ELSEVIER

Journal of Geometry and Physics 35 (2000) 1–34

JOURNAL OF GEOMETRY AND PHYSICS

A Morse theory for massive particles and photons in general relativity

Fabio Giannoni [a], Antonio Masiello [b,*], Paolo Piccione [c]

[a] *Dipartimento di Matematica e Fisica, Universitá di Camerino, Via Madonna delle Carceri, 20-62032-Camerino (MC), Italy*

[b] *Dipartimento Interuniversitario di Matematica, Politecnico di Bari, Via E. Orabona 4, 70125-Bari, Italy*

[c] *Departamento de Matemática, Universidade de São Paulo, São Paulo, SP, Brazil*

Received 4 May 1999

Abstract

In this paper we develop a Morse theory for time-like geodesics parameterized by a constant multiple of proper time. The results are obtained using an extension to the time-like case of the relativistic Fermat principle, and techniques from Global Analysis on infinite dimensional manifolds. In the second part of the paper we discuss a limit process that allows to obtain also a Morse theory for light rays.

Subj. Class: *General relativity*

1991 MSC: *83C22; 58F05*

Keywords: *Morse theory; General relativity*

1. Introduction

In an arbitrary relativistic space–time, modeled by a four-dimensional time-oriented Lorentzian manifold $(\mathcal{M}, g)$, the trajectories of massive objects or massless particles, like photons, that move freely under the action of the gravitational field, are geodesics. These geodesics are *time-like* in the massive case, representing the motion of objects traveling slower than the speed of light, and null, or light-like, in the case of (massless) particles moving at the speed of light. They can be characterized by variational principles which can be interpreted as extensions to General Relativity of the Fermat principle in classical optics.

* Corresponding author.

E-mail addresses: giannoni@campus.unicam.it (F. Giannoni), masiello@pascal.dm.uniba.it (A. Masiello), piccione@ime.usp.br (P. Piccione)

PII: S0393-0440(99)00045-5

Some of them can be used to describe the so-called *gravitational lens effect* that occurs in Astrophysics whenever multiple images of point-like sources (e.g. quasars) are observed (cf. e.g. [22]). In mathematical terminology, a gravitational lensing situation can be modeled in the following way. We consider a Lorentzian manifold $(\mathcal{M}, g)$ as a mathematical model for the space–time, we fix a time-like curve γ as the worldline of a light source and a point p as the event where the observation takes place. Now, the number of images seen by the observer equals the number of future pointing light-like geodesics from p to γ. Whenever there are two or more such geodesics, we are in a gravitational lensing situation. Alternatively, one could interpret p as an instantaneous point-like source of light and γ as the worldline of a receiver. Since the two problems can be treated in the same way from a mathematical point of view, we shall focus our attention only on this second case.

It should be remarked that different approaches to the mathematical modeling of the gravitational lensing effect are possible. For instance, in [19–21], the authors use a *thin lens* approximation; in [10] also nonthin lenses are considered.

In a recent paper, Kovner has suggested a very general version of the Fermat principle to study time-like and light-like geodesics (cf. [9]). Kovner's principle, justified by plausible arguments in [9] and rigorously proven by Perlick in [18] for the light-like case, can be stated as follows. Among all future pointing curves $z : [0, 1] \to \mathcal{M}$ joining p and γ and satisfying $g(z)[\dot{z}, \dot{z}] \equiv a$, with $a \leq 0$ fixed, i.e., all possibilities to go from p to γ at speed less than ($a < 0$) or equal to ($a = 0$) the (vacuum) speed of light, the geodesics are characterized as stationary points for the *arrival time* (defined using a smooth parameterization of γ). In the light-like case ($a = 0$), this principle generalizes the Fermat's principle for light rays in classical optics.

In an absolutely similar fashion, one could give a time-reversed version of the principle, by interpreting p as an instantaneous receiver and γ the worldline of a source. In this case, the geodesics are characterized by stationary *departure time*.

The aim of this paper is twofold. In the first part we shall develop a Morse theory for future pointing time-like geodesics with a prescribed parameterization (proportional to the proper time) and joining a given event with a time-like curve in a time-oriented Lorentzian manifold.

In the second part of the article, using a limit process, we shall prove the Morse relations for future pointing light-like geodesic (light rays), giving a new and simpler proof with respect to the ones of Giannoni et al. [6,7], where the existence of a smooth time function was assumed. In this paper we shall only assume the existence of a time-orientation for the Lorentzian manifold.

In order to state our results, we now give the basic definitions and we introduce the notations needed for our setup.

Let $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ be a time oriented Lorentz manifold and let Y be a smooth time-like vector field giving the time orientation (we refer to [1,15] for the basic notions of Lorentzian geometry that will be used). We set $m = \dim(\mathcal{M})$; the physical interesting case is $m = 4$.

Fix an event $p \in \mathcal{M}$ and a time-like curve $\gamma : \mathbb{R} \to \mathcal{M}$. On the curve γ we shall make the following assumptions:

- γ is of class C^2;
- γ is time-like and future pointing;
- γ is injective;
- $\gamma(\mathbb{R})$ does not contain p;
- $\gamma(\mathbb{R})$ is not entirely contained in $I^+(p)$, the *causal future* of p.

We recall that the causal future of a point p is defined as

$$I^+(p) = \{q \in \mathcal{M} | \text{there exists a future pointing causal curve}\\ z : [a,b] \mapsto \mathcal{M} \text{ with } z(a) = p \text{ and } z(b) = q\}. \tag{1.1}$$

As customary, if $I \subseteq \mathbb{R}$ is any interval, we will denote by $H^{1,2}(I, \mathbb{R}^n)$ the Sobolev space of all absolutely continuous curves $z : I \mapsto \mathbb{R}^n$ having square integrable derivative on I. Given any differentiable manifold N, with $n = \dim(N)$, we define $H^{1,2}([0,1], N)$ as the set of all absolutely continuous curves $z : [0,1] \mapsto N$ such that for every local chart (V, φ) on N, with $\varphi : U \mapsto \mathbb{R}^n$ a diffeomorphism, and for every closed subinterval $I \subseteq [0,1]$ such that $z(I) \subset V$, it is $\varphi \circ z \in H^{1,2}(I, \mathbb{R}^n)$ (cf. [16]).

It is not difficult to see that this definition of $H^{1,2}([0,1], N)$ may be given equivalently in the following two ways:

- a curve $z : [0,1] \mapsto N$ belongs to $H^{1,2}([0,1], N)$ if and only if there exists a finite sequence $I_1, \dots, I_k$ of closed subintervals of $[0,1]$ and a finite number of charts $\varphi_i : U_i \mapsto \mathbb{R}^n$ on N, $i = 1, \dots, k$, such that $\bigcup_{i=1}^k I_k = [0,1]$, $z(I_i) \subset U_i$, and $\varphi_i \circ z \in H^{1,2}(I_i, \mathbb{R}^n)$ for all $i = 1, \dots, k$;
- a C^1-curve $z : [0,1] \mapsto N$ is in $H^{1,2}([0,1], N)$ if and only if for one (hence for every) Riemannian metric $g_{(\mathrm{R})}$ on N, the integral $\int_0^1 g_{(\mathrm{R})}(\dot{z}, \dot{z})\,\mathrm{d}t$ is finite.

A classical result of Global Analysis (see [17]) states that $H^{1,2}([0,1], N)$ has the structure of an infinite dimensional manifold modeled on the Hilbert space $H^{1,2}([0,1], \mathbb{R}^n)$. Similarly, one defines the Banach manifolds $H^{k,p}([0,1], N)$, $k \in \mathbb{N}$, $1 \le p \le +\infty$, modeled on the Sobolev spaces $H^{k,p}([0,1], \mathbb{R}^n)$. In particular, in this paper we will be concerned with the manifolds $H^{k,p}([0,1], \mathcal{M})$ and $H^{k,p}([0,1], T\mathcal{M})$, where $T\mathcal{M}$ is the tangent bundle of $\mathcal{M}$.

If $g_{(\mathrm{R})}$ is any given Riemannian metric on $\mathcal{M}$, for $1 \le p \le +\infty$ we also define the spaces $L^p([0,1], T\mathcal{M})$ as the set of functions $\zeta : [0,1] \mapsto T\mathcal{M}$ such that the real valued function $g_{(\mathrm{R})}(\zeta, \zeta)^{1/2}$ is in $L^p([0,1], \mathbb{R})$. It is easy to see that, by the compactness of $[0,1]$, the definition of $L^p([0,1], T\mathcal{M})$ does not depend on the choice of a specific Riemannian metric $g_{(\mathrm{R})}$.

The natural setting to study future pointing light rays joining p and γ is the following space:

$$\mathcal{L}^+_{p,\gamma} = \{z : [0,1] \to \mathcal{M} | z \in H^{1,2}([0,1], \mathcal{M}),\\ \langle Y, \dot{z} \rangle < 0 \text{ for any } s \text{ such that } \dot{z}(s) \text{ exists and it is different from zero},\\ \langle \dot{z}, \dot{z} \rangle = 0 \text{ a.e., } z(0) = p, z(1) \in \gamma(\mathbb{R})\}.$$

Here the $H^{1,2}$-regularity is used because it is the simplest one if we want to give an infinite dimensional approach to the Morse theory.

Unfortunately, $\mathcal{L}^+_{p,\gamma}$ is not a C^1-submanifold of $H^{1,2}([0,1],\mathcal{M})$, but it only has a Lipschitz regularity. For this reason we shall approximate it by the family of smooth submanifolds of $H^{1,2}([0,1],\mathcal{M})$, parameterized by a positive number ϵ, given by

$$\mathcal{L}^+_{p,\gamma,\epsilon} = \{\, z : [0,1] \mapsto \mathcal{M} | z \in H^{1,2}([0,1],\mathcal{M}), \langle Y(z), \dot{z}\rangle < 0 \text{ a.e.}, \\ \langle \dot{z}, \dot{z}\rangle = -\epsilon^2 \text{ a.e.}, z(0) = p, z(1) \in \gamma(\mathbb{R})\}.$$

To complete our variational framework we introduce the *arrival time* functional τ which assigns to each curve ending on γ the value of the parameter of γ at the arrival point. The functional τ is defined on the manifold

$$\Omega^{1,2}_{p,\gamma} = \{z : [0,1] \to \mathcal{M} | z \in H^{1,2},\ \ z(0) = p, z(1) \in \gamma(\mathbb{R})\}$$

as

$$\tau(z) = \gamma^{-1}(z(1)).$$

Observe that τ is well defined because γ is injective.

Some relativistic versions of the Fermat principle have been already used (cf. e.g. [7] and the references therein) to develop a Morse theory for light rays. However, the Morse relations for time-like geodesics with prescribed parameterization has not been obtained yet. Moreover, the results for light rays in [6,7] have been proven under the extra assumption of stable causality for $\mathcal{M}$, i.e., assuming the existence of a smooth global time function $T : \mathcal{M} \mapsto \mathbb{R}$ on $\mathcal{M}$, and using the following functional:

$$Q(z) = \int_0^1 \langle \dot{z}, \nabla T\rangle^2 \, \mathrm{d}s,$$

where ∇T is the Lorentzian gradient of T.

In spite of the analogy with the energy functional in Riemannian manifolds, the critical points of Q on the approximating manifolds $\mathcal{L}^+_{p,\gamma,\epsilon}$ do not have a clear geometrical or physical meaning; moreover, the Euler–Lagrange equations for the Lagrangian function of Q are very complicated. This is one of the main reasons making the proof of Morse theory in [7] quite involved.

In this paper, due to the use of the arrival time functional τ on the manifolds $\mathcal{L}^+_{p,\gamma,\epsilon}$, we first obtain the Morse relations for the time-like geodesics, then, using a limit process as $\epsilon \to 0$, we extend the results to the case of light-like geodesics.

In order to avoid technical difficulties that could make not completely clear the advantages of this new approach, we will consider only the case where $\mathcal{M}$ is a manifold without boundary. It is worthy to observe here that the techniques presented in this paper can be employed also in the study of causal geodesics in manifolds having a causally convex boundary.

Before stating the main results of the present paper, let us recall the notions of conjugate point along a geodesic and the notion of geometric index.

We denote by D the Levi-Civita connection of the metric g; moreover, let R be the *curvature tensor* of g, defined with the following sign convention:

$$R(X,Y)Z = D_X D_Y Z - D_Y D_X Z - D_{[X,Y]} Z,$$

where X, Y, Z are vector fields on $\mathcal{M}$.

Definition 1.1. Let $z : [0, 1] \to \mathcal{M}$ be a geodesic. The point $z(s)$, $s \in]0, 1]$, is said to be *conjugate* to $z(0)$ along z if there exists a nonzero smooth vector field ζ along $z_{|[0,s]}$ (called *Jacobi field*) such that

$$D_s^2\zeta + R(\zeta, \dot{z})\dot{z} = 0, \tag{1.2}$$

and satisfying the boundary condition

$$\zeta(0) = 0, \quad \zeta(s) = 0. \tag{1.3}$$

The *multiplicity* of the conjugate point $z(s)$ is the maximal number of linearly independent Jacobi fields satisfying (1.3). The *geometric index* $\mu(z)$ of the geodesic z is the number of points $z(s)$ conjugate to $z(0)$ along z, counted with their multiplicity.

We shall prove that the functional τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ is of class C^2. Let z be a critical point of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$. The Morse index $m(z, \tau)$ is defined as the maximal dimension of a subspace of $T_z\mathcal{L}^+_{p,\gamma,\epsilon}$ (the tangent space to $\mathcal{L}^+_{p,\gamma,\epsilon}$ at z), where the Hessian of τ at z is negative definite.

The first result concerns the Fermat principle in $\mathcal{L}^+_{p,\gamma,\epsilon}$.

Theorem 1.2. *A curve z is a critical point of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ if and only if z is a future pointing time-like geodesic joining p with γ such that $\langle\dot{z}, \dot{z}\rangle = -\epsilon^2$. Moreover, if $z(1)$ is nonconjugate to $p = z(0)$ along z, then $m(z, \tau) = \mu(z)$.*

To write Morse relations for τ in $\mathcal{L}^+_{p,\gamma,\epsilon}$ (where $\epsilon > 0$ is fixed), we need to assume that $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ is *strongly causal*. This means that for any point $q \in \mathcal{M}$, there is no future pointing causal curves starting arbitrarily close to q, leaving some fixed neighborhood of p and returning arbitrarily close to q (cf. [1,15]).

Moreover, we need to recall some topological definitions. Let X be a topological space, $\mathcal{K}$ an algebraic field, for any $q \in \mathbb{N}$ we denote by $H_q(X, \mathcal{K})$ the qth homology group of X with coefficient in $\mathcal{K}$ (cf. [23]). Since $\mathcal{K}$ is a field, $H_q(X, \mathcal{K})$ is a vector space. The dimension $\beta_q(X, \mathcal{K})$ of $H_q(X, \mathcal{K})$ is called the q^{th} Betti number of X (with coefficients in $\mathcal{K}$). Finally the Poincaré polynomial of X is the formal series with coefficients in $\mathbb{N} \cup \{+\infty\}$ defined as

$$\mathcal{P}_r(X, \mathcal{K}) = \sum_{q=0}^{\infty}\beta_q(X, \mathcal{K})r^q.$$

Theorem 1.3. *Let $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ be strongly causal, γ a curve in $\mathcal{M}$ satisfying* (1.1) *and*

1. $\mathcal{L}^+_{p,\gamma,\epsilon} \neq \emptyset$;
2. *for any geodesic $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$, $z(1)$ is nonconjugate to $z(0) = p$;*
3. *the functional τ is pseudo-coercive on $\mathcal{L}^+_{p,\gamma,\epsilon}$, namely: for any $c \in \mathbb{R}$, there exists K_c compact subset of $\mathcal{M}$ such that $z([0, 1]) \subset K_c$ for any $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$ satisfying $\tau(z) \leq c$.*

Then, for any coefficient field $\mathcal{K}$, there exists a formal series $S(r)$ with coefficients in $\mathbb{N} \cup \{+\infty\}$ such that

$$\sum_{z\in\mathcal{G}^+_{p,\gamma,\epsilon}} r^{\mu(z)} = \mathcal{P}_r(\mathcal{L}^+_{p,\gamma,\epsilon};\mathcal{K}) + (1+r)S(r). \tag{1.4}$$

Here $\mathcal{G}^+_{p,\gamma,\epsilon}$ is the set of the time-like geodesics in $\mathcal{L}^+_{p,\gamma,\epsilon}$.

Remark 1.4. *The set of assumptions* (1.1) *on the curve γ imply immediately that τ is bounded from below in $\mathcal{L}^+_{p,\gamma,\epsilon}$.*

Taking the limit as $\epsilon \to 0$, we obtain also the Morse relations for light rays.

Theorem 1.5. *Let $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ be strongly causal, γ a curve satisfying* (1.1) *and*
1. $\mathcal{L}^+_{p,\gamma} \neq \emptyset$;
2. *for any geodesic $z \in \mathcal{L}^+_{p,\gamma}$, $z(1)$ is nonconjugate to $z(0) = p$;*
3. *the functional τ is pseudo-coercive on $\mathcal{L}^+_{p,\gamma}$, namely, for any $c \in \mathbb{R}$, there exists K_c compact subset of $\mathcal{M}$ such that $z([0,1]) \subset K_c$ for any $z \in \mathcal{L}^+_{p,\gamma}$ satisfying $\tau(z) \le c$.*

Then, for any coefficient field $\mathcal{K}$, there exists a formal series $S(r)$ with coefficients in $\mathbb{N} \cup \{+\infty\}$ such that

$$\sum_{z\in\mathcal{G}^+_{p,\gamma}} r^{\mu(z)} = \mathcal{P}_r(\mathcal{L}^+_{p,\gamma};\mathcal{K}) + (1+r)S(r). \tag{1.5}$$

Here $\mathcal{G}^+_{p,\gamma}$ is the set of the light-like geodesics in $\mathcal{L}^+_{p,\gamma}$.

For the limit process the following results are crucial.

Theorem 1.6. *Assume that $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ is strongly causal and τ is pseudo-coercive on $\mathcal{L}^+_{p,\gamma}$. Let $c \in \mathbb{R}$, $(\epsilon_m)_{m\in\mathbb{N}}$ any sequence in $\mathbb{R}^+$ with $\epsilon_m \to 0$, and $(z_m)_{m\in\mathbb{N}}$ a sequence of (time-like) geodesics in $\mathcal{L}^+_{p,\gamma,\epsilon_m}$, satisfying $\tau(z_m) \le c$ for all $m \in \mathbb{N}$. Then, z_m has a subsequence which is convergent (with respect to the C^2*-norm) *to a future pointing light-like geodesic joining p and γ.*

Theorem 1.7. *Let $(z_m)_{m\in\mathbb{N}}$ be a sequence of time-like geodesics convergent with respect to the C^2 -norm to a light-like geodesic z such that $z(0)$ and $z(1)$ are nonconjugate. Then*

$$\mu(z_m) = \mu(z) \quad \textit{for any } m \textit{ sufficiently large.}$$

Theorem 1.6 will be proved in Section 6. The proof of Theorem 1.7 involves the notion of *Maslov* index for a semi-Riemannian geodesic (see [8,13]). For causal Lorentzian geodesics, the Maslov index coincides with the geometric index of the geodesic, while in the general case it is given by a sort *algebraic* count of the multiplicities of the conjugate points along the geodesic. The Maslov index can be characterized as the intersection number between a curve and a codimension one subvariety of the *Lagrangian–Grassmannian* of a symplectic space, and thus it is stable by homotopies. The stability of the geometric index can be proven in more general contexts; details of the proof may be found in [13].

The Morse relations provide a global description of the multiple image effect for point-like sources. Some information about the physical phenomenon can be obtained directly using them: for instance the information about the odd number of images predicted by astrophysicists (cf. [7, Theorem 1.16]).

2. Existence of minimizers

Fix $\epsilon > 0$. In order to develop a Morse theory on $\mathcal{L}^+_{p,\gamma,\epsilon}$ using the functional τ, we should need the Palais–Smale condition for τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$. Namely, we should need that any sequence $(z_m)_{m\in\mathbb{N}}$ such that $\tau(z_m)_{m\in\mathbb{N}}$ is uniformly bounded with respect to m and $\mathrm{d}\tau(z_m) \to 0$ as $m \to \infty$ had a converging subsequence in $\mathcal{L}^+_{p,\gamma,\epsilon}$. Unfortunately, τ has homogeneity 1 as a length functional on a Riemannian manifold (as it can be proved using local coordinates). Therefore, the natural space to study the Palais–Smale condition is the space

$$\begin{aligned}\hat{\mathcal{L}}^+_{p,\gamma,\epsilon} &= \{ z \in H^{1,1}([0,1],\mathcal{M}) | \langle \dot z, \dot z\rangle \\ &= -\epsilon^2 \text{ a.e., } \langle Y(z), \dot z\rangle < 0 \text{ a.e., } z(0)=p, z(1)\in\gamma(\mathbb{R})\},\end{aligned} \tag{2.1}$$

where $H^{1,1}([0,1],\mathcal{M})$ denotes the space of the absolutely continuous curves (on any local chart) whose first derivative is integrable.

But to develop a Morse theory it is really more convenient to work on the Hilbert manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$. For this reason we shall use a curve shortening procedure working on the curve space $\mathcal{L}^+_{p,\gamma,\epsilon}$. The space $\mathcal{L}^+_{p,\gamma,\epsilon}$ is equipped with a structure of infinite dimensional manifold and its tangent space at a point z is given by

$$\begin{aligned}T_z\mathcal{L}^+_{p,\gamma,\epsilon} &= \{\zeta \in H^{1,2}([0,1],T\mathcal{M}) : \zeta(0)=0, \zeta(1)\|\dot\gamma(z(1)), \quad \langle z, D_s\zeta\rangle \\ &= 0 \text{ a.e., } \zeta(s)\in T_{z(s)}\mathcal{M} \text{ for any } s\in[0,1]\},\end{aligned} \tag{2.2}$$

where $T\mathcal{M}$ is the tangent bundle of $\mathcal{M}$ (cf. [6] replacing there ∇T by Y).

We introduce a Riemannian structure on $\mathcal{M}$ setting for any $p\in\mathcal{M}$ and $\zeta\in T_p\mathcal{M}$,

$$\langle\zeta,\zeta\rangle_{(\mathrm{R})} = \langle\zeta,\zeta\rangle - 2\frac{\langle\zeta, Y(z)\rangle^2}{\langle Y(z), Y(z)\rangle}. \tag{2.3}$$

The *wrong way Schwartz's inequality* (cf. [15]) shows that (2.3) is a Riemannian structure on $\mathcal{M}$. We shall denote by d_{R} the distance function induced by (2.3).

A Riemannian structure can be introduced on the manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$, setting for any $z\in\mathcal{L}^+_{p,\gamma,\epsilon}$ and $\zeta\in T_z\mathcal{L}^+_{p,\gamma,\epsilon}$,

$$\langle\zeta,\zeta\rangle_1 = \int_0^1 \langle D_s\zeta, D_s\zeta\rangle_{(\mathrm{R})}\,\mathrm{d}s. \tag{2.4}$$

The proof is formally the same as in [6], where the existence of a time function is assumed.

Now, for any $[a,b]\subset[0,1]$, $-\infty<\alpha<\beta<+\infty$, $q\in\mathcal{M}$ and $\delta :]\alpha,\beta[\to\mathcal{M}$ smooth time-like curve, we set

$$\begin{aligned}&\mathcal{L}^+_{q,\delta,\epsilon}([a,b]) = \{ z\in H^{1,2}([a,b],\mathcal{M} : z(a)=q, z(b)\in\delta(]\alpha,\beta[), \\ &\langle\dot z,\dot z\rangle = -\epsilon^2 \text{ a.e., } \langle\dot z, Y(z)\rangle < 0 \text{ a.e.}\}.\end{aligned} \tag{2.5}$$

Note that δ is injective because $\mathcal{M}$ is strongly causal.

The main result of this section is the following result on the existence and the uniqueness of minimizers of the arrival time τ between a point and a "sufficiently close" integral curve δ of the vector field Y (obviously δ is a time-like curve).

Theorem 2.1. *Fix $-\infty < \alpha < \beta < +\infty$. For any $q \in \mathcal{M}$ there exists a positive number $\rho(q)$ having the following property: for any integral curve δ $:]\alpha,\beta[\to \mathcal{M}$ of Y such that $d_{\mathrm{R}}(q, \delta((\alpha+\beta)/2) \le \rho(q)$, and for any interval $[a,b]$ such that $0 < |b-a| \le \rho(q)$, there exists one and only one $z \in \mathcal{L}^{+}_{q,\delta,\epsilon}([a,b])$ which minimizes the arrival time on $\mathcal{L}^{+}_{q,\delta,\epsilon}([a,b])$.*

Note that in the statement of theorem 2.1, the arrival time is given by $\tau(z) = \delta^{-1}(z(b))$. Set

$$T_z\hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b]) = \{\zeta \in H^{1,1}([a,b], T\mathcal{M} : \zeta(a)=0,\ \zeta(b)\|\dot\delta(z(b)),\ \langle z, D_s\zeta\rangle = 0 \text{ a.e.},$$
$$\zeta(s) \in T_{z(s)}\mathcal{M},\ \text{for any } s \in [0,1]\}. \tag{2.6}$$

Note that the space $T_z\hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b])$ must be considered as a tangent space, but only in a "Gateaux" sense. This is what we need to prove Theorem 2.1.

In order to prove Theorem 2.1, some preliminary results are needed. The first says that τ satisfies the Palais–Smale condition with respect to the *admissible variations* in $T_z\hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b])$ and with respect to the "Finsler" structure on $\hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b])$ defined in the following way: for any $z \in \hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b])$ and for any $\zeta \in T_z\hat{\mathcal{L}}^{+}_{q,\delta,\epsilon}([a,b])$, we set

$$\|\zeta\|_{1,a,b} \equiv \|\zeta\| = \int_a^b (\langle D_s^{\mathrm{R}}\zeta, D_s^{\mathrm{R}}\zeta\rangle_{(\mathrm{R})} + \langle\zeta,\zeta\rangle_{(\mathrm{R})})^{1/2}\,\mathrm{d}s, \tag{2.7}$$

where D_s^{R} denotes the Levi-Civita connection with respect to the Riemannian metric (2.3).

Remark 2.2. *Note that since δ is a curve of class C^2 and τ is characterized by the relation $\delta(\tau(z)) = z(b)$, we have that τ is a functional of class C^2 on the space of the curves parameterized on the interval $[a,b]$ and joining q and δ. Moreover, its differential along a direction ζ is given by*

$$\dot\delta(\tau(z))\,\mathrm{d}\tau(z)[\zeta] = \zeta(b).$$

Therefore

$$\mathrm{d}\tau(z)[\zeta] = \frac{\langle\dot\delta(\tau(z)), \zeta(b)\rangle}{\langle\dot\delta(\tau(z)), \dot\delta(\tau(z))\rangle}. \tag{2.8}$$

Remark 2.3. *In the rest of the paper the parallel transport of $\dot\delta(z(b))$ along z will be often used, namely the solution $U(z)$ of the Cauchy problem*

$$D_sU(z) = 0, \qquad U(b) = \dot\delta(\tau(z)), \tag{2.9}$$

where D_s is the covariant derivative along $z(s)$. Note that if z has a $H^{1,r}$-regularity, then also $U(z)$ is of class $H^{1,r}$ $(r \in [1,\infty])$.

Since the parallel transport is an isometry, the vector field $U(z)$ along z is time-like and for any $s \in [a, b]$,

$$\langle \dot{\delta}(\tau(z)), \dot{\delta}(\tau(z)) \rangle = \langle U(z)(s), U(z)(s) \rangle.$$

Moreover, any vector field ζ along z such that $\zeta(a) = 0$, $\zeta(b) = 0$ can be projected on $T_z \hat{\mathcal{L}}^+_{q,\delta,\epsilon}([a, b])$ using $U(z)$. Indeed, set

$$V_\zeta(s) = \zeta(s) - \mu(s) U(z)(s), \qquad \mu(s) = \int_a^s \frac{\langle D_s \zeta, \dot{z} \rangle}{\langle U(z), \dot{z} \rangle} \mathrm{d}r. \tag{2.10}$$

Clearly $V_\zeta \in T_z \hat{\mathcal{L}}^+_{q,\delta,\epsilon}([a, b])$, and by (2.8),

$$\mathrm{d}\tau(z)[V_\zeta] = \mathrm{d}\tau(z)[\zeta - \mu U(z)] = -\mu(b) = -\int_a^b \frac{\langle D_s \zeta, \dot{z} \rangle}{\langle U(z), \dot{z} \rangle} \mathrm{d}r. \tag{2.11}$$

Note that zero-homogeneity of the map

$$\theta \to \frac{\theta}{\langle U(z), \theta \rangle}$$

shows that the vector field $\langle U(z), \dot{z} \rangle^{-1} \dot{z}$ is uniformly bounded, and therefore $\mu(s) \in H^{1,1}([a, b], \mathbb{R})$.

Proposition 2.4. *Let $(z_m)_{m \in \mathbb{N}}$ be a sequence of curves of class C^1 and such that $z_m \in \hat{\mathcal{L}}^+_{q,\delta,\epsilon}([a, b])$ for any $m \in \mathbb{N}$. Assume that*
1. *$\tau(z_m) \to c \in]\alpha, \beta[$, as $m \to \infty$, where $]\alpha, \beta[$ is the interval where δ is defined;*
2. *$\sup\{|\mathrm{d}\tau(z_m)[\zeta]| : \zeta \in T_{z_m} \hat{\mathcal{L}}^+_{q,\delta,\epsilon}([a, b]), \|\zeta\|_{a,b,1} \leq 1\} \to 0$ as $m \to \infty$.*

Then the sequence $(z_m)_{m \in \mathbb{N}}$ contains a subsequence converging to a curve z with respect to the C^1 -norm.

In order to prove Proposition 2.4, the following remarks and lemmas are needed.

Remark 2.5. *It is not difficult to verify that for any $z_0 \in \mathcal{M}$ there exists a local chart (U, φ) of $\mathcal{M}$ containing z_0 such that $\varphi(U) = V \times I$, where V is a convex open subset of $\mathbb{R}^n$, $n = m - 1$, I is an open interval,*

$$\varphi(U) = \left\{ (x, t) : x = (x_1, \ldots, x_n), \text{ the distribution generated by the } \frac{\partial}{\partial x_i}\text{'s is space-like and } \frac{\partial}{\partial t} = Y \right\},$$

and the Lorentzian metric g on $\varphi(U)$ can be written as

$$\mathrm{d}s^2 = \langle \alpha(x, t)\xi, \xi \rangle_0 + 2\langle \Gamma(x, t), \xi \rangle_0 \theta - \beta(x, t)\theta^2, \tag{2.12}$$

where $\langle \cdot, \cdot \rangle_0$ is a Riemann structure on V, $\alpha(x, t)$ is a positive linear operator, Γ is a smooth vector field, $\beta(x, t)$ is a smooth positive scalar field, and $(\xi, \theta) \in \mathbb{R}^n \times \mathbb{R}$.

Lemma 2.6. *Assume that τ is pseudo-coercive on $\hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$ (or equivalently on $\mathcal{L}^+_{p,\gamma,\epsilon}$). Then, for any $c \in \mathbb{R}$ there exists $D(c) > 0$ such that*

$$\tau(z) \le c \Rightarrow \int_0^1 \sqrt{\langle \dot{z}, \dot{z}\rangle_{\mathrm{R}}}\,\mathrm{d}s \le D(c).$$

Proof. Assume by contradiction that there exists a sequence $(z_m)_{m\in\mathbb{N}}$ in $\hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$ such that $\tau(z_m) \le c$ for any $m \in \mathbb{N}$ and

$$\int_0^1 \sqrt{\langle \dot{z}_m, \dot{z}_m\rangle_{\mathrm{R}}}\,\mathrm{d}s \to +\infty. \tag{2.13}$$

Set $\hat{z}_m(s) = z_m(s/\lambda_m)$, where $\lambda_m = \sup\{\langle \dot{z}_m, \dot{z}_m\rangle_{\mathrm{R}}^{1/2} : s \in [0,1]\}$. By pseudo-coercivity (and Ascoli–Arzelá's theorem), up to passing to a subsequence there exists a curve $z : \mathbb{R}^+ \to \mathcal{M}$ such that

$$\hat{z}_m \to z \quad \text{uniformly on the compact subsets of } \mathbb{R}^+. \tag{2.14}$$

Now fix $r > 0$ and consider the interval $[0, r]$. Suppose that $z(r+1)$ does not intersect $z([0, r])$. (Since $\mathcal{M}$ is strongly causal and any $\hat{z}_m$ is causal, this means that z is not constant on the interval $[r, r+1]$.) The strong causality of $\mathcal{M}$ implies (arguing by contradiction) that $z(s)$ is uniformly far from $z([0, r])$ on $[r+1, +\infty[$. Therefore, we can use a countable set of local charts (U_j, φ_j), $j = 1, \ldots, k$, as in Remark 2.5 and the t-coordinate on any $\varphi_j(U_j)$ to construct, without ambiguity, a smooth map T on a relatively compact neighborhood $\mathcal{U}$ of $z(\mathbb{R}^+)$ such that for any $q \in \mathcal{U}$,

$$\langle \nabla T(q), \nabla T(q)\rangle < 0 \quad \text{and} \quad \langle \nabla T(q), Y(q)\rangle < 0.$$

Now any z_m is time-like and $\langle Y(z_m), \dot{z}_m\rangle < 0$ for m and for any $s \in [0, 1]$. Then, for any m sufficiently large, $\langle \nabla T(z_m), \dot{z}_m\rangle > 0$ for any $s \in [0, 1]$. Moreover, since $\tau(z_m) \le c$, we have (unless to consider a subsequence) that

$$T(z_m(1)) \text{ is bounded}.$$

Now

$$T(z_m(1)) - T(p) = T(z_m(1)) - T(0) = \int_0^1 \langle \nabla T(z_m), \dot{z}_m\rangle\,\mathrm{d}s, \tag{2.15}$$

while, by (2.3) and the choice of the orientation of $\nabla T(z)$, there exists ν_0 such that

$$\langle \nabla T(z_m), \dot{z}_m\rangle \ge \nu_0 \sqrt{\langle \dot{z}_m, \dot{z}_m\rangle_{(\mathrm{R})}} \tag{2.16}$$

for any $s \in [0, 1]$ and m sufficiently large (recall that $z_m \in \hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$).

Since $T(z_m(1))$ is bounded, combining (2.14) and (2.15) gives the boundedness of

$$\int_0^1 \sqrt{\langle \dot{z}_m, \dot{z}_m\rangle_{(\mathrm{R})}}\,\mathrm{d}s,$$

in contradiction with (2.13). □

Proof of Proposition 2.4. The proof will be carried out assuming $[a, b] = [0, 1]$. Since $(z_m)_{m\in\mathbb{N}}$ is a Palais–Smale sequence,

$$\lim_{m\to\infty}(\sup\{\|\tau'(z_m)[\zeta]\|_1 : \zeta \in T_{z_m}\hat{\mathcal{L}}^+_{p,\gamma,\epsilon},\ \|\zeta\|_1 \leq 1\}) = 0.$$

By assumptions (i) and pseudo-coercivity, there exists K compact subset of $\mathcal{M}$ such that

$$z_m([0, 1]) \subset K \quad \text{for any } m.$$

Moreover, well-known results on dual Sobolev spaces (cf. [2]) imply that

$$\tau'(z_m)[\zeta] = \int_0^1 \langle \alpha_m, D^{\mathrm{R}}_s\zeta\rangle_{(\mathrm{R})}\,\mathrm{d}s + \int_0^1 \langle \beta_m, \zeta\rangle_{(\mathrm{R})}\,\mathrm{d}s, \tag{2.17}$$

where α_m and β_m are L^∞-vector fields along z_m and

$$\alpha_m \to 0, \qquad \beta_m \to 0 \text{ uniformly}.$$

Now

$$D^{\mathrm{R}}_s\zeta - D_s\zeta = \Gamma(z_m)[\dot{z}, \zeta],$$

where $\Gamma(z_m)$ is a bilinear map, whose components are smooth functions of z_m. Then there exists a vector field $\hat{\beta}_m$ along z_m (of class $H^{1,1}$) and a bilinear map $B(z_m)[\cdot,\cdot]$ such that $\hat{\beta}_m \to 0$ uniformly and

$$\tau'(z_m)[\zeta] = \int_0^1 \langle \alpha_m, D_s\zeta\rangle\,\mathrm{d}s + \int_0^1 \langle B(z_m)[\alpha_m, \dot{z}_m], \zeta\rangle\,\mathrm{d}s + \int_0^1 \langle \hat{\beta}_m, \zeta\rangle\,\mathrm{d}s. \tag{2.18}$$

Then, if μ and $U(z_m)$ are as in (2.9) and (2.10), with ζ replaced by W, for every $W \in C^\infty_0([0, 1], T\mathcal{M})$ such that $W(s) \in T_{z_m(s)}\mathcal{M}$ for any s, we have

$$\begin{aligned}\tau'(z_m)[W - \mu U(z_m)] = {} & \int_0^1 \langle \alpha_m, D_s(W - \mu U(z_m))\rangle\,\mathrm{d}s \\ & + \int_0^1 \langle B(z_m)[\alpha_m, \dot{z}_m] + \hat{\beta}_m, W - \mu U(z_m)\rangle\,\mathrm{d}s.\end{aligned}$$

Since $D_sU(z_m) = 0$ and

$$\mu(s) = \int_0^s \frac{\langle D_sW, \dot{z}_m\rangle}{\langle U(z_m), \dot{z}_m\rangle}\mathrm{d}s,$$

by (2.10) we have

$$\begin{aligned} & -\int_0^1 \left\langle D_sW, \frac{\dot{z}_m}{\langle U(z_m), \dot{z}_m\rangle}\right\rangle \mathrm{d}s \\ & \quad = \int_0^1 \left\langle D_sW - \frac{\langle D_sW, \dot{z}_m\rangle}{\langle U(z_m), \dot{z}_m\rangle}U(z_m), \alpha_m\right\rangle \mathrm{d}s + \int_0^1 \langle B(z_m)[\alpha_m, \dot{z}_m] + \hat{\beta}_m, W\rangle\,\mathrm{d}s \\ & -\int_0^1\int_0^s \left(\frac{\langle D_\sigma W, \dot{z}_m\rangle}{\langle U(z_m), \dot{z}_m\rangle}\mathrm{d}\sigma\right)\langle B(z_m)[\alpha_m, \dot{z}_m]\hat{\beta}_m, U(z_m)\rangle\,\mathrm{d}s.\end{aligned}$$

Now, since $z_m([0, 1]) \subset K$ for all $m \in \mathbb{N}$,

$$U(z_m) \text{ and } \frac{\dot z_m}{\langle U(z_m), \dot z_m\rangle} \text{ are uniformly bounded.}$$

Moreover, by Lemma 2.6 the sequence $(\dot z_m)_{m\in\mathbb{N}}$ is bounded in $L^1([0, 1], T\mathcal{M})$. Since α_m and $\hat\beta_m \to 0$ uniformly, the covariant primitive

$$\int_0^s (B(z_m)[\alpha_m, \dot z_m] + \hat\beta_m)\, d\sigma$$

tends uniformly to 0. Therefore, an integration by parts shows the existence of a vector field A_m along z_m such that A_m tends uniformly to 0 and

$$\int_0^1 \left\langle D_s W, \frac{\dot z_m}{\langle U(z_m), \dot z_m\rangle}\right\rangle ds + \int_0^1 \langle D_s W, A_m\rangle\, ds = 0$$

for any vector field $W \in C_0^\infty([0, 1], T\mathcal{M})$ such that $W(s) \in T_{z_m(s)}\mathcal{M}$ for any s.

The arbitrariness of W gives the existence of a vector field $Z_m \in T_{z_m}\hat{\mathcal{L}}^+_{q,\delta,\epsilon}$ such that

$$D_s Z_m = 0 \quad \text{and} \quad \frac{\dot z_m}{\langle U(z_m), \dot z_m\rangle} + A_m = Z_m. \tag{2.19}$$

Since $D_s Z_m = 0$, the function $C_m = \langle Z_m, Z_m\rangle$ is constant. Moreover, since $\langle \dot z_m, \dot z_m\rangle = -\epsilon^2$, we obtain the existence of a sequence of functions $\hat A_m$ such that $\hat A_m \to 0$ uniformly and

$$C_m = \frac{-\epsilon^2}{\langle U(z_m), \dot z_m\rangle} + \hat A_m. \tag{2.20}$$

We show now that the functions $\langle U(z_m), \dot z_m\rangle$ are bounded uniformly with respect to $m \in \mathbb{N}$ and $s \in [0, 1]$. Assume by contradiction that there exists a sequence $(s_m)_{m\in\mathbb{N}}$ such that $\langle U(z_m(s_m)), \dot z_m(s_m)\rangle \to +\infty$. By (2.20), $C_m \to 0$ and

$$\frac{-\epsilon^2}{\langle U(z_m), \dot z_m\rangle} \to 0 \text{ uniformly.}$$

This means that

$$|\langle U(z_m), \dot z_m\rangle| \to +\infty \text{ uniformly.} \tag{2.21}$$

Since $U(z_m)$ is an uniformly bounded sequence of time-like vector fields along the curve z_m and $\dot z_m$ is time-like, $\|\dot z_m(s)\|_{\mathrm{R}} \to +\infty$ uniformly, in contradiction with Lemma 2.6. Then $\langle U(z_m), \dot z_m\rangle$ is uniformly bounded with respect to $m \in \mathbb{N}$ and $s \in [0, 1]$, and since $U(z_m)$ and $\dot z_m$ are time-like, there exists a positive constant D such that

$$\|\dot z_m(s)\|_{\mathrm{R}} \le D, \quad \forall n \in \mathbb{N}, \ \forall s \in [0, 1]. \tag{2.22}$$

By the Ascoli–Arzelá theorem, up to subsequences, we have that the sequence $(z_m)_{m\in\mathbb{N}}$ is uniformly convergent.

Now the sequence $(C_m)_{m\in\mathbb{N}}$ converges (up to subsequences) to $C \in \mathbb{R}$. Therefore, the sequence $(\langle U(z_m), \dot{z}_m\rangle)_{m\in\mathbb{N}}$ is convergent in L^∞.

Now $\langle Z_m, Z_m\rangle$ is bounded, z_m is uniformly convergent and $D_s Z_m = 0$. Then using (2.22) and the Ascoli–Arzelá theorem gives that the sequence Z_m has a subsequence which is uniformly convergent. By (2.19) there exists a subsequence $(\dot{z}_{m_k})_{m\in\mathbb{N}}$ which converges uniformly. □

The manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$ is only of class C^1 (cf. [6]). However, the restriction of the arrival time τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ is of class C^2. This fact is essential to develop a Morse theory on $\mathcal{L}^+_{p,\gamma,\epsilon}$ (in particular for the study of the behavior of τ nearby its critical point).

More precisely consider the C^1-bundle W_ϵ over the manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$, whose fiber $W_\epsilon(z)$ is given by the whole tangent space $T_z\Omega^{1,2}_{p,\gamma}$, $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$, namely

$$W_\epsilon(z) = \{(z,\zeta) : z \in \mathcal{L}^+_{p,\gamma,\epsilon}, \zeta \in T_z\Omega^{1,2}_{p,\gamma}\}.$$

Moreover we set

$$W^0_\epsilon = \{(z,\zeta) \in W_\epsilon : \zeta(1) = 0\}.$$

We are thinking of W_ϵ as a regular extension of tangent bundle $T\mathcal{L}^+_{p,\gamma,\epsilon}$. They are related by the bundle map $V : W_\epsilon \to T\mathcal{L}^+_{p,\gamma,\epsilon}$,

$$V(z,\zeta) = (z, V_\zeta),$$

where V_ζ is defined by (2.10).

Remark 2.7. *It is immediately checked that V is a continuous map and it is a C^1-map considered as a map from W_ϵ into itself (with image in $T\mathcal{L}^+_{p,\gamma,\epsilon}$). Moreover, its restriction to the tangent bundle $T\mathcal{L}^+_{p,\gamma,\epsilon}$ is the identity map and for every $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$, V is surjective from W^0_ϵ to $T\mathcal{L}^+_{p,\gamma,\epsilon}$.*

By Remark 2.7, the following proposition easily follows.

Proposition 2.8. *The functional τ is of class C^2 on $\mathcal{L}^+_{p,\gamma,\epsilon}$ in the sense that the map*

$$(z,\zeta) \to \tau'(z)[V_\zeta]$$

is of class C^1 on W_ϵ.

Corollary 2.9. *For any local chart of the manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$, the restriction of τ to the domain of the chart is of class C^2.*

We prove now the time-like version of the Fermat principle for curves of class $H^{1,1}$. It will be fundamental to prove Theorem 2.1.

Theorem 2.10. *A curve z is a critical point of τ on $\hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$ in the sense that $\mathrm{d}\tau(z)[\zeta] = 0$ for any $\zeta \in T_z\hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$ if and only if z is a (smooth) geodesic.*

Proof. Let $U(z)$ be the vector field along z given by (2.9). By (2.10) and (2.11), z is a critical point of τ if and only if for any vector field $W \in T_z H^{1,1}([0,1], \mathcal{M})$ such that $W(0) = 0$, $W(1) = 0$,

$$\int_0^1 \frac{\langle D_s W, \dot{z}\rangle}{\langle U(z), \dot{z}\rangle} \mathrm{d}s = 0. \tag{2.23}$$

Now, assume that z is a geodesic. Then $\langle U(z), \dot{z}\rangle$ is a constant since $D_s U = 0$ and $D_s \dot{z} = 0$. Such a constant is nonzero because $U(z)$ and $\dot{z}$ are both time-like.

Moreover, if z is a geodesic, integration by parts gives

$$\int_0^1 \langle D_s W, \dot{z}\rangle \, \mathrm{d}s = 0$$

for all $W \in H^{1,1}([0,1], T\mathcal{M})$, with $W(s) \in T_{z(s)}\mathcal{M}$ for all s, and such that $W(0) = 0$, $W(1) = 0$. Hence, (2.23) holds.

Conversely, assume that (2.23) holds. Then, setting

$$\lambda(s) = \frac{1}{\langle U(z), \dot{z}\rangle},$$

we have by an usual boot-strap argument that the vector field $\lambda(z)\dot{z}$ is of class C^1. Moreover, $D_s(\lambda \dot{z}) = 0$. Then

$$\langle \lambda \dot{z}, \lambda \dot{z}\rangle = -\lambda^2 \epsilon^2 \quad \text{is constant},$$

showing that λ is constant (and nonzero). Then $D_s \dot{z} = 0$. □

Remark 2.11. *By* (2.10) and (2.11), *z is a critical point of τ if and only if $\mu(1) = 0$, for any vector field $W \in T_z H^{1,1}([0,1], T\mathcal{M})$ along z, with $W(0) = 0$, $W(1) = 0$. Therefore, by* (2.11) *and Remark* 2.7, *z is a critical point of τ if and only if $\zeta(1) = 0$, for any $\zeta \in T_z \hat{\mathcal{L}}^+_{p,\gamma,\epsilon}$.*

We give now the statement of the well-known *Ekeland's variational principle* (cf. [4]). It will be used in the proof of Theorem 2.1.

Theorem 2.12. *Let (X, d) be a complete metric space and $E : X \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous functional, bounded from below, $E \not\equiv +\infty$.*

Then, for any $\nu, \mu > 0$ and for any $u \in X$ such that

$$E(u) \le \inf_X E + \mu,$$

there exists an element $v \in X$ strictly minimizing the functional

$$E_u(w) = E(w) + \frac{\nu}{\mu} d(u, w).$$

Moreover, we have

$$E(v) \le E(u) \quad \text{and} \quad d(u, v) \le \mu.$$

We are finally ready to prove Theorem 2.1.

Proof of Theorem 2.1. Fix $q \in \mathcal{M}$ and choose a local chart (U, φ) as in Remark 2.5 and including q. Then we can reduce us to work on the space $V \times I$, where V is a bounded open subset of $\mathbb{R}^n$, $n = m - 1$, $I =]-\lambda_0, \lambda_0[$ is an open interval of $\mathbb{R}$, $q = (q_0, 0) \in V \times I$ and the metric g satisfies (2.12). Since δ is an integral curve of the vector field Y, if $d_{\mathrm{R}}(q, \delta((\alpha+\beta)/2))$ is sufficiently small, we can assume that

$$\delta(s) = (q_\delta, s), \quad \forall s \in]-\lambda_0, \lambda_0[\subset]\alpha, \beta[,$$

where $q_\delta \in V$ and $d_{\mathrm{R}}(q_0, q_\delta) \to 0$ as $d_{\mathrm{R}}(q, \delta((\alpha+\beta)/2)) \to 0$.

If $z \in \mathcal{L}^+_{q,\delta,\epsilon}$ is a curve with values in U, unless we consider the chart $\varphi(U) = V \times I$, $z = (x, t)$, $x(a) = q$, $x(b) = q_\delta$ and t satisfies the Cauchy problem

$$\dot{t} = \left\langle \frac{\Gamma}{\beta}(x,t), \dot{x} \right\rangle + \sqrt{\left\langle \frac{\alpha}{\beta}(x,t)\dot{x}, \dot{x} \right\rangle + \left\langle \frac{\Gamma}{\beta}(x,t), \dot{x} \right\rangle^2 + \epsilon^2}, \qquad t(a) = 0 \tag{2.24}$$

Moreover,

$$\begin{aligned} \tau(z) &= t_x(b) \\ &= \int_a^b \left\langle \frac{\Gamma}{\beta}(x,t_x), \dot{x} \right\rangle + \sqrt{\left\langle \frac{\alpha}{\beta}(x,t_x)\dot{x}, \dot{x} \right\rangle + \left\langle \frac{\Gamma}{\beta}(x,t_x), \dot{x} \right\rangle^2 + \epsilon^2}\, \mathrm{d}s, \end{aligned} \tag{2.25}$$

where t_x is the solution of (2.24). Using as a test function the chord joining q_0 with q_δ in the interval $[a, b]$, we see that

$$\inf_{\mathcal{L}^+_{q,\delta,\epsilon}} \tau \to 0 \ \text{ as } \ |b-a| \to 0 \quad \text{and} \quad d_{\mathrm{R}}(q_0, q_\delta) \to 0. \tag{2.26}$$

Therefore, if $|b - a|$ and $d_{\mathrm{R}}(q_0, q_\delta)$ are sufficiently small,

$$\text{any minimizing sequence } (z_m)_{m\in\mathbb{N}} \text{ for } \tau \text{ in } \mathcal{L}^+_{q,\delta,\epsilon} \text{ is contained in } \varphi(U). \tag{2.27}$$

The Cauchy problem (2.24) can obviously be written as

$$\dot{t} = \langle A(x,t), \dot{x} \rangle + \sqrt{\langle L(x,t)\dot{x}, \dot{x}\rangle^2 + \epsilon^2}, \qquad t(0) = 0,$$

where L is a smooth definite operator and A is a smooth vector field. Using the above position and the Gronwall lemma shows that the map $\Phi : H^{1,1}([a,b], \mathbb{R}^n) \to L^1([0,1], \mathbb{R})$ such that $\Phi(x)$ is the unique solution of (2.25) (whenever it is defined in all the interval $[a, b]$) is a continuous map (cf. also [5]).

We claim that for any $\zeta \in C^1([0,1], \mathbb{R}^n)$,

$$\Phi \text{ is differentiable along the direction } \zeta. \tag{2.28}$$

Towards this goal consider the map

$$G(x,t) = \dot{t} - \langle A(x,t), \dot{x}\rangle - \sqrt{\langle L(x,t)\dot{x}, \dot{x}\rangle^2 + \epsilon^2}.$$

Fix ζ of class C^1. $G(x, \Phi(x)) = 0$ and for any $\lambda \in \mathbb{R}$, $G(x + \lambda\zeta, \Phi(x + \lambda\zeta)) = 0$. Since ζ is of class C^1, straightforward computations shows that there exists

$$\lim_{\lambda\to 0} \frac{G(x + \lambda\zeta, t) - G(x, t)}{\lambda} = \frac{\partial G}{\partial x}(x, t)[\zeta] \text{ uniformly in } t$$

with respect to the L^1-norm, and

$$\frac{\partial G}{\partial x}(x + \sigma\lambda\zeta, \Phi(x + \lambda\zeta))[\zeta] \to \frac{\partial G}{\partial x}(x, \Phi(x))[\zeta] \text{ in } L^1$$

as $\lambda \to 0$ uniformly on $\sigma \in [0, 1]$.

Moreover, for any $\theta \in H^{1,1}([0, 1], \mathbb{R})$,

$$\frac{\partial G}{\partial t}(x, t)[\theta] = \dot\theta - \left\langle \frac{\partial A}{\partial t}(x, t), \dot x \right\rangle \theta - \frac{1}{2\sqrt{\langle\langle L(x, t)\dot x, \dot x\rangle + \epsilon^2\rangle}} \left\langle \frac{\partial L}{\partial t}(x, t)\dot x, \dot x \right\rangle \theta.$$

This allows to show that the map

$$\frac{\partial G}{\partial t} : H^{1,1}([0, 1], \mathbb{R}) \to L^1([0, 1], \mathbb{R})$$

is invertible (the inverse can be evaluated solving a linear ordinary differential equation) and

$$\left[\frac{\partial G}{\partial t}(x, \Phi(x) + \sigma(\Phi(x + \lambda\zeta) - \Phi(x)))\right]^{-1} \to \left[\frac{\partial G}{\partial t}(x, \Phi(x))\right]^{-1}$$

in $H^{1,1}([0, 1], \mathbb{R})$ (uniformly with respect to σ, because $\Phi(x+\lambda\zeta) \to \Phi(x)$ in $L^\infty([0, 1], \mathbb{R})$.

Now, since $G(x, \Phi(x)) = 0$ and $G(x + \lambda\zeta, \Phi(x + \lambda\zeta)) = 0$, applying the Lagrange theorem we obtain

$$\begin{aligned} 0 = {} & \frac{\partial G}{\partial x}(x + \sigma_1\lambda\zeta, \Phi(x + \lambda\zeta))[\lambda\zeta] + \frac{\partial G}{\partial t}(x, \Phi(x) + \sigma_2(\Phi(x + \lambda\zeta) \\ & -\Phi(x)))[\Phi(x + \lambda\zeta) - \Phi(x)]. \end{aligned}$$

Dividing by λ and passing to the limit as $\lambda \to 0$ gives (2.28).

Take a sequence $(\nu_m)_{m\in\mathbb{N}}$ of positive numbers such that $\nu_m \to 0$. By virtue of (2.27), for any $m \in \mathbb{N}$ we can choose a curve x_m with support contained in V such that

$$\tau(x_m) \le \inf_{\mathcal{L}^+_{q,\delta,\epsilon}} \tau + \nu_m^2.$$

In Theorem 2.12 choose $\nu = \nu_m^2$, $\mu = \nu_m$ and $u = x_m$. Since V is relatively compact, by (2.27) we can assume to be on a complete metric space. So, by applying Theorem 2.12 we find a point y_m satisfying

$$\tau(y_m) \le \tau(y_m + w) + \nu_m \|w\|_1 \tag{2.29}$$

for any $w \in H^{1,1}([a, b], V)$, and therefore for any $w \in C^1([a, b], V)$. Now by a density argument, y_m can be chosen of class C^1. Then, by the arbitrariness of w, since τ is differentiable (in $H^{1,1}$) along the directions of class C^1, we deduce that

$$|\mathrm{d}\tau(y_m)[\zeta]| \le \epsilon_m \to 0 \tag{2.30}$$

for any ζ of class C^1 such that $\|\zeta\|_1 \le 1$.

Indeed, taking $w = \lambda\zeta$ in (2.28) we have

$$\frac{\tau(y_m) - \tau(y_m + \lambda\zeta)}{|\lambda|\|\zeta\|_1} = \frac{\tau(y_m) - \tau(y_m + \lambda\zeta)}{|\lambda|} \leq \epsilon_m,$$

from which we deduce (2.30) sending $\lambda \to 0$ (first choosing $\lambda > 0$ and then $\lambda < 0$). Note that y_m is a minimizing sequence (by Theorem 2.12).

Now by the uniqueness of the related Cauchy problems, we see that

$$\{(\zeta, \mathrm{d}\Phi(y_m)\zeta) : \zeta \in C^1([0,1], \mathbb{R}^n)\} = T_{y_m}\hat{\mathcal{L}}^+_{q,\delta,\epsilon} \cap C^1([0,1], T\mathcal{M}).$$

Then due to the density of C^1 in $H^{1,1}$ we see that the sequence $(y_m, \Phi(y_m))_{m\in\mathbb{N}}$ is a minimizing sequence for τ satisfying the assumptions of Proposition 2.4. Then by Proposition 2.4, there exists a subsequence of $(y_m)_{m\in\mathbb{N}}$ convergent to a curve y with respect to the C^1-topology. Then $(y, \Phi(y))$ is a C^1-curve minimizing τ on $\hat{\mathcal{L}}^+_{q,\delta,\epsilon}$. Finally, by Theorem 2.10 we obtain that $(y, \Phi(y))$ is a geodesic, while the uniqueness of the minimizer comes from the local invertibility of the exponential map. □

Remark 2.13. *Working in local coordinates shows immediately that for any fixed neighborhood $\mathcal{U}_q$ of q, there exists a positive number ρ_q such that the minimal geodesic for τ on $\mathcal{L}^+_{q,\delta,\epsilon}([a,b])$ is in $\mathcal{U}_q$ if*

$$d_{\mathrm{R}}\left(q, \delta\left(\frac{\alpha+\beta}{2}\right)\right) \leq \rho(q) \quad \text{and} \quad |b-a| \leq \rho(q).$$

3. A shortening method for τ on $\mathcal{L}^+_{q,\gamma,\epsilon}$

In this section we shall introduce a shortening flow for the functional $\tau(z)$. Such a flow will be used to get the deformations for the sublevels of τ (needed to develop a Morse theory) when we are far from the critical points of τ, i.e., time-like geodesics.

To construct the shortening flow we shall use the same ideas as in [14], adapting them to our case. Note that here we cannot use the same finite dimensional approach nearby critical curves (used in [14] for Riemannian geodesics) because we are not working with fixed point boundary conditions.

The shortening procedure, which is illustrated by five pictures appearing at the end of the paper, is constructed in the following way.

Fix $c > \inf\{\tau(z), z \in \mathcal{L}^+_{p,\gamma,\epsilon}\}$ and consider $D(c)$ as in Lemma 2.6. Let K_c be a compact subset of $\mathcal{M}$ including all the curves $z \in \mathcal{L}^+_{q,\gamma,\epsilon}$ such that $\tau(z) \leq c$.

Let $\rho_*(c) > 0$ be such that Theorem 2.1 holds with $\rho(q)$ replaced by $\rho_*(c)$ for any $q \in K_c$. Take $N = N(c)$ such that

$$\frac{1}{N} \leq \rho_*(c), \qquad \frac{D(c)}{N} \leq \rho_*(c).$$

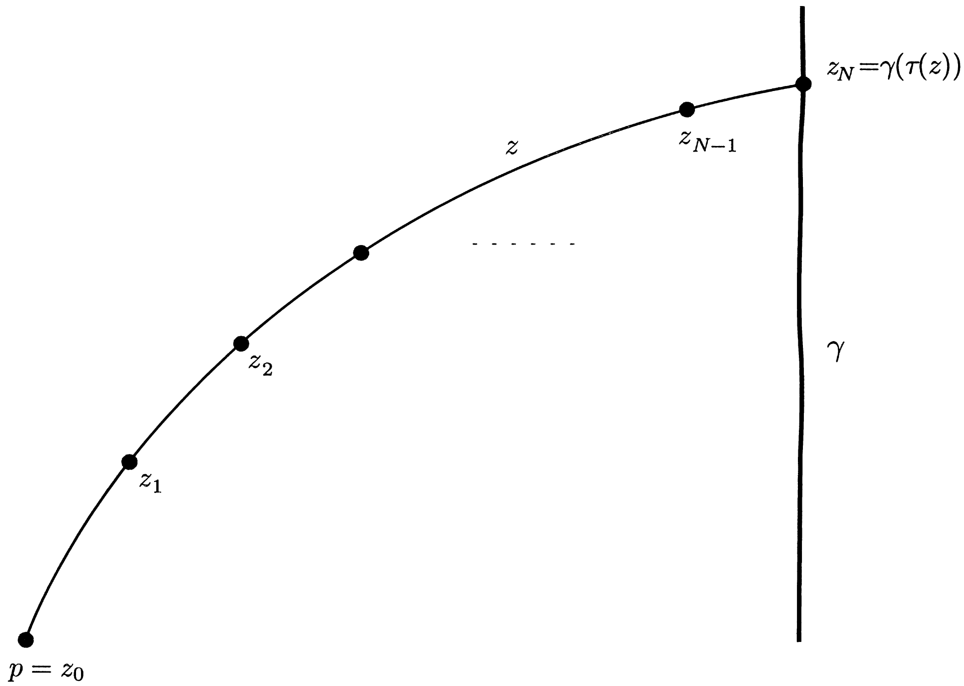

Fig. 1.

Choose a partition $\{0 = s_0 < s_1 < \cdots < s_{N-1} < s_N = 1\}$ of $[0, 1]$ such that for any $i \in \{1, \ldots, N\}$,

$$s_i - s_{i-1} = \frac{1}{N}.$$

For any $z \in \tau^c \cap \mathcal{L}^+_{p,\gamma,\epsilon}$, choose $N + 1$ points $z_0, z_1, \ldots, z_N$ on $z([0, 1])$ such that $z(0) = p$, $z_N = z(1)$ and $d_{\mathrm{R}}(z_i, z_{i-1}) = l(z)/N$, for any $i \in \{1, \ldots, N\}$, where $l(z)$ denotes the length of z with respect to the Riemannian structure (2.3) (see Fig. 1).

Denote by γ_i $(i = 1, \ldots, N)$ the maximal integral curve of W such that $\gamma_i(0) = z_i$ (see Fig. 2). Observe that $\gamma_N(s) = \gamma(s + \tau(z))$ for all s.

Let w_1 be the geodesic minimizing τ on $\mathcal{L}^+_{p,\gamma_1,\epsilon}([s_0, s_1])$ (recall that $z_0 = p$ and $s_0 = 0$), w_2 the light-like geodesic minimizing τ on $\mathcal{L}^+_{w_1(s_1),\gamma_2,\epsilon}([s_1, s_2])$, and so on (see Fig. 3).

In Figs. 3–5 the points $w_i(s_i)$ are denoted by $\overline{w}_i$. Note that $w_i(s_i)$ is located “under” z_i. This can be seen under a comparison theorem in ordinary differential equations using the “spatial coordinate” of z to compare the solutions of (2.24) having initial data z_{i-1} and $w_{i-1}(s_{i-1})$, respectively, and recalling that w_i is a minimizer. Note also that the number N can be chosen large enough so that $d_{\mathrm{R}}(w_i(s_i), z_{i+1}) \le \rho_*(c)$ for any $i = 1, \ldots, N - 1$ and for any $z \in \tau^c$.

Remark 3.1. *Let $K = K(c)$ be a compact subset of $\mathcal{M}$ containing the images of the curves of the curves $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$ with $\tau(z) \le c$. By compactness, $K(c)$ can be covered*

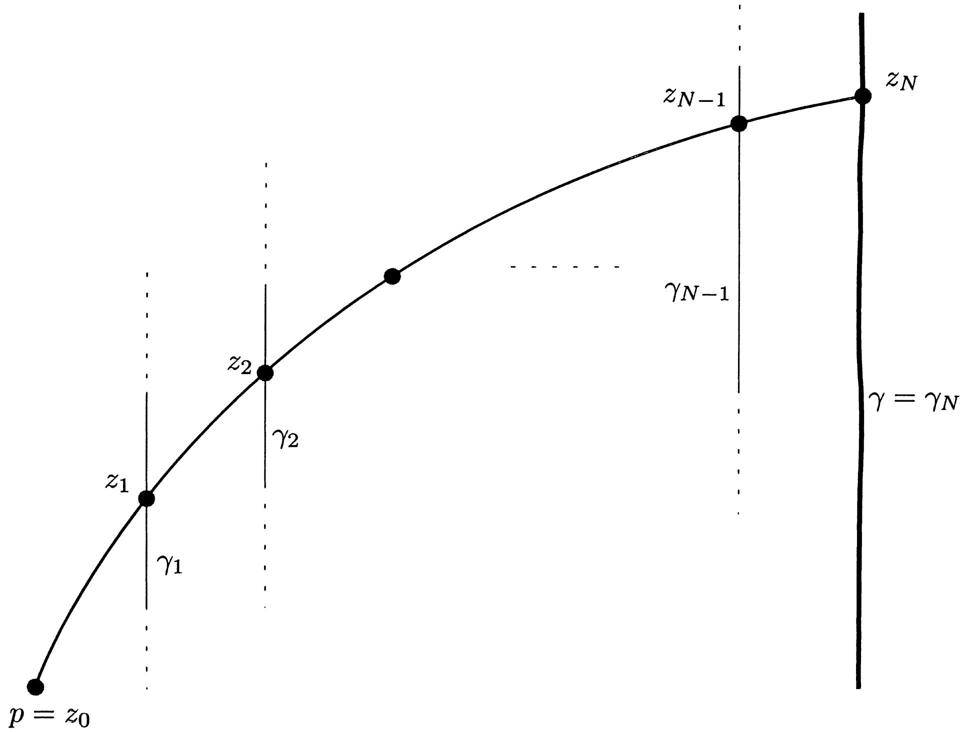

Fig. 2.

by a finite family (U_j) *as in Remark* 2.5, *and the Lorentzian metric* g *is described by* (2.12).

Moreover, N can be chosen so large that $z([s_{i-1}, s_i])$ and the minimizer of τ on $\mathcal{L}^+_{w_{i-1}(s_{i-1}),\gamma_i,\epsilon}([s_{i-1}, s_i])$ are contained in some U_j.

With the notation of Remark 2.5, for any future pointing curve z with image contained in some U_j, the condition $\langle \dot z, \dot z\rangle = -\epsilon^2$ holds if and only if

$$\dot t = \left\langle \frac{\Gamma_j}{\beta_j}(x,t), \dot x \right\rangle_0 + \sqrt{\left\langle \frac{\alpha_j}{\beta_j}(x,t), \dot x, \dot x \right\rangle_0 + \left\langle \frac{\Gamma_j}{\beta_j}(x,t), \dot x \right\rangle_0^2 + \epsilon^2}. \tag{3.1}$$

Moreover, any γ_i is an integral curve of W, so, in U_j, it has the form $s \mapsto (x_j, t_j + s)$ if $z_j = (x_j, t_j)$.

Note that $\mathcal{L}^+_{p,\gamma_1,\epsilon}([s_0, s_1])$ is nonempty since it contains the restriction $z_{|[s_0,s_1]}$.

Now, using elementary comparison theorems for ordinary differential equations allow to deduce that also any space $\mathcal{L}^+_{w_{i-1}(s_{i-1}),\gamma_i,\epsilon}([s_{i-1}, s_i])$ is nonempty for any $i \in \{2, \dots, N\}$.

Note also that if η_1 is the curve defined by setting $\eta_1([s_{i-1}, s_i]) = w_i$, then $\tau(\eta_1) \le \tau(z) \le c$ (always by comparison theorems in ODE). In particular $\eta_1([0, 1])$ is contained in $K(c)$.

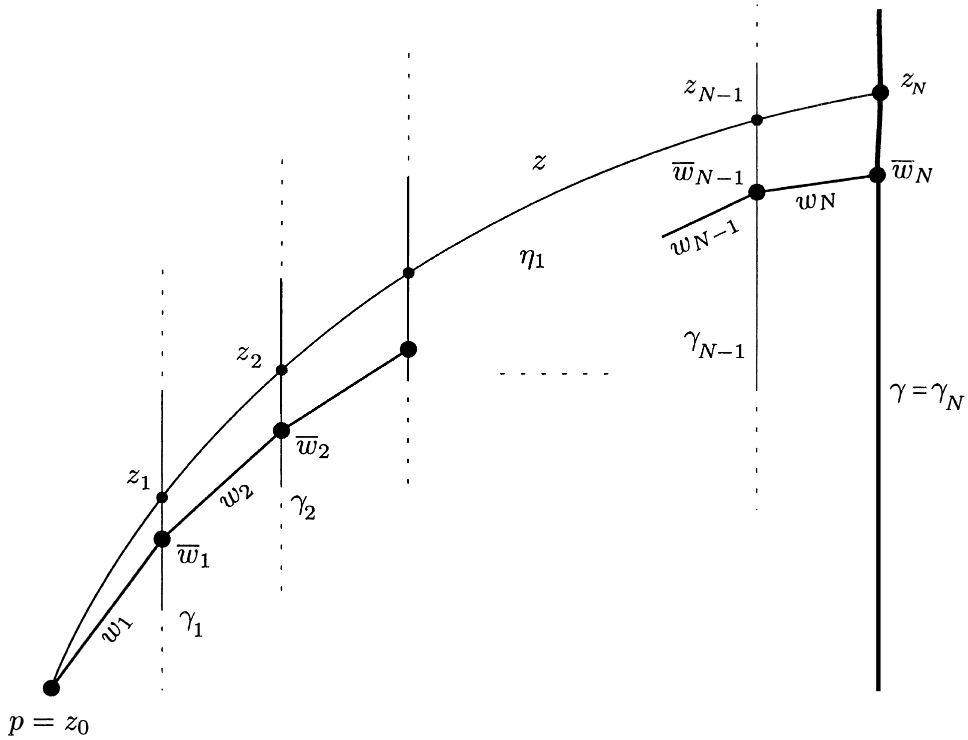

Fig. 3.

Remark 3.2. *A second curve η_2 will be constructed in the following way starting from η_1. On any minimizer w_i ($i = 1, \dots, N$) consider the point m_i such that $d(w_i(s_{i-1}), m_i) = d(m_i, w(s_i))$.*

For $i = 1, \dots, N$, we denote by λ_i the maximal integral curve of W such that $\lambda_i(0) = m_i$; moreover, we set $\lambda_{N+1}(s) = \gamma(s + \tau(\eta_1))$ (see Fig. 4).

Consider now the following subdivision of the interval [0, 1]. Let $\sigma_0 = 0$, $\sigma_1 = 1/2N$, $\sigma_j = (2j-1)/2N$ for $j = 2, \dots, N$, and $\sigma_{N+1} = 1$.

Denote by u_1 the minimizer of τ on $\mathcal{L}^+_{p,\lambda_1,\epsilon}([\sigma_0, \sigma_1])$, by u_2 the minimizer of τ on $\mathcal{L}^+_{u_1(\sigma_1),\lambda_2,\epsilon}([\sigma_1, \sigma_2])$ and so, inductively, we denote by u_j the minimizer of τ in $\mathcal{L}_{u_{j-1}(\sigma_{j-1}),\lambda_j,\epsilon}([\sigma_{j-1}, \sigma_j])$, $j = 2, \dots, N+1$.

Finally (see Fig. 5), we denote by η_2 the curve such that $\eta_2|_{[\sigma_{j-1},\sigma_j]} = u_j$.

Using again comparison theorems in ordinary differential equations one proves that $\tau(\eta_2) \le \tau(\eta_1)$.

The continuous flow $\eta(\sigma, z)$ can be constructed as follows. Fix $\sigma \in [0, 1]$ and consider for instance the interval $[s_0, s_1]$. We choose $\eta(\sigma, z)_{|[s_0,s_1]}$ as follows. Set $p = (x_0, 0)$ and $\gamma_1(s) = (x_1, t_1 + s)$ (in some neighborhood U_j as in Remark 3.1). Since $z(s) = (x(s), t(s))$, the curve $x(s)$ joins x_0 with x_1.

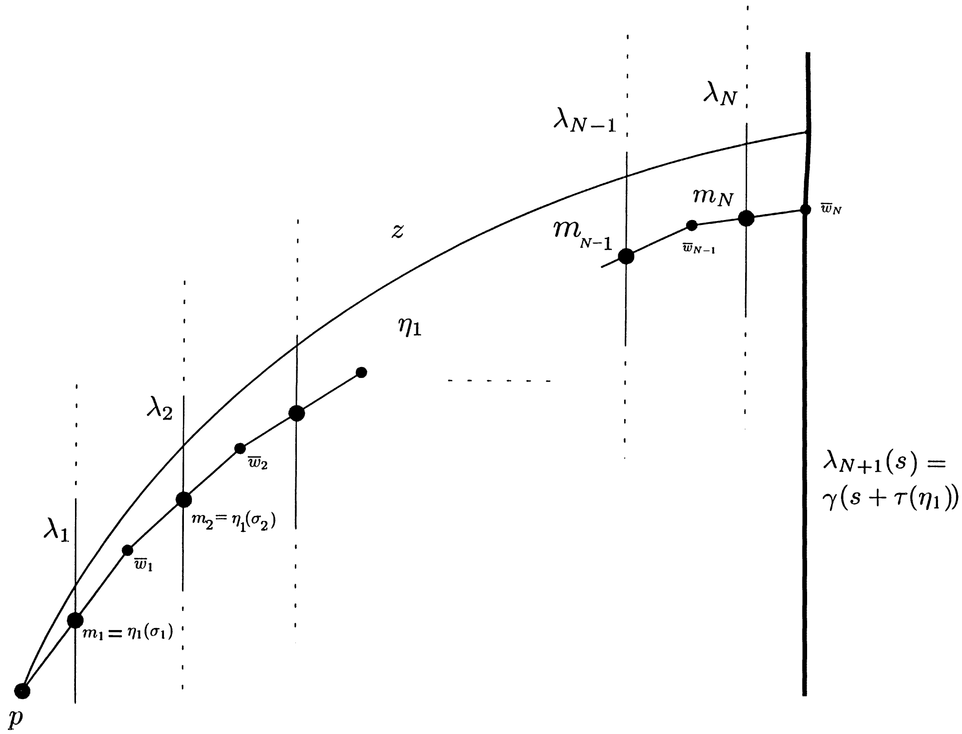

Fig. 4.

Let $y(\sigma)$ be the minimizer of the functional

$$y \mapsto \int_{s_0}^{\sigma s_1} \left\langle \frac{\Gamma_i}{\beta_i}(y, t_y), \dot{y} \right\rangle_0 \mathrm{d}s + \int_{s_0}^{\sigma s_1} \sqrt{\left\langle \frac{\alpha_i}{\beta_i}(y, t_y), \dot{y}, \dot{y} \right\rangle_0 + \left\langle \frac{\Gamma_i}{\beta_i}(y, t_y), \dot{y} \right\rangle_0^2} \mathrm{d}s \quad (3.2)$$

with boundary conditions $y(0) = x_0$ and $y(\sigma s_1) = x(\sigma s_1)$, where t_y is the solution of (3.2) with $t_y(0) = 0$ in the interval $[0, \sigma s_1]$.

Denote by $\hat{y}(\sigma)$ the extension of $y(\sigma)$ to $[s_0, s_1]$ taking $\hat{y}(s) = x(s)$ for $s \in [\sigma s_1, s_1]$. Finally, denote by $\hat{t}_y$ the corresponding solution of (3.1) in the interval $[s_0, s_1]$. The curve $(\hat{y}(\sigma), \hat{t}_y(\sigma))$ will be $\eta(\sigma, z)$ in the interval $[s_0, s_1]$. In the same way we can construct $\eta(\sigma, z)$ on the other intervals $[s_{i-1}, s_i]$. Note that, by construction, $\eta(1, z) = \eta_1$. Similarly, we can extend the flow η to a map defined on $[0, 2] \times \tau^c$ in such a way that $\eta(2, z) = \eta_2$.

Now we iterate the shortening argument above, replacing the original curve z with the curve η_2. Successively, we apply the above construction starting from η_2. By induction we obtain a flow $\eta(\sigma, z)$ defined on $\mathbb{R}^+ \times \tau^c$. Note that $\tau(\eta(\sigma, z)) \leq \tau(z)$ for any σ and for any z.

Suppose that $\tau(\eta_1) = \tau(\eta_2)$ and consider the situation is a single interval $[\sigma_j, \sigma_{j+1}]$. Since $\tau(\eta_1) = \tau(\eta_2)$ simple comparison theorems in ODE show that η_1 is a minimizer on the interval $[\sigma_j, \sigma_{j+1}]$. Suppose that it consists of two (nonconstant) light-like geodesics. If it is not a light-like geodesic, by the above construction it has a discontinuity at $s_{j+1} = (\sigma_{j+1} + \sigma_j)/2$. Denote by U_{η_1} the parallel transport of $\dot{\gamma}(\tau(\eta_1))$ along the curve η_1. Since

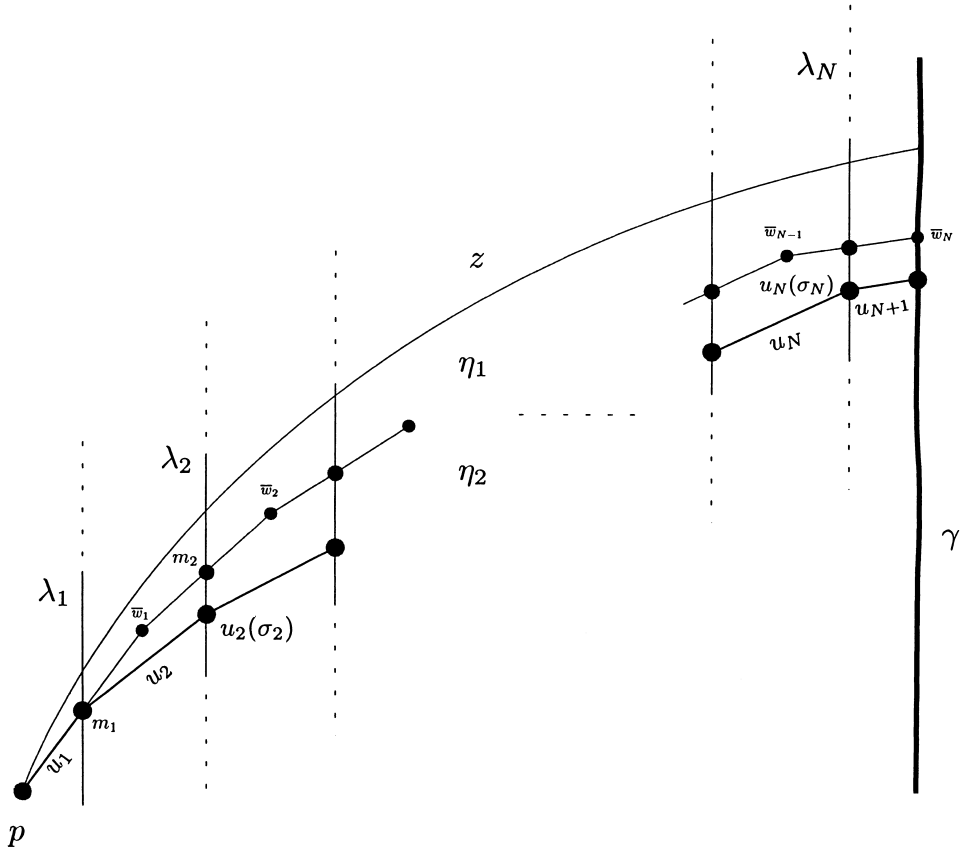

Fig. 5.

η_1 is a minimizer, by (2.23),

$$\int_{\sigma_j}^{\sigma_{j+1}} \frac{\langle D_s V, \dot\eta_1\rangle}{\langle U_{\eta_1}, \dot\eta_1\rangle} \mathrm{d}s = 0$$

for any C^∞-vector field along η_1 such that $V(0) = 0$, $V(1) = 0$. In particular $\frac{\dot\eta_1}{\langle U_{\eta_1}, \dot\eta_1\rangle}$ is a C^1 curve and also

$$\frac{-\epsilon^2}{\langle U_{\eta_1}, \dot\eta_1\rangle^2}$$

is of class C^1, and this implies that η_1 is of class C^1 because $\langle U_{\eta_1}, \dot\eta_1\rangle$ never changes its sign.

Then, whenever we are far from critical points of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$, $\tau(\eta_2) < \tau(\eta_1)$.

Finally compactness arguments similar to the ones used for the shortening method for Riemannian geodesics (cf. [14]) allows to obtain the analog of the classical deformation results (cf e.g. [12,24]) for the functional τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$.

For any $d \in \mathbb{R}$ set $\tau^d = \{z \in \mathcal{L}^+_{p,\gamma,\epsilon} : \tau(z) \le d\}$.

Proposition 3.3. *Let c be a regular value for τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ (namely $\tau^{-1}(\{c\})$ does not contain geodesics).*

Then, there exists a positive number $\delta = \delta(c)$ and a continuous map $H \in C^0([0,1] \times \tau^{c+\delta}, \tau^{c+\delta})$ such that
1. $H(0, z) = z$ *for every* $z \in \tau^{c+\delta}$;
2. $H(1, \tau^{c+\delta}) \subseteq \tau^{c-\delta}$;
3. $H(\sigma, z) \in \tau^{c-\delta}$ *for any* $\sigma \in [0, 1]$ *and* $z \in \tau^{c-\delta}$.

Proposition 3.4. *Let Z_c be the set of the time-like geodesics on $\tau^{-1}(\{c\}) \cap \mathcal{L}^+_{p,\gamma,\epsilon}$. Then for any open neighborhood $\mathcal{U}$ of Z_c, there exists a positive number $\delta = \delta(\mathcal{U}, c)$ and a homotopy $H \in C^0([0,1] \times \tau^{c+\delta}, \tau^{c+\delta})$ such that*
1. $H(0, z) = z$ *for any* $z \in \tau^{c+\delta}$;
2. $H(1, \tau^{c+\delta} \setminus \mathcal{U}) \subset \tau^{c-\delta}$;
3. $H(\sigma, z) \in \tau^{c-\delta}$ *for every* $\sigma \in [0, 1]$ *and* $z \in \tau^{c-\delta}$.

Remark 3.5. *There are two main differences between the shortening method described above and the classical shortening method for Riemannian geodesics. In our case, we locally minimize a functional which is is not given in an integral form. Secondly, we minimize the functional in the space of curves joining a point with a curve, and not two fixed points.*

Remark 3.6. *The flow used in proving Propositions* 3.3 *and* 3.4 *are just what we need for a Ljusternik–Schnirelmann theory. Then, without using the nondegeneracy assumption of Theorem* 1.3 *we can obtain the existence of at least* $\mathrm{cat}(\mathcal{L}^+_{p,\gamma,\epsilon})$ *future pointing time-like geodesics in* $\mathcal{L}^+_{p,\gamma,\epsilon}$. (Here cat X denotes the minimal number of contractible subsets of X covering it.) Moreover, if $\mathrm{cat}(\mathcal{L}^+_{p,\gamma,\epsilon}) = +\infty$ there is a sequence z_n of future pointing time-like geodesics in $\mathcal{L}^+_{p,\gamma,\epsilon}$ such that $\tau(z_n) \to +\infty$. (Recall that we are assuming that γ is defined on $\mathbb{R}$.)

4. The index theorem and the Morse relations on $\mathcal{L}^+_{p,\gamma,\epsilon}$

In this section we shall prove the Morse relations on $\mathcal{L}^+_{p,\gamma,\epsilon}$ and the second part of Theorem 1.2, namely Theorem 4.1.

Theorem 4.1. *Let z be a geodesic in $\mathcal{L}^+_{p,\gamma,\epsilon}$ such that $z(1)$ is nonconjugate to p along z. Then*

$$\mu(z) = m(z, \tau),$$

where $\mu(z)$ is the geometric index of z and $m(z, \tau)$ is the Morse index of z considered as a critical point of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$.

In order to prove Theorem 4.1, we first need to evaluate the Hessian of τ at z,

$$H_\tau(z)[\zeta,\zeta] = \frac{\mathrm{d}^2}{\mathrm{d}\sigma^2}(\tau(\eta(\sigma,\cdot)))_{\sigma=0},$$

where $\zeta \in T_z\mathcal{L}^+_{p,\gamma,\epsilon}$ and $\eta :]-\sigma_0,\sigma_0[\to \mathcal{L}^+_{p,\gamma,\epsilon}$ is a variation of z with variational vector field ζ, i.e.

$$\eta(0,s) = z(s) \ \text{ for any } s \in [0,1], \qquad \eta_\sigma(0,s) = \zeta(s) \ \text{ for any } s \in [0,1].$$

Here η_σ denotes the partial derivative with respect to σ.

Proposition 4.2. *In the notation above, for all* $\zeta \in T_z\mathcal{L}^+_{p,\gamma,\epsilon}$,

$$H_\tau(z)[\zeta,\zeta] = \frac{-1}{\langle\dot\gamma(\tau(z)),\dot z(1)\rangle}\int_0^1 (\langle D_s\zeta, D_s\zeta\rangle - \langle R(\zeta,\dot z)\dot z,\zeta\rangle)\,\mathrm{d}s. \tag{4.1}$$

Proof. Since $\eta(\sigma,\cdot) \in \mathcal{L}^+_{p,\gamma,\epsilon}$ for any σ, we have

$$\langle \eta_s(s,\sigma),\eta_s(s,\sigma)\rangle = -\epsilon^2, \quad \text{for any } s \text{ and for any } \sigma.$$

Here η_s denotes the partial derivative of η with respect to s. Since z is of class C^2, it suffices to prove (4.1) whenever ζ (and therefore η) is of class C^2 and apply standard density arguments. We have

$$\frac{\partial}{\partial\sigma}\left(\int_0^1 \langle\eta_s,\eta_s\rangle\,\mathrm{d}s\right) = 0$$

and therefore

$$\begin{aligned} 0 &= \int_0^1 \langle D_\sigma\eta_s,\eta_s\rangle\,\mathrm{d}s = \int_0^1 \langle D_s\eta_\sigma,\eta_s\rangle\,\mathrm{d}s \\ &= \langle\eta_\sigma(\sigma,1),\eta_s(\sigma,1)\rangle - \langle\eta_\sigma(\sigma,0),\eta_s(\sigma,0)\rangle - \int_0^1\langle\eta_\sigma, D_s\eta_s\rangle\,\mathrm{d}s. \end{aligned} \tag{4.2}$$

Now, since $\gamma(\tau(\eta(\sigma,\cdot))) = \eta(\sigma,1)$, we have

$$\dot\gamma(\tau(\eta(\sigma,\cdot)))\frac{\mathrm{d}\tau}{\mathrm{d}\sigma}(\eta(\sigma,\cdot)) = \eta_\sigma(\sigma,1),$$

therefore, since $\eta_\sigma(\sigma,0) = 0$ for any σ, by (4.2) we have

$$\begin{aligned} \frac{\mathrm{d}\tau}{\mathrm{d}\sigma}(\eta(\sigma,\cdot)) &= \frac{\langle\eta_\sigma(\sigma,1),\eta_s(\sigma,1)\rangle}{\langle\dot\gamma(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle} \\ &= \frac{1}{\langle\dot\gamma(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\int_0^1\langle\eta_\sigma, D_s\eta_s\rangle\,\mathrm{d}s. \end{aligned}$$

Note that $\langle\dot\gamma(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle \neq 0$ because both $\dot\gamma(\tau(\eta(\sigma,\cdot)))$ and $\eta_s(\sigma,1)$ are time-like vectors.

Then, since $D_s\dot{z}=0$, we get

$$\begin{aligned}\frac{\mathrm{d}^2\tau}{\mathrm{d}\sigma^2}(\eta(\sigma,\cdot))_{|\sigma=0} &= \frac{\mathrm{d}}{\mathrm{d}\sigma}\left(\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\right)\int_0^1\langle\zeta,D_s\eta_s\rangle\,\mathrm{d}s\\ &\quad+\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\frac{\mathrm{d}}{\mathrm{d}\sigma}\left(\int_0^1\langle\eta_\sigma,D_s\eta_s\rangle\,\mathrm{d}s\right)_{\sigma=0}\\ &=\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\\ &\quad\times\left(\int_0^1(\langle D_\sigma\eta_\sigma,D_s\eta_s\rangle+\langle\eta_\sigma,D_\sigma D_s\eta_s\rangle)\,\mathrm{d}s\right)_{\sigma=0}\\ &=\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\left(\int_0^1\langle\eta_\sigma,D_\sigma D_s\eta_s\rangle\,\mathrm{d}s\right)_{\sigma=0}.\end{aligned}$$

Since $D_\sigma D_s\eta_s = D_s D_\sigma\eta_s + R(\eta_\sigma,\eta_s)\eta_s$ (cf. [1]), we have

$$\begin{aligned}H^\tau(z)[\zeta,\zeta] &= \frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\int_0^1(\langle\eta_\sigma,D_sD_\sigma\eta_s+R(\eta_\sigma\eta_s)\eta_s\rangle\,\mathrm{d}s)_{\sigma=0}\\ &=\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}(\langle\eta_\sigma(\sigma,1),D_\sigma\eta_s(\sigma,1)\rangle\\ &\quad-\langle\eta_\sigma(\sigma,0),D_\sigma\eta_s(\sigma,0)\rangle)_{\sigma=0}+\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\\ &\quad\times\left[\left(-\int_0^1\langle D_s\eta_\sigma,D_\sigma\eta_s\rangle\,\mathrm{d}s+\int_0^1\langle R(\eta_\sigma,\eta_s)\eta_s,\eta_\sigma\rangle\,\mathrm{d}s\right)\right]_{\sigma=0}\\ &=\frac{1}{\langle\dot{\gamma}(\tau(\eta(\sigma,\cdot))),\eta_s(\sigma,1)\rangle}\Bigg(\langle\zeta(1),D_{\zeta(1)}\dot{z}(1)\rangle-\langle\zeta(0),D_{\zeta(0)}\dot{z}(0)\rangle\\ &\quad-\int_0^1\langle D_s\zeta,D_s\zeta\rangle\,\mathrm{d}s+\int_0^1\langle R(\zeta,\dot{z})\dot{z},\zeta\rangle\,\mathrm{d}s\Bigg).\end{aligned}$$

Finally, $\zeta(0)=0$ and by Remark 2.11, $\zeta(1)=0$. □

Let z be a geodesic in $\mathcal{L}^+_{p,\gamma,\epsilon}$. For any $\theta\in\,]0,1]$, set

$$\begin{aligned}&A_\theta=\{\zeta\in H^{1,2}([0,\theta],T\mathcal{M}):\zeta(s)\in T_{z(s)}\mathcal{M}\ \text{ for any }\ s\in[0,\theta],\\ &\langle D_s\zeta,\dot{z}\rangle=0\ \text{ a.e.},\ \zeta(0)=0,\ \zeta(\theta)=0\}.\end{aligned}$$

We consider the bilinear form on A_θ given by

$$J_\theta(z)[\zeta,\zeta]=\int_0^\theta(\langle D_s\zeta,D_s\zeta\rangle-\langle R(\zeta,\dot{z})\dot{z},\zeta\rangle)\,\mathrm{d}s. \tag{4.3}$$

Lemma 4.3. *Let $\zeta_0\in A_\theta$. In the above notations, $J_\theta(z)[\zeta_0,\cdot]=0$ in A_θ, if and only if ζ_0 solves* (1.2) *in* $[0,\theta]$.

Proof. Let $V \in C_0^\infty([0,\theta], T\mathcal{M})$ such that $V(s) \in T_{z(s)}\mathcal{M}$, for any $s \in [0,\theta]$. Since z is a geodesic, we can describe all the elements of A_θ by

$$\zeta = V + \frac{\langle V, \dot{z}\rangle}{\epsilon^2}\dot{z}.$$

Indeed, $\zeta(0) = 0$, $\zeta(\theta) = 0$ and $\langle D_s\zeta, \dot{z}\rangle = 0$ because $\langle \dot{z}, \dot{z}\rangle = -\epsilon^2$ and $D_s\dot{z} = 0$.

Then, $J_\theta(z)[\zeta_0, \zeta] = 0$ for any $\zeta \in A_\theta$ if and only if

$$\int_0^\theta \left(\left\langle D_s\zeta_0, D_sV + \frac{\langle D_sV, \dot{z}\rangle\dot{z}}{\epsilon^2}\dot{z}\right\rangle - \left\langle R(\zeta_0, \dot{z})\dot{z}, V + \frac{\langle V, \dot{z}\rangle\dot{z}}{\epsilon^2}\right\rangle\right) \mathrm{d}s = 0$$

for any $V \in C_0^\infty([0,1], \mathcal{M})$ such that $V(s) \in T_{z(s)}\mathcal{M}$ for any $s \in [0,\theta]$.

But $\langle D_s\zeta_0, \dot{z}\rangle = 0$ because $\zeta_0 \in A_\theta$ and $\langle R(\zeta_0, \dot{z})\dot{z}, \dot{z}\rangle = 0$ by well-known properties of the Riemann tensor. Therefore, $J_\theta(z)[\zeta_0, \zeta] = 0$ for any $\zeta \in A_\theta$ if and only if

$$\int_0^\theta (\langle D_s\zeta_0, D_sV\rangle - \langle R(\zeta_0, \dot{z})\dot{z}, V\rangle)\,\mathrm{d}s = 0$$

for any $V \in C_0^\infty([0,1], \mathcal{M})$ with $V(s) \in T_{z(s)}\mathcal{M}$ for any $s \in [0,\theta]$. Then, an integration by parts completes the proof. □

We are finally ready to prove Theorem 4.1.

Proof of Theorem 4.1. Recalling (2.4), since $\dot{z}$ is a time-like vector field, a simple compactness argument shows the existence of $\nu = \nu(z) > 0$ such that

$$\langle w, w\rangle \geq \nu(z)\langle w, w\rangle_{(\mathrm{R})}$$

for any vector field w along z such that $\langle w(s), \dot{z}(s)\rangle = 0$ for any s. Moreover, since γ is an integral curve of Y and $\dot{z}(1)$ is future pointing, $\langle\dot{\gamma}(\tau(z)), \dot{z}(1)\rangle < 0$. Therefore, by (4.1), for any $\theta \in]0,1]$ the linear operator associated to the bilinear form J_θ is a compact perturbation of the identity operator if we equip A_θ with the natural Riemannian structure given by

$$\int_0^\theta \langle D_s^{(\mathrm{R})}\zeta, D_s^{(\mathrm{R})}\zeta\rangle_{(\mathrm{R})}\,\mathrm{d}s.$$

Then we can use the methods of Milnor in [14] (cf. also [11]) and Lemma 4.3 to conclude the proof. □

Now we can prove the classical Morse relations on the sublevels of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$. They can be stated in the following way. For any $b \in \mathbb{R} \cup \{+\infty\}$ set,

$$\begin{aligned}\mathcal{G}^{+,b}_{p,\gamma,\epsilon} = \{\, & z \in C^2([0,1], \mathcal{M}) : z \text{ is a future pointing geodesic such that } z(0) \\ & = p, z(1) \in \gamma(\mathbb{R}), \langle\dot{z}, \dot{z}\rangle \equiv -\epsilon^2, \tau(z) \leq b\}.\end{aligned}$$

Theorem 4.4. *Assume that* $(\mathcal{M}, \langle\cdot,\cdot\rangle)$ *is strongly causal and that assumptions* 1—3 *of Theorem* 1.3 *hold true. Then, for any field* $\mathcal{K}$ *and for any regular value* b *of* τ *on* $\mathcal{L}^+_{p,\gamma,\epsilon}$,

$b \in]\inf \tau, +\infty]$, there exists a formal series $\mathcal{S}(\lambda)$ with non negative integer coefficients (possibly $+\infty$ if $b = +\infty$) such that

$$\sum_{z \in \mathcal{G}^{+,b}_{p,\gamma,\epsilon}} \lambda^{\mu(z)} = \mathcal{P}_\lambda(\tau^b, \mathcal{K}) + (1+\lambda)\mathcal{S}(\lambda), \tag{4.4}$$

where $\mathcal{P}_\lambda(\tau^b, \mathcal{K})$ is the Poincaré polynomial of τ^b with coefficients in $\mathcal{K}$.

Proof. By Lemma 4.3 and assumption 2 of Theorem 1.3, any critical point of τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ is nondegenerate (and therefore isolated). Moreover, using the geodesic equation, it is not difficult to prove that for every $b \in \mathbb{R}$, the set $\mathcal{G}^{+,b}_{p,\gamma,\epsilon}$ is compact with respect to the C^2-topology. Hence, for all $b \in \mathbb{R}$, the set $\mathcal{G}^{+,b}_{p,\gamma,\epsilon}$ is finite.

By the deformation results of Propositions 3.3 and 3.4, since τ is of class C^2 on the Hilbert manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$ we can apply the classical Morse theory (cf. [3,12]) to describe the topology nearby the geodesics obtaining the classical Morse relations

$$\sum_{z \in \mathcal{G}^{+,b}_{p,\gamma,\epsilon}} \lambda^{m(z,\tau)} = \mathcal{P}_\lambda(\tau^b, \mathcal{K}) + (1+\lambda)\mathcal{S}(\lambda).$$

Here $m(z,\tau)$ denotes the Morse index of the critical points z for the functional τ in the Hilbert manifold $\mathcal{L}^+_{p,\gamma,\epsilon}$. Finally, thanks to Theorem 4.1, the Morse relations (4.4) follow. □

Proof of Theorem 1.3. By Theorem 4.4, setting $b = +\infty$ we have

$$\sum_{z \in \mathcal{G}^{+}_{p,\gamma,\epsilon}} \lambda^{\mu(z)} = \mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma,\epsilon}, \mathcal{K}) + (1+\lambda)\mathcal{S}(\lambda)$$

obtaining the proof. □

5. Some relations between $\mathcal{L}^+_{p,\gamma}$ and $\mathcal{L}^+_{p,\gamma,\epsilon}$

In this section we will discuss the method of approximation of the space $\mathcal{L}^+_{p,\gamma}$ with the regular manifolds $\mathcal{L}^+_{p,\gamma,\epsilon}$, pointing the results needed to obtain the Morse relations on $\mathcal{L}^+_{p,\gamma}$ as limit of the Morse relations on $\mathcal{L}^+_{p,\gamma,\epsilon}$.The first result, which is stated in the following proposition, is concerned with the existence of transition functions between $\mathcal{L}^+_{p,\gamma}$ and $\mathcal{L}^+_{p,\gamma,\epsilon}$.

Proposition 5.1. *Suppose that τ is pseudo-coercive in $\mathcal{L}^+_{p,\gamma}$.* Then, for any $c > \inf \tau$, *there exists a positive number $\epsilon_0 = \epsilon_0(c) > 0$ such that for every $\epsilon \in]0, \epsilon_0]$ there exist two injective maps:*

$$\phi_\epsilon : \tau^c \cap \mathcal{L}^+_{p,\gamma} \to \mathcal{L}^+_{p,\gamma,\epsilon}, \qquad \psi_\epsilon : \mathcal{L}^+_{p,\gamma,\epsilon} \to \mathcal{L}^+_{p,\gamma}$$

such that

1. *ϕ_ϵ and ψ_ϵ are continuous with respect to the $H^{1,1}$-norm;*
2. *for every $z \in \tau^c \cap \mathcal{L}^+_{p,\gamma}$, $\psi_\epsilon(\phi_\epsilon(z)) = z$;*

3. *for every* $z \in \mathcal{L}^{+}_{p,\gamma,\epsilon}$ *such that* $\tau(\psi_\epsilon(z)) \le c$, $\phi_\epsilon(\psi_\epsilon(z)) = z$;
4. *if* $\epsilon_1 < \epsilon_2$, *then* $\tau(\phi_{\epsilon_1}(z)) \le \tau(\phi_{\epsilon_2}(z))$ *and* $\tau(\psi_{\epsilon_1}(z)) \ge \tau(\psi_{\epsilon_2}(z))$;
5. $\tau(\phi_\epsilon(z)) \ge \tau(z)$ *and* $\tau(\psi_\epsilon(z)) \le \tau(z)$;
6. *there exists a positive constant* $M = M(c)$ *such that* $d_2(\phi_\epsilon(z), z) \le M \cdot \epsilon$ *for every* $z \in \tau^c \cap \mathcal{L}^{+}_{p,\gamma}$, *where* d_2 *is the metric induced by the Hilbert structure* (2.4).

Proof. We fix c and we find a compact subset K such that the support of every $z \in \tau^c \cap \mathcal{L}^{+}_{p,\gamma}$ lies in K. Let δ be a positive number such that the flow $\Phi(s, q)$ of the vector field Y is defined on $[-\delta, \delta] \times K$. By definition, the curve $\eta_q(s) = \Phi(s, q)$ is the maximal solution of the Cauchy problem:

$$\dot{\eta} = Y(\eta), \qquad \eta(0) = q.$$

For $z \in \tau^c \cap \mathcal{L}^{+}_{p,\gamma}$, we define

$$z_\epsilon(s) = \phi_\epsilon(z)(s) = \Phi(\sigma_{z,\epsilon}(s), z(s))$$

for some function $\sigma_{z,\epsilon}(s) = \sigma(s)$ on $[0, 1]$ and with values in $[0, \delta)$ to be determined in such a way that

$$\sigma_{z,\epsilon}(0) = 0$$

(which means that $z_\epsilon(0) = p$),

$$\langle \dot{z}_\epsilon, Y(z_\epsilon)\rangle < 0 \tag{5.1}$$

and

$$\langle \dot{z}_\epsilon, \dot{z}_\epsilon\rangle = -\epsilon^2.$$

Observe that any such curve automatically satisfies $z_\epsilon(1) \in \gamma(\mathbb{R})$ since γ is an integral curve of Y and $\Phi(0, z(1)) = z(1) \in \gamma(\mathbb{R})$.

We compute $\dot{z}_\epsilon$ as follows:

$$\dot{z}_\epsilon = \Phi_q[\dot{z}] + \Phi_\sigma[\dot{\sigma}] = \Phi_q[\dot{z}] + Y(z_\epsilon)\dot{\sigma},$$

where Φ_q and Φ_σ denote the partial derivatives of Φ. So, we have

$$\langle \dot{z}_\epsilon, \dot{z}_\epsilon\rangle = \langle Y, Y\rangle \dot{\sigma}^2 + 2\dot{\sigma}\langle Y(z_\epsilon), \Phi_q[\dot{z}]\rangle + \langle \Phi_q[\dot{z}], \Phi_q[\dot{z}]\rangle = -\epsilon^2. \tag{5.2}$$

Formula (5.2) contains a quadratic equation on $\dot{\sigma}$; observe that, by the wrong way Schwartz inequality, the discriminant Δ of (5.1) is positive:

$$\frac{\Delta}{4} = \langle Y(z_\epsilon), \Phi_q[\dot{z}]\rangle^2 - \langle Y(z_\epsilon), Y(z_\epsilon)\rangle \langle \Phi_q[\dot{z}], \Phi_q[\dot{z}]\rangle + \epsilon^2 \ge \epsilon^2 > 0. \tag{5.3}$$

Take the solution σ of (5.2) given by

$$\dot{\sigma} = -\langle Y(z_\epsilon), Y(z_\epsilon)\rangle^{-1}(\langle Y(z_\epsilon), \Phi_q[\dot{z}]\rangle + \tfrac{1}{2}\sqrt{\Delta}),$$

where Δ is given by (5.3). Notice that, with this choice

$$\langle \dot{z}_\epsilon, Y(z_\epsilon)\rangle = \dot{\sigma}\langle Y(z_\epsilon), Y(z_\epsilon)\rangle + \langle Y(z_\epsilon), \Phi_q[\dot{z}]\rangle = -\tfrac{1}{2}\sqrt{\Delta} < 0,$$

and (5.1) is satisfied. Observe also that the coefficients of (5.2) clearly depend continuously on ϵ. The function σ has to satisfy the Cauchy problem:

$$\dot{\sigma} = -\langle Y(z_\epsilon), Y(z_\epsilon)\rangle^{-1}(\langle Y(z_\epsilon), \Phi_q[\dot{z}]\rangle + \tfrac{1}{2}\sqrt{\Delta}), \qquad \sigma(0) = 0. \tag{5.4}$$

Observe that, for $\epsilon = 0$, (5.4) has the null solution, which is defined on the whole real line. Hence, for ϵ small enough, (5.4) admits a unique solution defined on all the interval [0, 1]. Moreover, if ϵ is chosen small enough, we can also assume that the solution σ of (5.4) takes values in $[-\delta, \delta]$ so that the curve $z_\epsilon = \Phi(\sigma, z)$ is well defined.

The construction of the map ψ_ϵ is done in a similar fashion considering the flow $\Psi(s, q)$ of the vector field Y, and setting

$$\psi_\epsilon(z)(s) = z^\epsilon(s) = \Psi(\sigma(s), z(s)),$$

where $\sigma = \sigma_{z,\epsilon}$ is to be determined with the conditions

$$\sigma(0) = 0, \quad \langle \dot{z}^\epsilon, \dot{z}^\epsilon\rangle = 0 \quad \text{and} \quad \langle \dot{z}^\epsilon, Y(z^\epsilon)\rangle \le 0.$$

An argument similar to the previous case shows the existence and the continuity properties of the map σ, which proves the first part of the proposition.

Elementary comparison theorems for ordinary differential equations allow to show that for all $z \in \mathcal{L}^+_{p,\gamma,\epsilon}$, the Cauchy problem (5.4) has solution defined on the whole interval [0, 1]. Therefore, the map ψ_ϵ is defined on the whole space $\mathcal{L}^+_{p,\gamma,\epsilon}$.

Parts 2 and 3 follows immediately from the construction of ϕ_ϵ and ψ_ϵ.

Parts 4 and 5 follows from simple comparison theorems in ODE applied to (5.4), while part 6 follows from the Gronwall's lemma. □

We need also the following proposition.

Proposition 5.2. *Let z be a geodesic in $\mathcal{L}^+_{p,\gamma}$ with $z(1)$ nonconjugate to p along z. Then there exists $\epsilon_0 > 0$ such that for any $\epsilon \in]0, \epsilon_0]$ there exists one and only one geodesic $z_\epsilon \in \mathcal{L}^+_{p,\gamma,\epsilon}$ such that*

$$\lim_{\epsilon\to 0} z_\epsilon = z_0 \text{ in the } H^{1,2}\text{-norm}.$$

Remark 5.3. *Notice that if z_ϵ converges to z_0 in the $H^{1,1}$-norm, using the Cauchy problem related to the geodesic equation we immediately get that the convergence is also with respect to the C^2-norm, i.e., uniform up to the second derivative.*

Proof. Since $z(1)$ is nonconjugate to p, the map

$$v \to \exp_p v$$

is a local diffeomorphism between a neighborhood of $\dot{z}(0)$ in $T_p\mathcal{M}$ and a neighborhood of $z(1) = \exp_p(\dot{z}(0)) \in \gamma(\mathbb{R})$ in $\mathcal{M}$. Then there exists a C^1-map $\varphi :]-\delta_0, \delta_0[\to T_p\mathcal{M}$ such that

$$\varphi(0) = 0, \qquad \exp_p(\dot{z}(0) + \varphi(\delta)) = \gamma(\tau(z) + \delta). \tag{5.5}$$

Differentiating with respect to δ and setting $\delta = 0$, we obtain

$$d\exp_p(\dot{z}(0))[\varphi'(0)] = \dot{\gamma}(\tau(z)). \tag{5.6}$$

The following lemma is needed. □

Lemma 5.4. *Fix $v_0 \in T_p\mathcal{M}$ light-like and future pointing. Set $V = d\exp_p(v_0)[v]$. Assume that V is time-like and future pointing. Then $\langle v, v_0\rangle < 0$.*

Proof. Denote by z the geodesic such that $z(0) = p$ and $\dot{z}(0) = v_0$. As known, since $v_0 = \dot{z}(0)$, $d\exp_p(v_0)[v]$ is given by $Z(1)$, where Z is the unique Jacobi field along z such that $Z(0) = 0$ and $D_s Z(0) = v$. Since $\zeta(s) = s\dot{z}(s)$ is the unique Jacobi field along z such that $\zeta(0) = 0$ and $D_s\zeta(0) = \dot{z}(0) = v_0$, we have

$$d\exp_p(v_0)[v_0] = \dot{z}(1). \tag{5.7}$$

Now, by the Gauss lemma (cf. [1]), for any $v \in T_p\mathcal{M}$, we have

$$\langle d\exp_p(v_0)[v_0], d\exp_p(v_0)[v]\rangle = \langle v_0, v\rangle. \tag{5.8}$$

By (5.7), $V_0 = d\exp_p(v_0)[v_0]$ is light-like and future pointing. Indeed, $\dot{z}(1)$ is light-like and future pointing since $\dot{z}(0) = v_0$ is light-like and future pointing. Moreover, $V = d\exp_p(v_0)[v]$ is time-like and future pointing by assumption, therefore $\langle V_0, V\rangle < 0$. Then, by (5.8) the proof is complete. □

Now, let us go back to the proof of Proposition 3.2

Since $\dot{\gamma}(\tau(z(1)))$ is time-like and future pointing, by (5.6) and Lemma 5.4 we get

$$\langle \varphi'(0), \dot{z}(0)\rangle < 0. \tag{5.9}$$

By (5.9), since $\varphi(0) = 0$ and $\langle \dot{z}(0), \dot{z}(0)\rangle = 0$ up to the choice of a smaller δ_0, we immediately obtain

$$\langle \dot{z}(0) + \varphi(\delta), \dot{z}(0) + \varphi(\delta)\rangle < 0 \quad \forall \delta > 0. \tag{5.10}$$

Moreover, since $\varphi(0) = 0$, for any δ sufficiently small,

$$\langle \dot{z}(0) + \varphi(\delta), Y(z(0))\rangle < 0. \tag{5.11}$$

Then we can conclude the proof taking $\epsilon_0 = \epsilon_0(\delta_0)$ and

$$\epsilon = \epsilon(\delta) = \sqrt{-\langle \dot{z}(0) + \varphi(\delta), \dot{z}(0) + \varphi(\delta)\rangle}, \tag{5.12}$$

which is well defined because of (5.10).

Indeed the geodesic z_ϵ such that $z_\epsilon(0) = p$ and $\dot{z}_\epsilon(0) = z(0) + \varphi(\delta)$ is in $\mathcal{L}^+_{p,\gamma,\epsilon}$ since by (5.12) $\langle \dot{z}_\epsilon, \dot{z}_\epsilon \rangle = -\epsilon^2$.

Moreover, by (5.11) $\dot{z}_\epsilon(0)$ is future pointing so that $z_\epsilon(s)$ is time-like and future pointing for any s. Finally, the C^2 convergence of z_ϵ to z is obvious by the continuous dependence of the solutions of differential equations on its data. □

We conclude the section with an useful result for the proof of Theorem 1.6.

Proposition 5.5. *Suppose that τ is pseudo-coercive on $\mathcal{L}^+_{p,\gamma}$. Let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to* 0 *and $z_n \in \mathcal{L}^+_{p,\gamma,\epsilon_n}$ be a sequence of curves such that*

$$\sup_n \tau(z_n) = \overline{c} < +\infty.$$

Then, denoting by $l(z_n)$ the length of z_n with respect to the Riemannian metric (2.4),

$$\sup_n l(z_n) < +\infty.$$

Moreover, there exists K, compact subset of $\mathcal{M}$, such that $z_n([0, 1]) \subset K$ for any n.

Proof. Denote by $\tilde{z}_n$ the sequence

$$\tilde{z}_n = \psi_{\epsilon_n}(z_n) \in \hat{\mathcal{L}}^+_{p,\gamma}. \tag{5.13}$$

By 5 of Proposition 5.1, $\tau(\tilde{z}_n) \leq \tau(z_n) \leq \overline{c}$. Then, by the same arguments used in the proof of Lemma 2.6, we see that the pseudo-coercivity of τ implies that

$$l(\tilde{z}_n) \leq \tilde{c} < +\infty \quad \text{for any } n, \tag{5.14}$$

and there exists a compact subset of $\mathcal{M}$ containing the images of all the $\tilde{z}_n$'s. Moreover, since $z_n = \phi_{\epsilon_n}(\psi_{\epsilon_n}(z_n))$, by (6) of Proposition 5.1 there exists $M > 0$ such that

$$d_2(\tilde{z}_n, z_n) \leq M \cdot \epsilon_n.$$

Therefore, it follows that $l(z_n)$ is bounded and there exists a compact subset of $\mathcal{M}$ containing the images of the z_n's. □

6. The limit process and the Morse relations on $\mathcal{L}^+_{p,\gamma}$

In this section we shall prove Theorems 1.6 and 1.4.

Proof of Theorem 1.6. Let z_n be as in the statement of Theorem 1.6. Since $\tau(z_n) \leq c$ for all n, by Proposition 5.5 there exists a compact subset K of $\mathcal{M}$ and a positive constant C such that

$$z_n([0, 1]) \subset K \quad \text{and} \quad l(z_n) \leq C \ \ \forall n \in \mathbb{N}.$$

Then, the proof is obtained passing to the limit as $\varepsilon \to 0$ in the Cauchy problem related to the geodesic equation satisfied by the z_n's. □

Proof of Theorem 1.4. Let c be a regular value for τ on $\mathcal{L}^+_{p,\gamma}$, i.e., $\tau^{-1}(c)\cap\mathcal{L}^+_{p,\gamma}$ does not contain geodesics. By assumption 2, all the geodesics in $\tau^{-1}(c)\cap\mathcal{L}^+_{p,\gamma}$ are isolated. A simple compactness argument shows that they are finite. By Proposition 5.2 there exists a positive number $\epsilon(c)$ such that for any geodesic z_i in $\tau^{-1}(c)\cap\mathcal{L}^+_{p,\gamma}$ and for any $\epsilon\in]0,\epsilon(c)]$, there exists a unique geodesic $z^i_\epsilon\in\tau^{-1}(c)\cap\mathcal{L}^+_{p,\gamma,\epsilon}$ approaching z_i for any $i=1,\dots,k$. Choose $\epsilon(c)\le\epsilon_0(c)$ given by Proposition 5.1 and denote (for any $\epsilon\in]0,\epsilon_0]$) by c_ϵ the minimal real number such that

$$\phi_\epsilon(\tau^c\cap\mathcal{L}^+_{p,\gamma})\subset\tau^{c_\epsilon}\cap\mathcal{L}^+_{p,\gamma,\epsilon},$$

where ϕ_ϵ is defined in Proposition 5.1.

If $\epsilon(c)$ is sufficiently small, any c_ϵ is a regular value for τ on $\mathcal{L}^+_{p,\gamma,\epsilon}$ for all $\epsilon\in[0,\epsilon(c)]$ and for any geodesic in $\mathcal{L}^+_{p,\gamma,\epsilon}$ "correspond" to a unique geodesic on $\mathcal{L}^+_{p,\gamma}$ (having the same geometric index) (cf. Proposition 5.2 and Theorem 1.7).

Moreover, choosing $\epsilon(c)$ small enough, by the pseudo-coercivity of τ on $\mathcal{L}^+_{p,\gamma}$ we have the existence of a compact subset $K=K(c)$ of $\mathcal{M}$ and of a positive constant $L=L(c)$ such that $z([0,1])\subset K$ and $l(z)\le L(c)$ for all $\epsilon\in[0,\epsilon(c)]$ and for all $z\in\tau^{c_\epsilon}\cap\mathcal{L}^+_{p,\gamma,\epsilon}$ (cf. Proposition 5.5).

This allows us to use the curve shortening method at every level $b\le c_\epsilon$ and Propositions 3.3 and 3.4.

Arguing as in the proof of Theorem 4.4 and using Theorem 4.1, we can write the following Morse relations valid for every $\epsilon\in]0,\epsilon(c)]$ and every coefficients field $\mathcal{K}$:

$$\sum_{z_\epsilon\in\mathcal{G}^{+,c_\epsilon}_{p,\gamma,\epsilon}}\lambda^{\mu(z_\epsilon)}=\mathcal{P}_\lambda(\tau^{c_\epsilon}\cap\mathcal{L}^+_{p,\gamma,\epsilon},\mathcal{K})+(1+\lambda)\mathcal{S}_\epsilon(\lambda),\tag{6.1}$$

where $\mathcal{G}^{+,d}_{p,\gamma,\epsilon}=\mathcal{G}^+_{p,\gamma}\cap\tau^d$.

Now choose a monotone sequence c_m of regular values for τ on $\mathcal{L}^+_{p,\gamma}$ such that $c_m\to+\infty$. For any m let $\epsilon_m=\epsilon(c_m)$ as above. Let d_m be the minimal real number such that

$$\phi_\epsilon(\tau^{c_m}\cap\mathcal{L}^+_{p,\gamma})\subset\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon}\quad\text{for any }\epsilon\in[0,\epsilon_m].$$

(Note that $d_m\ge c_m$.) By (6.1), Propositions 5.1 and 5.2 and Theorem 1.7 we deduce

$$\sum_{z\in\mathcal{G}^{+,d_m}_{p,\gamma}}\lambda^{\mu(z)}=\mathcal{P}_\lambda(\psi_{\epsilon_m}(\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon_m}),\mathcal{K})+(1+\lambda)\mathcal{S'}_m(\lambda),$$

where $\mathcal{S'}_m$ is a polynomial with nonnegative integer coefficients.

By the exactness in singular homology of the pair $(\mathcal{L}^+_{p,\gamma},\psi_{\epsilon_m}(\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon_m}))$, there exists a formal series R_m (with coefficients in $\mathbb{N}\cup\{+\infty\}$) such that (cf. e.g. [12])

$$\mathcal{P}_\lambda(\psi_{\epsilon_m}(\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon_m}))+\mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma},\psi_{\epsilon_m}(\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon_m})=\mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma})+(1+\lambda)R_m(\lambda).$$

Then, there exists a formal series S_m such that

$$\sum_{z\in\mathcal{G}^{+,d_m}_{p,\gamma}}\lambda^{\mu(z)}+\mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma},\psi_{\epsilon_m}(\tau^{d_m}\cap\mathcal{L}^+_{p,\gamma,\epsilon_m}))=\mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma})+(1+\lambda)\mathcal{S}_m(\lambda).\tag{6.2}$$

Let $N(l,m)$ be the number of light-like geodesics in $\psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})$ having geometric index equal to l. By Proposition 5.1, the subsets $\psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})$ are ordered by inclusion. Then $N(l,m)$ is nondecreasing in m and tends, as $m \to +\infty$, to the number $N(l)$ of the light-like geodesics in $\mathcal{L}^+_{p,\gamma}$ having geometric index equal to l. Since $\mathbb{N} \cup \{+\infty\}$ is compact (with respect to its usual convergence), a diagonalization argument shows the existence of a subsequence $(m_k)_{k\in\mathbb{N}}$ such that for any $l \in \mathbb{N}$ the sequences (b_{l,m_k}) of the formal series S_{m_k} in (6.2) converges to $b_l \in \mathbb{N} \cup \{+\infty\}$. Then, up to considering subsequences, every coefficient $b_{l,m}$ of S_m is convergent to b_l. We shall prove (1.4) arguing for any coefficient $l \in \mathbb{N}$. If $N(l) = +\infty$, either the lth coefficient β_l of $\mathcal{P}_\lambda(\mathcal{L}^+_{p,\gamma}, \mathcal{K})$ is equal to $+\infty$, or at least one between b_{l-1} and b_l is equal to $+\infty$. In any case,

$$N(l) = \beta_l + b_{l-1} + b_l, \tag{6.3}$$

obtaining (1.4) relatively to the lth coefficient.

Assume now that $N(l) < +\infty$. Let

$$b_* = \max\{\tau(z) : z \in \mathcal{G}^+_{p,\gamma}, \mu(z) = q\}. \tag{6.4}$$

By (6.2), in order to prove (6.3), it suffices to show the vanishing of the Betti number:

$$\beta_l(\mathcal{L}^+_{p,\gamma}, \psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})) = 0 \quad \forall m \text{ such that } c_m > b_*. \tag{6.5}$$

Assume by contradiction that (6.5) does not hold. Let Δ_m be a nontrivial element of the homology group $H_l(\mathcal{L}^+_{p,\gamma}, \psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m}))$ and let K_m be its compact support. Now for any $\epsilon \in]0, \epsilon_m]$, by Proposition 5.1, there exists $\mu_m > 0$ (infinitesimal as ϵ_m tends to 0) such that

$$\begin{aligned}\psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m}) &\subset \psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}) \subset \psi_{\epsilon_m}(\tau^{d_m+\mu_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})\\ &\subset \psi_\epsilon(\tau^{d_m+\mu_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}).\end{aligned}$$

Now, choosing ϵ_m small enough, we can assume that there are no geodesics in the strip $\tau^{-1}([d_m, d_m+\mu_m]) \cap \mathcal{L}^+_{p,\gamma,\epsilon}$ for all $\epsilon \in]0, \epsilon_m]$. Then, if ϵ_m is small, $\psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})$ is a strong deformation retract of $\psi_{\epsilon_m}(\tau^{d_m+\mu_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})$ and $\psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon})$ is a strong deformation retract of $\psi_\epsilon(\tau^{d_m+\mu_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon})$ for any $\epsilon \in]0, \epsilon_m]$. (Recall that $Y \subset X$ is a strong deformation retract of X if there exists a continuous map $H : [0,1] \times X$ such that $H(0,\cdot)$ is the identity on X, $H(s,\cdot)$ is the identity on Y for all s, and $H(1,X) \subset Y$.)

Then, by standard techniques in Algebraic Topology we have that for any $k \in \mathbb{N}$,

$$i_k^* : H_k(\mathcal{L}^+_{p,\gamma}, \psi_{\epsilon_m}(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon_m})) \to H_k(\mathcal{L}^+_{p,\gamma}, \psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}))$$

(where i denotes the inclusion map) is an isomorphism. Therefore, there exists $\Delta_\epsilon \in H_l(\mathcal{L}^+_{p,\gamma,\epsilon}, \psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma})) \setminus \{0\}$ with support K_m. Finally, choose

$$C_m > \sup\{\tau(z) : z \in K_m \cap \mathcal{L}^+_{p,\gamma}\}, \quad C_m \text{ regular value for } \tau \text{ on } \mathcal{L}^+_{p,\gamma}.$$

(Clearly, C_m can be chosen larger than d_m.) Using the exactness of the triple $(\mathcal{L}^+_{p,\gamma}, \psi_\epsilon(\tau^{C_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}), \psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}))$ gives the existence of

$$\Gamma_\epsilon \in H_l(\psi_\epsilon(\tau^{C_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}), \psi_\epsilon(\tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon})) \setminus \{0\}$$

with support K_m.

Since ψ_ϵ is an homeomorphism there exists

$$\hat{\Gamma}_\epsilon \in H_l(\tau^{C_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}, \tau^{d_m} \cap \mathcal{L}^+_{p,\gamma,\epsilon}) \setminus \{0\}$$

with support K_m. Using the curve shortening method and the classical Morse theory nearby critical points shows the existence of $\hat{\epsilon} \in]0, \epsilon_m]$ such that for any $\epsilon \in \hat{\epsilon}$ we have the existence of a geodesic $z_\epsilon \in \tau^{-1}([d_m, C_m]) \cap \mathcal{L}^+_{p,\gamma,\epsilon}$ having index l (see Theorem 4.4). Finally sending ϵ to 0, by Theorems 1.6 and 1.7 we obtain the existence of a geodesic z in $\mathcal{L}^+_{p,\gamma}$ such that

$$\mu(z) = q, \quad \tau(z) \in [d_m, C_m].$$

In particular $\tau(z) \geq d_m \geq c_m > b_*$ in contradiction with (6.4). □

References

[1] J.K. Beem, P.E. Ehrlich, K.L. Easley, Global Lorentzian Geometry, Marcel Dekker, New York, 1996.
[2] H. Brezis, Analyse Fonctionelle, Masson, Paris, 1984.
[3] K.C. Chang, Infinite Dimensional Morse Theory and Multiple Solutions Problems, Birkhaüser, Boston, 1993.
[4] I. Ekeland, On the variational principle J. Math. Anal. Appl. 47 (1974) 324–353.
[5] F. Giannoni, A. Masiello, On a Fermat principle in General Relativity, A Ljusternik–Schnirelmann theory for light rays, Ann. Math. Pure Appl. (IV) CLXXIV (1998) 161–207.
[6] F. Giannoni, A. Masiello, P. Piccione, A variational theory for light rays in stably causal Lorentzian manifolds: regularity and multiplicity results Comm. Math. Phys. 187 (1997) 375–415.
[7] F. Giannoni, A. Masiello, P. Piccione, A Morse Theory for light rays on stably causal Lorentzian manifolds Ann. Inst. H. Poincaré, Physique Theorique 69 (1998) 359–412.
[8] A. Helfer, Conjugate points on space-like geodesics or pseudo-self-adjoint Morse–Sturm–Liouville operators Pac. J. Math. 164 (1994) 321–340.
[9] I. Kovner, Fermat principles in arbitrary gravitational fields Astrophys. J. 351 (1990) 114–120.
[10] M. Lombardi, An application of the topological degree to gravitational lenses Modern Phys. Lett. A 13 (1997) 83–86.
[11] A. Masiello, Variational Methods in Lorentzian Geometry, Pitman Research Notes in Mathematics Series, vol. 309, Longman, London, 1994.
[12] J. Mawhin, M. Willem, Critical Point Theory and Hamiltonian Systems, Springer, Berlin, 1988.
[13] F. Mercuri, P. Piccione, D.V. Tausk, Stability of the focal and geometric index in semi-Riemannian geometry via the Maslov index, Technical Report RT-MAT/99, Department of Mathematics, University of São Paulo, Brazil.
[14] J. Milnor, Morse Theory, Ann. Math. Stud., vol. 51, Princeton University Press, Princeton, NJ, 1963.
[15] B. O'Neill, Semi-Riemannian Geometry with Applications to Relativity, Academic Press, New York, 1983.
[16] R.S. Palais, Morse theory on Hilbert manifolds Topology 2 (1963) 299–340.
[17] R.S. Palais, Foundations of Global Nonlinear Analysis, Benjamin, New York, 1968.
[18] V. Perlick, On Fermat's principle in General Relativity: I. The general case Class. Quantum Grav. 7 (1990) 1319–1331.
[19] A. Petters, Multiplane gravitational lensing I. Morse theory and image counting J. Math. Phys. 36 (1995) 4263–4275.
[20] A. Petters, Multiplane gravitational lensing II. Global geometry of caustics J. Math. Phys. 36 (1995) 4276–4295.
[21] P. Schneider, A new formulation of gravitational lens theory, time-delay and Fermat's principle, Astr. Astrophys. 143 (1985) 413–420.
[22] P. Schneider, J. Ehlers, E.E. Falco, Gravitational Lenses, Springer, Berlin, 1992.
[23] H. Spanier, Algebraic Topology, McGraw-Hill, New York, 1966.
[24] M. Struwe, Variational Methods, Springer, Berlin, 1996.