,$$

where $D_a = \partial_a + ieA_a$ is a covariant derivative which makes the $U(1)$ symmetry global if A is a one-form gauge field. One then identifies e with a charge associated with the field ψ. Expanding the covariant derivative in the expression above, we find

$$\mathcal{L} = \bar{\psi}\,[i\Gamma^a\partial_a - m]\,\psi - e\bar{\psi}\Gamma^a\psi A_a\,.$$

We already know that area metric electrodynamics provides the most general gauge invariant action which gives rise to linear field equations for a one-form gauge field A. So restricting attention to area metrics G giving rise to bimetric dispersion relations $P_G = g^{-1}h^{-1}$ (which have been identified in [65]), we obtain the complete Lagrangian for bimetric area metric electrodynamics

$$\mathcal{L}(\psi, A) = \bar{\psi}\,[i\Gamma^a\partial_a - m]\,\psi - \frac{1}{8}\,G^{abcd}F_{ab}F_{cd} - e\bar{\psi}\Gamma^a\psi A_a\,,$$

which describes the interaction of the electromagnetic field A_a and the Dirac field ψ on a bimetric area metric spacetime. One might proceed with the quantization of this Lagrangian in order to study, for instance, refinements to quantum electrodynamics on hyperbolic, time-orientable and energy-distinguishing spacetimes.

CHAPTER 9

Towards gravity

In this chapter, we preview how building on the foundational work presented in this thesis, it could be shown in subsequent work [14, 15] *that the problem of finding modified gravitational dynamics for tensorial spacetimes that are predictive, time-orientable and energy-distinguishing reduces to solving a system of homogeneous linear partial differential equations. Thus the formidable physical problem of constructing alternative diffeomorphism-invariant gravity theories is reduced to a well-defined mathematical task.*

9.1. Deformation algebra of hypersurfaces.

The aim of this chapter is to preview how one can find the gravitational dynamics for a bi-hyperbolic and energy-distinguishing geometry (M, G), that is, how to determine the values of G throughout the manifold. More precisely, one wants to provide "adequate" geometric data on suitable initial value hypersurfaces for matter field dynamics and then evolve these geometric data to a neighbouring hypersurface, such that by repeating this procedure, one is able to reconstruct the geometry G on the entire manifold M. Since adequate geometric data are provided in terms of projections of the spacetime geometry to an initial data surface with the help of an embedding map for this hypersurface, one wants to study how functionals of this embedding map change under normal and tangential deformations of the hypersurface. A representation (loosely speaking) of the deformation operators on the geometric phase space (spanned by the geometric degrees of freedom on a hypersurface and canonically conjugate momenta) then yields first class Hamiltonian and diffeomorphism constraints whose sum determines the pure constraint Hamiltonian dynamics.

To execute this programme, one needs to derive the commutation algebra of these deformation operators, and this is where the findings of the present thesis play the pivotal rôle. So we consider $X : \Sigma \hookrightarrow M$ to be a smooth embedding map of a smooth manifold Σ of dimension $\dim M - 1$ with local coordinates $\{y^\alpha\}$ into the spacetime manifold M with coordinates $\{x^a\}$. Furthermore, we consider that the co-normals $n(y)$ at each point $X(y)$ on the hypersurface $X(\Sigma)$ are hyperbolic with respect to P, i.e., $n(y)$ is contained in the hyperbolicity cone $C_{X(y)}$ of P. Only such hypersurfaces can possibly carry initial data. Moreover, we also require that the co-normals $n(y)$ are normalized with respect to P as

$$P(n(y)) = 1\,. \tag{9.1}$$

Only thanks to the duality theory for massive particles, which we were only able to show to hold using the entire technology developed in this thesis, one can now use the Legendre map

$L : C_{X(y)} \to L(C_{X(y)})$ at each point $X(y)$ of the hypersurface $X(\Sigma)$ in order to define

$$T(y) := L_{X(y)}(n(y))$$

as the spacetime tangent vectors *normal* to the hypersurface $X(\Sigma)$. But then the $\dim M$ vectors

$$T(y) := L_{X(y)}(n(y)), \quad e_1(y) := \frac{\partial X^a(y)}{\partial y^1}\frac{\partial}{\partial x^a}, \quad \dots, \quad e_{\dim M-1}(y) := \frac{\partial X^a(y)}{\partial y^{\dim M-1}}\frac{\partial}{\partial x^a} \tag{9.2}$$

constitute a basis of each tangent space $T_{X(y)}M$ of M along the hypersurface Σ, where the co-normals $n(y)$ are uniquely determined by the conditions

$$n \in C, \qquad P(n) = 1, \qquad n(e_\alpha) = 0 \text{ for all } \alpha = 1, \dots, \dim M - 1\,.$$

The basis vectors $e_1(y), \cdots e_{\dim M-1}(y)$ are (spacetime) tangent vectors to $X(\Sigma)$. The uniquely determined dual basis to (9.2) then takes the form

$$n(y), \quad \epsilon^1(y), \quad \dots, \quad \epsilon^{\dim M-1}(y)\,. \tag{9.3}$$

Using the bases (9.2) and (9.3) we can now determined the projections of any tensorial geometry G to the hypersurface Σ. These are obtained by inserting the ϵ^α and n into the co-vector slots and e_α and T into the vector slots of the geometric tensor G. In order to obtain the independent geometric configuration variables which one can give dynamics to, one then needs to eliminate those contributions to the projections that are eliminated by the frame normalization conditions

$$P(n) = 1\,, \quad \text{and} \quad P(\epsilon^\alpha, n, \cdots, n) = 0\,.$$

This will result in a set of tensor fields $\hat{G}^A$ on Σ to which one adjoins canonically conjugate momenta $\hat{\pi}_A$. So we recognize that the geometry enters in three different ways into gravitational dynamics: (i) it identifies the admissible initial data hypersurfaces for matter field dynamics, which have hyperbolic co-normals with respect to P, (ii) it is used in order to calculate the normal vectors to admissible initial data hypersurfaces, and (iii) it provides the geometry degrees of freedom after eliminating those fixed by the frame normalization conditions.

The change of functionals F of the embedding map X under infinitesimal deformations of the hypersurface $X(\Sigma)$ is now studied by first implementing the deformation of the hypersurface $X(\Sigma)$ by considering a family of hypersurfaces X_t labelled by the parameter t such that $X_{t=0}$ recovers the undeformed hypersurface $X(\Sigma)$. One can then uniquely decompose the spacetime vector $\dot{X} = \partial_t X$ (connecting two hypersurfaces of the family X_t) along the undeformed hypersurface $X(\Sigma)$ as

$$\dot{X}^a(y) = N(y)T^a(y) + N^\alpha(y)e^a_\alpha(y)\,.$$

In this expression,

$$N(y) := n_a(y)(\dot{X}(y)) \qquad \text{and} \qquad N^\alpha(y) := \epsilon^\alpha(y)(\dot{X}(y)) \tag{9.4}$$

are the normal and tangential parts of the deformation, respectively. Any infinitesimal deformation $\delta X = dt\,\dot{X}_t$ is therefore completely parametrized by provision of a scalar field $N(y)$ and vector field $N^\alpha(y)e^a_\alpha(y)$.

The normal deformation operator acting on functionals F of the hypersurface X has the form

$$\mathcal{H}(N) := \int_\Sigma dy N(y) T^a(y) \frac{\delta}{\delta X^a(y)}\,, \tag{9.5}$$

while the tangential deformation operator is defined as

$$\mathcal{D}(N^\alpha \partial_\alpha) := \int_\Sigma dy N^\alpha(y) e^a_\alpha(y) \frac{\delta}{\delta X^a(y)}\,. \tag{9.6}$$

Thus the linear deformation of a functional F in the normal direction of $\dot{X}$ is given by $\mathcal{H}(N)F$, while its linear deformation in the purely tangential direction of $\dot{X}$ is given by $\mathcal{D}(N^\alpha\partial_\alpha)F$. This immediately yields the commutation algebra that these deformation operators satisfy:

THEOREM 9.1.1. (*First theorem of Lecture VIII in* [14]*) The normal and tangential deformation operators, $\mathcal{H}$ and $\mathcal{D}$, acting on functionals on an initial data hypersurface, satisfy the hypersurface deformation algebra*

$$\begin{aligned}
[\mathcal{H}(N), \mathcal{H}(M)] &= -\mathcal{D}((\deg P - 1) P^{\alpha\beta}(M\partial_\beta N - N\partial_\beta M)\partial_\alpha)\,,\\
[\mathcal{D}(N^\alpha\partial_\alpha), \mathcal{H}(M)] &= -\mathcal{H}(N^\alpha \partial_\alpha M)\,,\\
[\mathcal{D}(N^\alpha\partial_\alpha), \mathcal{D}(M^\beta\partial_\beta)] &= -\mathcal{D}((N^\beta\partial_\beta M^\alpha - M^\beta \partial_\beta N^\alpha)\partial_\alpha)\,,
\end{aligned} \tag{9.7}$$

where

$$P^{\alpha\beta} = P(n, \dots, n, \epsilon^\alpha, \epsilon^\beta)\,. \tag{9.8}$$

Notice that the background geometry only appears in the first commutation relation between the normal deformation operators through the spatial tensor $P^{\alpha\beta}$, while the other two commutation relations are independent of it.

9.2. Dynamical evolution of the geometry.

The problem of finding gravitational dynamics is to find differential equations which determine the spatial geometry on a neighbouring hypersurface from the data on the initial hypersurface such that all these hypersurface geometries generate a predictive, time-orientable and energy-distinguishing tensorial geometry on the spacetime manifold M.

Now it is clear why we have to introduce canonically conjugate momenta $\hat{\pi}_A$ in addition to the hypersurface geometry tensor field $\hat{G}^A$, since in this dynamical picture, they are needed to compensate the lack of information on how the geometry looks like away from the hypersurface. The space of tensor fields (G^A, π_A) is therefore called the geometric phase space if equipped with the Poisson brackets

$$\{\hat{F}, \hat{G}\} = \int_\Sigma dy \left[\frac{\delta \hat{F}}{\delta \hat{G}^A(y)} \frac{\delta \hat{G}}{\delta \hat{\pi}_A(y)} - \frac{\delta \hat{G}}{\delta \hat{G}^A(y)} \frac{\delta \hat{F}}{\delta \hat{\pi}_A(y)}\right]\,,$$

where $\hat{F}$ and $\hat{G}$ are arbitrary functionals of the geometric phase space variables $(\hat{G}^A, \hat{\pi}_A)$.

More precisely, one constructs the dynamics by requiring the evolution of the initial data on the initial hypersurface to neighbouring hypersurfaces is consistent with the action of the deformation operators on the initial data. This is achieved by requiring that the dynamics for observables on the geometric phase space are provided by the Hamiltonian

$$H = \int_\Sigma dy \, \left[\hat{\mathcal{H}}(N) + \hat{\mathcal{D}}(N^\alpha e_\alpha)\right] ,$$

where the so called superhamiltonian $\hat{\mathcal{H}}$ and supermomentum $\hat{\mathcal{D}}$ are functionals of the geometric phase space variables (P^A, π_A) that represent the hypersurface deformation algebra by virtue of

$$\begin{aligned} \{\hat{\mathcal{H}}(N), \hat{\mathcal{H}}(M)\} &= \hat{\mathcal{D}}((\deg P - 1)\hat{P}^{\alpha\beta}(M\partial_\beta N - N\partial_\beta M)\partial_\alpha)\,, \\ \{\hat{\mathcal{D}}(N^\alpha\partial_\alpha), \hat{\mathcal{H}}(M)\} &= \hat{\mathcal{H}}(N^\alpha\partial_\alpha M)\,, \\ \{\hat{\mathcal{D}}(N^\alpha\partial_\alpha), \hat{\mathcal{D}}(M^\beta\partial_\beta)\} &= \hat{\mathcal{D}}((N^\beta\partial_\beta M^\alpha - M^\beta\partial_\beta N^\alpha)\partial_\alpha)\,. \end{aligned} \tag{9.9}$$

Solving this algebra for the superhamiltonian $\hat{\mathcal{H}}$ and the supermomentum $\hat{\mathcal{D}}$ then yields the gravitational dynamics. For the familiar simple case of a Lorentzian metric spacetime structure, this has been shown in seminal work by Hojman, Kuchar and Teitelboim. In our general case of predictive, time-orientable and energy-distinguishing tensorial geometries, solving this algebra at first appears to be the rather difficult problem of solving a non-linear functional-differential equation. However, in [14, 15] it was shown that this can be remarkably be reduced. It is fair to say that none of this would have been possible without the groundwork laid in this thesis.

CHAPTER 10

Conclusions

In this thesis, we characterized all tensorial geometries (M, G), where M is a finite-dimensional smooth manifold M and G a smooth tensor field of arbitrary rank, that qualify as spacetime geometries in the sense that they allow for linear matter field dynamics that are predictive, interpretable and quantizable. We showed that these physical conditions on matter field dynamics translate into three corresponding simple algebraic conditions — hyperbolicity, time-orientability and energy-distinguishability — on the underlying tensor field G that defines the geometry. More precisely, these three algebraic conditions are to be imposed on G not directly, but by way of a totally symmetric contravariant field P of even rank, which is extracted from the principal polynomial of the matter field dynamics: hyperbolicity means that the polynomial P_x defined in each cotangent space is hyperbolic, time-orientability means that the dual polynomial $P_x^\#$ to P_x defined in each tangent space is also hyperbolic, and the energy-distinguishing condition means that all observers, defined in terms of $P^\#$, agree on the sign of the energy of massless momenta. Thus the conditions we identified for a tensorial structure to qualify as a spacetime geometry are based on the study of given matter field dynamics. This need to specify matter field dynamics is not seen as a weakness, but rather as an insight of the general theory presented in this thesis. After all, it was also the study of a matter field, namely the electromagnetic field, that ultimately led Einstein to conclude that the geometry of spacetime could be encoded in a Lorentzian metric. And superluminal neutrinos, or indeed any other matter that does not mimick the structure of Maxwell theory, will force us to choose another tensorial geometry.

We showed that all kinematical constructions known from general relativity exist, for precisely the same fundamental reasons, on any hyperbolic, time-orientable and energy-distinguishing geometry. In particular, a complete theory of massive and massless point particles was constructed on any geometric background satisfying these conditions. On each cotangent spaces, the geometry seen by massive and massless point particles is encoded in the hyperbolic polynomial P_x. However, in tangent spaces, the geometry seen by massless point particles is encoded in the dual polynomial $P_x^\#$, while the geometry seen by massive particles is encoded in the generically non-polynomial structure P^*. Both structures in tangent space ($P^\#$ and P^*) are determined by the polynomial P_x in cotangent space and are needed there due to the different duality theories necessary to associate vector velocities with momenta of massless and massive particles: the inverse duality map for massless particles is provided in terms of $P^\#$, while the inverse duality map for massive particles is provided in terms of P^*. In Lorentzian geometry, the conceptually different rôle that these mathematical structures play is seriously hidden, since in this case, the tensorial structure coupling to matter fields is given by the inverse metric, the hyperbolic polynomial P is also given by the inverse the metric, and $P^{\#}$ and P^{*} are again both given by the metric. In any other spacetime geometry with $\deg P > 2$, however, P, $P^{\#}$ and P^{*} are provided by entirely structurally different mathematical objects, as we illustrated for the case of area metric geometry. Furthermore, the three fundamental conditions on the geometry can equally be used in order to probe the viability of modified dispersion relations, since we know that a modified dispersion relation, if viable, must be written in terms of a hyperbolic, time-orientable and energy-distinguishing polynomial P.

This also clarifies why most approaches of Finsler geometry attempting to describe spacetime beginning with a single geometric structure in the tangent bundle are not entirely compelling. While these approaches have been right in speculating about a generically non-polynomial structure on the tangent bundle (by means of a pseudo-norm), we now know that the behaviour of massless and massive particles cannot be encoded in a *single* geometric structure in the tangent bundle. However, for massive particles, we also showed that a hyperbolic, time-orientable and energy-distinguishing geometry automatically provides a pseudo-Finslerian analogue of Lorentzian geometry. A generically non-linear connection was then introduced in terms of such pseudo-Finslerian structure in order to provide a notion of freely falling and non-rotating frames.

It turned out that massive particles can propagate faster than some massless particles on spacetime geometries with $\deg P > 2$. This is due to the multiple cone structure of the set of massless momenta that one has for spacetime geometries with $\deg P > 2$. The result is that massive particles whose tangent vectors lie outside the cone of observers propagate faster than massless particles whose tangent vectors lie on the boundary of the cone of observers. A superluminal massive particle can radiate off energy in form of massless particles, by means of a vacuum Cherenkov process, but only until its tangent vector reaches the cone of observers, which turns out to be equivalent to being infraluminal. Thus, although the propagation of massive particles with velocities higher than some of massless particles is generically allowed, massive particles always tend to infraluminal propagation. On Lorentzian geometry, superluminal propagation of massive particles is kinematically forbidden, thus if the observation of superluminal neutrinos by the OPERA collaboration is confirmed, we are forced to consider more general geometric backgrounds. But, as we learnt in this thesis, not any. Only those that are hyperbolic, time-orientable and energy-distinguishing allow superluminal propagation and, at the same time, are consistent with causality. Fortunately, we now have these geometries under good mathematical control.

However, in order to have a complete description of the mentioned Cherenkov process, it is necessary to be able to compute the corresponding decay rate. This requires the development of a quantum theory for tensorial spacetime geometries. In this thesis, we provided essential steps towards the construction of such a quantum theory. In particular, we performed the canonical quantization of free massive particles and of a special class of area metric electrodynamics. For the case of massive particles, it was crucial to recognize that only energy solutions lying on the hyperbolicity cones of P are physical and, therefore, that only they must be considered. However, for the canonical quantization of a massive matter field, if one wants to preserve microcausality, all energy solutions must be considered. In that case, a physical interpretation of the energy solutions lying outside the hyperbolicity cones must be provided, and this remains a question for further research. For area metric electrodynamics, a canonical quantization was performed on a particular area metric background. Here further research has to show how to develop a frame-independent quantization scheme, which is essentially the problem of finding a suitable gauge. Moreover, we identified necessary and sufficient conditions to generate gauge-free classical linear field equations of arbitrary differential order with a hyperbolic, time-orientable and energy-distinguishing dispersion relation. In particular, this led us to the construction of the algebra of generalized Dirac matrices and corresponding field equations on flat tensorial spacetimes.

Building on the kinematical constructions presented in this thesis, it was achieved in subsequent work to find a complete characterization of the gravitational dynamics of the tensorial spacetimes cosidered here. This was done by means of obtaining the deformation algebra of hypersurfaces in any hyperbolic, time-orientable and energy-distinguishing geometry. This algebra describes how the geometry induced on a hypersurface changes when the hypersurface is deformed into normal and tangential directions. Our constructions played the key rôle in that work by allowing for the association of normal vectors to hypersurface normal co-vectors by means of the Legendre map, whose existence is indeed only ensured by the hyperbolicity, time-orientability and energy-distinguishability properties of the underlying geometry. So this allowed to compute the deformation algebra satisfied by the deformation operators. The gravitational dynamics one desires are then determined by finding a representation of this deformation algebra in terms of geometric canonical variables which are functionals of the embedding map of a hypersurface. This problem appears to amount to solve a system of non-linear functional-differential equations. But fortunately, using the techniques of this thesis, the problem can be significantly reduced namely to a system of homogeneous linear partial differential equations. In combination with the kinematical results presented in this thesis, one thus has converted the physical question of what alternative tensorial spacetime geometries there could be, what their dynamics are and how they are to be interpreted kinematically into a purely mathematical task. Namely to solve a system of homogeneous linear partial differential equations.

Exciting open problems that can now be attacked based on the present work are

- to find explicit solutions to the linear partial differential equations that determine modified gravity theories, and, therefore, to obtain a concrete candidate to refine general relativity.
- to construct a complete quantum theory on hyperbolic, time-orientable and energy-distinguishable spacetime geometries.

This will be needed to further probe the alternative spacetime geometries one might well be forced to consider in the light of the problems of general relativity which motivated this thesis.

Bibliography

[1] A. Einstein, "On the electrodynamics of moving bodies," *Annalen Phys.*, vol. 17, pp. 891–921, 1905.

[2] A. Rendall, *Partial differential equations in general relativity.* Oxford University Press, 2008.

[3] J. Rauch, "Hyperbolic Partial Differential Equations and Geometric Optics," *http://www.math.lsa.umich.edu/ rauch/nlgonotes.pdf.*

[4] S. Hojman, K. Kuchar, and C. Teitelboim, "Geometrodynamics regained," *Ann. Phys.(NY);(United States)*, vol. 96, no. 1, 1976.

[5] C. Lämmerzahl, O. Preuss, and H. Dittus, "Is the physics within the solar system really understood?," *Lasers, Clocks and Drag-Free Control*, pp. 75–101, 2008.

[6] G. Starkman, "Modifying gravity: you cannot always get what you want," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1957, pp. 5018–5041, 2011.

[7] S. Carroll, *Spacetime and geometry. An introduction to general relativity.* Addison Wesley Publishing Company, 2004.

[8] S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. Castro, S. Deustua, S. Fabbro, A. Goobar, D. Groom, *et al.*, "Measurements of and λ from 42 high-redshift supernovae," *The Astrophysical Journal*, vol. 517, p. 565, 1999.

[9] A. Riess, A. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. Garnavich, R. Gilliland, C. Hogan, S. Jha, R. Kirshner, *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *The Astronomical Journal*, vol. 116, p. 1009, 1998.

[10] T. Sumner, "Experimental searches for dark matter," *Living Rev. Relativity*, vol. 5, no. 4, 2002.

[11] T. Adam *et al.*, "Measurement of the neutrino velocity with the opera detector in the cngs beam," 2011.

[12] A. G. Cohen and S. L. Glashow, "Pair Creation Constrains Superluminal Neutrino Propagation," *Phys.Rev.Lett.*, vol. 107, p. 181803, 2011.

[13] D. Rätzel, S. Rivera, and F. Schuller, "Geometry of physical dispersion relations," *Physical Review D*, vol. 83, no. 4, p. 044047, 2011.

[14] F. Schuller, "All spacetimes beyond Einstein (Obergurgl Lectures)," *Arxiv preprint arXiv:1111.4824*, 2011.

[15] K. Giesel, F. P. Schuller, C. Witte, and M. Wohlfarth, "Gravitational dynamics for tensorial spacetimes carrying predictive, interpretable and quantizable matter dynamics," *in preparation*, 2012.

[16] V. Y. Ivrii and V. M. Petkov, "Necessary conditions for the cauchy problem for non-strictly hyperbolic equations to be well-posed," *Russian Mathematical Surveys*, vol. 29, no. 5, p. 1, 1974.

[17] L. Hörmander, "The Cauchy problem for differential equations with double characteristics," *Journal d'Analyse Mathématique*, vol. 32, no. 1, pp. 118–196, 1977.

[18] F. Schuller, C. Witte, and M. Wohlfarth, "Causal structure and algebraic classification of non-dissipative linear optical media," *Annals of Physics*, vol. In Press, Corrected Proof, 2010.

[19] L. Gårding, "An inequality for hyperbolic polynomials," *J. Math. Mech*, vol. 8, no. 6, pp. 957–965, 1959.

[20] O. Gueler, "Hyperbolic polynomials and interior point methods for convex programming," *Mathematics of Operations Research*, vol. 22, no. 2, pp. 350–377, 1997.

[21] R. Benedetti and J. Risler, *Real algebraic and semi-algebraic sets.* Hermann, 1990.

[22] J. Bochnak, M. Coste, and M. Roy, *Real Algebraic Geometry.* Springer Verlag, 1998.

[23] H. Bauschke, O. Gueler, A. Lewis, and H. Sendov, "Hyperbolic polynomials and convex analysis," *Canad. J. Math*, vol. 53, no. 3, pp. 470–488, 2001.

[24] C. Lämmerzahl, "The geometry of matter fields," *Quantum mechanics in curved space-time/ edited by Juergen Audretsch and Venzo De Sabbata*, p. 23, 1990.

[25] Y. Egorov and M. Shubin, *Partial Differential Equations 2: Elements of the Modern Theory. Equations with Constant Coefficients.* Springer, 1994.

[26] V. Perlick, *Ray Optics, Fermat's Principle, and Applications to General Relativity.* Springer Verlag, 2000.

[27] J. Audretsch and C. Lämmerzahl, "Establishing the riemannian structure of space-time by means of light rays and free matter waves," *Journal of Mathematical Physics*, vol. 32, pp. 2099–2105, Aug. 1991.

[28] R. Beig, "Concepts of hyperbolicity and relativistic continuum mechanics," *Analytical and Numerical Approaches to Mathematical Relativity*, pp. 101–116, 2006.

[29] M. Fedoryuk, *Partial Differential Equations: Asymptotic Methods for Partial Differential Equations.* Springer, 1999.

[30] B. Hassett, *Introduction to algebraic geometry.* Cambridge Univ. Press, 2007.

[31] D. Dubois and G. Efroymson, "Algebraic theory of real varieties. I," pp. 107–135.

[32] R. Rockafellar, *Convex analysis*. Princeton University Press, Princeton, N.J., 1970.

[33] H. Rund, *Differential Geometry of Finsler Spaces*. Springer, 1959.

[34] Z. Shen, *Lectures on Finsler geometry*. World Scientific Pub Co Inc, 2001.

[35] Y. Nesterov and M. Todd, "Self-scaled barriers and interior-point methods for convex programming," *Mathematics of Operations Research*, pp. 1–42, 1997.

[36] J. Skakala and M. Visser, "Birefringence in pseudofinsler spacetimes," *Journal of Physics: Conference Series*, vol. 189, no. 1, p. 012037, 2009.

[37] J. Skakala and M. Visser, "Bi-metric pseudo-finslerian spacetimes," *Journal of Geometry and Physics*, 2011.

[38] F. Schuller and M. Wohlfarth, "Geometry of manifolds with area metric," *Nucl. Phys.*, vol. B747, pp. 398–422, 2006.

[39] F. Schuller and M. Wohlfarth, "Canonical differential geometry of string backgrounds," *Journal of High Energy Physics*, vol. 2006, no. 02, p. 059, 2006.

[40] F. W. Hehl, Y. N. Obukhov, and G. F. Rubilar, "Light propagation in generally covariant electrodynamics and the fresnel equation," *Int. J. Mod. Phys.*, vol. A17, pp. 2695–2700, 2002.

[41] F. Hehl, Y. Obukhov, and I. Obukhov, *Foundations of classical electrodynamics: Charge, flux, and metric*. Birkhauser, 2003.

[42] V. Perlick, "On the hyperbolicity of maxwell's equations with a local constitutive law," *Journal of Mathematical Physics*, vol. 52, p. 042903, 2011.

[43] R. Punzi, F. Schuller, and M. Wohlfarth, "Propagation of light in area metric backgrounds," *Classical and Quantum Gravity*, vol. 26, no. 3, p. 035024, 2009.

[44] R. Gambini and J. Pullin, "Nonstandard optics from quantum spacetime," *Phys. Rev.*, vol. D59, p. 124021, 1999.

[45] M. Atiyah, R. Bott, and L. Gårding, "Lacunas for hyperbolic differential operators with constant coefficients I," *Acta Mathematica*, vol. 124, no. 1, pp. 109–189, 1970.

[46] R. Gambini and J. Pullin, "Lorentz violations in canonical quantum gravity," 2001.

[47] J. Alfaro, H. A. Morales-Tecotl, and L. F. Urrutia, "Loop quantum gravity and light propagation," *Phys.Rev.*, vol. D65, p. 103509, 2002.

[48] R. C. Myers and M. Pospelov, "Ultraviolet modifications of dispersion relations in effective field theory," *Phys. Rev. Lett.*, vol. 90, p. 211601, May 2003.

[49] A. Hanson, T. Regge, and C. Teitelboim, *Constrained hamiltonian systems*. Accademia nazionale dei Lincei, 1976.

[50] K. Sundermeyer, "Constrained Dynamics (Lecture Notes in Physics 169)," *Sprlnger-Verlag, Berlin*, 1982.

[51] P. Dirac, *Lectures on quantum mechanics*. Dover Pubns, 2001.

[52] D. Gitman and I. Tyutin, *Quantization of fields with constraints*, vol. 288. Springer-Verlag, 1990.

[53] M. Amooshahi and B. Nasr Esfahani, "Canonical quantization of the electromagnetic field in the presence of non-dispersive bi-anisotropic inhomogeneous magnetodielectric media," *Annals of Physics*, vol. 325, no. 9, pp. 1913–1930, 2010.

[54] F. Gantmakher, *The theory of matrices*, vol. 1,2. Chelsea publishing company, 1959.

[55] S. Basu, R. Pollack, and M. Roy, *Algorithms in real algebraic geometry*, vol. 10. Springer-Verlag New York Inc, 2006.

[56] H. Casimir, "On the attraction between two perfectly conducting plates," in *Proc. K. Ned. Akad. Wet*, vol. 51, p. 793, 1948.

[57] S. Rivera and F. Schuller, "Quantization of general linear electrodynamics," *Physical Review D*, vol. 83, no. 6, p. 064036, 2011.

[58] P. Milonni and C. Eberlein, *The quantum vacuum: an introduction to quantum electrodynamics*. Academic press Boston, 1994.

[59] A. Favaro and L. Bergamin, "The non-birefringent limit of all linear, skewonless media and its unique light-cone structure," *Annalen der Physik*, 2011.

[60] L. Gårding, "Linear hyperbolic partial differential equations with constant coefficients," *Acta Mathematica*, vol. 85, no. 1, pp. 1–62, 1964.

[61] Y. Itin, "A generalized photon propagator," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, p. F737, 2007.

[62] H. Rund, *The Hamilton-Jacobi theory in the calculus of variations: its role in mathematics and physics*. Krieger Pub Co, 1966.

[63] A. Saa, "Canonical quantization of the relativistic particle in static spacetimes," *Classical and Quantum Gravity*, vol. 13, p. 553, 1996.

[64] J. Polchinski, *String Theory: Superstring theory and beyond*, vol. 2. Cambridge Univ Pr, 1998.

[65] M. Dahl, "Non-dissipative electromagnetic medium with a double light cone," *Arxiv preprint arXiv:1108.4207*, 2011.