

Theory of optimal transport for Lorentzian cost functions

Stefan Suhr

(Communicated by Linus Kramer)

Abstract. The optimal transport problem in the context of Lorentz–Finsler geometry is studied. Besides deducing the existence of optimal couplings a result on the intermediate regularity of optimal couplings is given. Furthermore, a solution to the Monge problem and an exact criterion for the existence of causal couplings are established. The results generalize parts of [6], [8] and [11].

1. INTRODUCTION

This article studies optimal transportation in Lorentz–Finsler manifolds from the geometric point of view. The geometric viewpoint necessitates that one passes to a spacetime as configuration space. In a spacetime the time parameter is part of the geometry. There are multiple choices for the time parameter and, as is well known in Lorentzian geometry, no choice is preferred. In other words it is not canonical which part of the spacetime is space, or equivalently which points are isochronous. Thus isochronicity in Lorentz–Finsler geometry is subject to a choice. Usually this choice is made via singling out a time function whose level sets are then thought of as constituting space. After choosing a time function a transport problem can be posed between two level sets of this time function. Solutions to such transport problems are provided for example in [6]. Transport problems originating in applications to relativity though, such as the early universe reconstruction problem (see below), cannot be brought into such a form thus motivating the spacetime perspective. This follows from the fact that the support of one martingale is the boundary of the causal past of a set. In this case the support is in general not acausal (see Section 2). Sets that are not acausal are never the level set of a time function.

The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013), ERC Grant Agreement 307062. The author is partially supported by the SFB/TRR 191: Symplectic Structures in Geometry, Algebra and Dynamics.

In the formulation of other transport problems the measures are not concentrated on a single level set of a time function, but are rather distributed on a continuum of level sets, i.e. distributed over a continuum of time parameters. This appears for example when the same transport problem relative to two different time functions is considered. For a discussion of the dependence on different time functions see [15].

Brenier [7] was the first who took notice of the problem of optimal transportation in Lorentzian geometry. He proposed a transportation problem which, only weakly disguised, is the problem of transportation between parallel spacelike hyperplanes in Minkowski space with respect to the negative Lorentzian distance extended by ∞ . Here a strong form of isochronicity is assumed for the support of each measure, i.e. being supported on level sets of a linear time function. Following this formulation, Bertrand and Puel [6] generalized the problem to a wider class of functions called *relativistic costs*, and gave inter alia a solution to the Monge problem while staying in the same basic geometric frame.

The *early universe reconstruction problem*, studied in [8] and [13] with methods of optimal transportation, asks whether one can construct the trajectories of masses from the big bang to their present day positions in Robertson–Walker spacetimes. A mathematical formulation for general globally hyperbolic spacetimes would read as follows: Given two measures, one concentrated on a Cauchy hypersurface, the other on the past cone of a point, what can be said about the trajectories of the minimizers in a dynamical optimal coupling (see Definition 2.8) of the two measures? Frisch et al. [13] gave a justification to why the problem can be studied with methods from optimal transportation.

The first question that comes to mind when studying a cost function which take an infinite value, such as the cost function considered here, is whether there exists a coupling of two given measures with finite cost. This problem was studied in [6]. Recently, in a more systematic approach, Eckstein and Miller [11, 15, 10] studied the problem and the causal evolution of measures in Lorentzian geometry. Theorem 2.6 below extends the existing results on the question to a more abstract setting including metric spaces.

The other results in this article generalize the previous approaches to the problem of Lorentzian optimal transportation in two directions: The first goal as already mentioned above is the structure of the support of the measures involved, i.e. passing from being supported on surfaces of isochronicity (level sets of time functions) to being distributed in space and time or on achronal sets, see Theorems 2.10, 2.12, 2.13, and 2.14. From the physics point of view this means that observations are not only made at a single point in time but rather over a stretch of time or cannot be brought into the form of a single time parameter.

Second, there is the extension to Lorentz–Finsler geometry. This category includes Lorentzian geometry. Thus one can now study transport problems in relativity in their full generality. The step from relativistic cost functions and

Robertson–Walker spacetimes, respectively, to globally hyperbolic Lorentz–Finsler spacetimes is comparable with passing from Euclidian space to Riemannian manifolds in the theory of optimal transport.

The study was motivated by these comments and the prospect of future developments mimicking the relations between optimal transport and fluid dynamics, Riemannian geometry and formulations of synthetic curvature. Besides these generalizations the article gives new results on the structure of optimal couplings not known even for relativistic cost functions. Section 2 describes the setting and states the main results. Section 3 provides the proofs.

2. THE RESULTS

Let M be a smooth manifold of dimension $m \geq 2$. Denote by $\pi_{TM}: TM \rightarrow M$ the canonical projection of tangent vectors to their base point. Throughout the article one fixes a complete Riemannian metric h on M . The norm $|\cdot|$ and distances $\text{dist}(\cdot, \cdot)$ are understood to be induced by h , unless noted otherwise. Recall that h induces a Riemannian metric on TM . Distances in TM are understood to be induced by this metric. The metric ball around y with radius r is denoted by $B_r(y)$. Set T^0M to be the image of the zero section of TM and 0_p the zero vector in TM_p .

Consider a continuous function $\mathbb{L}: TM \rightarrow \mathbb{R}$ smooth on $TM \setminus T^0M$ and positive homogenous of degree 2 such that the second fiber derivative is non-degenerate with index $m - 1$. One says that $\mathcal{C} \subset TM$ is a *closed cone field* if $\mathcal{C}_p := \mathcal{C} \cap TM_p$ is a closed convex cone for all $p \in M$ and $\mathcal{C} \cup T^0M$ is a closed subset of TM . A *causal structure* \mathcal{C} of (M, \mathbb{L}) is then a choice of a closed cone field \mathcal{C} with $\pi_{TM}(\mathcal{C}) = M$ such that $\text{int } \mathcal{C}$, the open interior of \mathcal{C} , is a connected component of $\{\mathbb{L} > 0\}$. For every point $p \in M$ every connected component of $TM_p \cap \{\mathbb{L} > 0\} \subset TM_p$ belongs to a unique causal structure up to a finite cover, see Section 3.1.

Fix a causal structure \mathcal{C} for (M, \mathbb{L}) . Define a new Lagrangian L on TM by setting

$$L(v) := \begin{cases} -\sqrt{\mathbb{L}(v)} & \text{if } v \in \mathcal{C}, \\ \infty & \text{otherwise.} \end{cases}$$

The function L is fiberwise convex, finite on its domain and positive homogenous of degree one. It further is smooth on $\text{int } \mathcal{C}$. The function L has the features of a Finsler metric of Lorentzian type. This justifies calling the pair (M, L) a *Lorentz–Finsler manifold*. The generality of Lorentz–Finsler geometry is chosen in view of recent developments in the area, see e.g. [14, 16, 17, 18], and the goal to achieve a scope comparable to the one of Tonelli–Lagrangian systems, see e.g. [3, 4, 12].

An absolutely continuous curve $\gamma: I \rightarrow M$ is called *(C-)causal* if $\dot{\gamma} \in \mathcal{C}$ whenever the tangent vector exists. A causal curve $\gamma: I \rightarrow M$ is *timelike* if for all $s \in I$ there exist $\varepsilon, \delta > 0$ such that $\text{dist}(\dot{\gamma}(t), \partial\mathcal{C}) \geq \varepsilon|\dot{\gamma}(t)|$, for every $t \in I$ for which $\dot{\gamma}(t)$ exists and $|s - t| < \delta$.

Denote by $J^+(p)$ the set of points $q \in M$ such that there exists a causal curve with initial point p and terminal point q . Denote by $J^-(p)$ the set of points $q \in M$ such that there exists a causal curve with initial point q and terminal point p . The sets $I^\pm(p)$ are defined in a similar way where causal curves are replaced by timelike ones. For $A \subset M$ set $J^\pm(A) := \bigcup_{p \in A} J^\pm(p)$. Define the set

$$J^+ := \{(p, q) \in M \times M \mid q \in J^+(p)\}.$$

For an open set $U \subseteq M$ define J_U^\pm and I_U^\pm as before for the restriction $(U, \mathcal{C}|_U)$, with $\mathcal{C}|_U := TU \cap \mathcal{C}$.

A Lorentz–Finsler manifold is said to be *causal* if it does not admit a closed causal curve.

Definition 2.1. A causal Lorentz–Finsler manifold (M, L) is *globally hyperbolic* if the sets $J^+(p) \cap J^-(q)$ are compact for all $p, q \in M$.

Every causal structure is a closed and nondegenerate cone field which is wider than an open nondegenerate cone field in the sense of [5]. Therefore Definition 2.1 implies that the causal structure of a globally hyperbolic Lorentz–Finsler manifold is globally hyperbolic in the sense of [5]. By [5, Thm. 3] there exists a smooth function $\tau: M \rightarrow \mathbb{R}$ (called a *splitting*) with

$$-d\tau(v) \leq \min\{L(v), -|v|\}$$

for all $v \in \mathcal{C}$. Corollary 1.8 of [5] implies that there exists a diffeomorphism (also called a *splitting*) $M \cong \mathbb{R} \times N$ such that

$$\tau: M \cong \mathbb{R} \times N \rightarrow \mathbb{R}, \quad p \cong (\theta, x) \mapsto \theta$$

if (M, L) is globally hyperbolic. Note that τ is by far not unique.

Remark 2.2. For a causal curve $\gamma: (a, b) \rightarrow M$ which leaves every compact subset of M for both $t \downarrow a$ and $t \uparrow b$, one has $\lim_{t \downarrow a} \tau \circ \gamma(t) = -\infty$ and $\lim_{t \uparrow b} \tau \circ \gamma(t) = \infty$. This follows from the completeness of h .

Define the Lagrangian action relative to L :

$$\mathcal{A}(\gamma) := \begin{cases} \int L(\dot{\gamma}) dt & \text{if } \gamma \text{ is } \mathcal{C}\text{-causal,} \\ \infty & \text{otherwise.} \end{cases}$$

A causal curve $\gamma: I \rightarrow M$ is an \mathcal{A} -*minimizer* between its endpoints $p, q \in M$ if

$$\mathcal{A}(\gamma) = \inf\{\mathcal{A}(\eta) \mid \eta \text{ connects } p \text{ and } q\}.$$

Proposition 2.3. *Let (M, L) be globally hyperbolic. Then for every pair of points $p, q \in M$ with $(p, q) \in J^+$ there exists an \mathcal{A} -minimizer $\gamma: I \rightarrow M$ with finite action connecting the two points. The minimizer γ solves the Euler–Lagrange equation of \mathbb{L} up to monotone reparameterization and one has $\dot{\gamma} \in \mathcal{C}$ whenever the tangent vector exists.*

Since the arguments are completely analogous to the Lorentzian case, only a brief summary of the proof is given for completeness.

Sketch of proof. For $(p, q) \in J^+$ consider the space $\mathcal{C}(p, q)$ of causal curves η from p to q with $\tau \circ \eta \equiv \text{id}$ and equipped with the uniform C^0 -topology. $\mathcal{C}(p, q)$ is nonempty and compact since (M, \mathcal{C}) is globally hyperbolic by [5, Prop. 5.15].

The restriction $\mathcal{A}: \mathcal{C}(p, q) \rightarrow \mathbb{R}$ is lower semi-continuous. Therefore there exists an \mathcal{A} -minimizer $\gamma: [\tau(p), \tau(q)] \rightarrow M$ in $\mathcal{C}(p, q)$ with $\dot{\gamma} \in \mathcal{C}$ whenever the tangent exists. Now by [17, Thm. 6] γ is a pregeodesic, i.e. solves the Euler–Lagrange equations of the action functional associated to \mathbb{L} up to a monotone reparameterization. \square

For a globally hyperbolic Lorentz–Finsler manifold define the *cost function relative to L* :

$$c_L: M \times M \rightarrow \mathbb{R} \cup \{\infty\},$$

$$(p, q) \mapsto \inf \{ \mathcal{A}(\gamma) \mid \gamma \text{ connects } p \text{ and } q \}.$$

This cost function satisfies the triangle inequality:

$$c_L(p, r) \leq c_L(p, q) + c_L(q, r)$$

for all $p, q, r \in M$. The inequality is nontrivial only if $(p, q), (q, r) \in J^+$. In this case the inequality follows from the observations that the causal relation J^+ is transitive and c_L is defined via an infimum. For an \mathcal{A} -minimizer $\gamma: [a, c] \rightarrow M$ and $a \leq b \leq c$ one has

$$c_L(\gamma(a), \gamma(c)) = c_L(\gamma(a), \gamma(b)) + c_L(\gamma(b), \gamma(c)).$$

For two Borel probability measures μ, ν on M call a Borel probability measure π on $M \times M$ a *coupling of μ and ν* if $(p_1)_\# \pi = \mu$ and $(p_2)_\# \pi = \nu$ where $p_1, p_2: M \times M \rightarrow M$ are the projections onto the first and second factor. Recall that the push-forward $(p_i)_\# \pi$ is defined as $(p_i)_\# \pi(A) := \pi(p_i^{-1}(A))$. The set of couplings of μ and ν is denoted by $\Pi(\mu, \nu)$.

The *cost of a coupling π* is

$$\int_{M \times M} c_L(p, q) d\pi(p, q).$$

Denote by $C_L(\mu, \nu)$ the *minimal cost relative to c_L* of couplings between μ and ν , i.e.

$$C_L(\mu, \nu) := \inf \left\{ \int c_L d\pi \mid \pi \in \Pi(\mu, \nu) \right\} \in \mathbb{R} \cup \{\infty\}.$$

A coupling π of two probability measures μ and ν is *optimal* if the cost of π is minimal, i.e.

$$\int c_L d\pi = C_L(\mu, \nu).$$

Denote by $\mathcal{P}(M)$ the set of Borel probability measures on M and set

$$\mathcal{P}_\tau(M) := \{ \mu \in \mathcal{P}(M) \mid \tau \in L^1(\mu) \}$$

for a splitting $\tau: M \rightarrow \mathbb{R}$.

Proposition 2.4. *Let $\mu, \nu \in \mathcal{P}_\tau(M)$. Then there exists an optimal coupling π of μ and ν .*

Proof. The statement is a direct consequence of [21, Thm. 4.1]. One thus has to check the assumption. Manifolds are Polish spaces and c_L is lower semi-continuous. For the other assumptions one has to find two upper semi-continuous functions $a, b: M \rightarrow \mathbb{R} \cup \{-\infty\}$ with $a \in L^1(\mu)$, $b \in L^1(\nu)$ and $a(p) + b(q) \leq c_L(p, q)$.

The inequality $-d\tau(v) \leq L(v)$ for all $v \in TM$ implies $\tau(p) - \tau(q) \leq c_L(p, q)$. Thus setting $a := \tau \in L^1(\mu)$ and $b := -\tau \in L^1(\nu)$ yields the claim. \square

The abstract existence result in Proposition 2.4 immediately raises the question: Under what assumptions does a coupling with finite cost exist? The simplest case is that of two Dirac measures $\mu = \delta_p$ and $\nu = \delta_q$. A coupling of δ_p and δ_q with finite cost exists if and only if $(p, q) \in J^+$. In turn $(p, q) \in J^+$ if and only if $\delta_q(J^+(A)) \geq \delta_p(A)$ and $\delta_p(J^-(B)) \geq \delta_q(B)$ for all measurable $A, B \subset M$. The necessity of the condition was noticed in [6] for relativistic cost functions and general measures.

The problem can be formulated in a more abstract setting though. Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be locally compact Polish spaces. Denote by $\pi_{\mathcal{X}}: \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$ and $\pi_{\mathcal{Y}}: \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Y}$ the canonical projections. For $\mathcal{J} \subseteq \mathcal{X} \times \mathcal{Y}$, $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$ define

$$\begin{aligned}\mathcal{J}^+(A) &:= \pi_{\mathcal{Y}}(\pi_{\mathcal{X}}^{-1}(A) \cap \mathcal{J}) \subset \mathcal{Y}, \\ \mathcal{J}^-(B) &:= \pi_{\mathcal{X}}(\pi_{\mathcal{Y}}^{-1}(B) \cap \mathcal{J}) \subset \mathcal{X}.\end{aligned}$$

Definition 2.5. Two probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ are \mathcal{J} -related if there exists a coupling π of μ and ν with $\pi(\mathcal{J}) = 1$.

For $\mathcal{X} = \mathcal{Y} = M$ and $\mathcal{J} = J^+$ a coupling π with $\pi(J^+) = 1$ is called a *causal coupling*. Further for two probability measures $\mu, \nu \in \mathcal{P}_{\tau}(M)$ the J^+ -relation is equivalent to the finiteness of the optimal cost, i.e. $|C_L(\mu, \nu)| < \infty$. Indeed, if $|C_L(\mu, \nu)| < \infty$, there exists a coupling π of μ and ν with $|\int c_L d\pi| < \infty$, i.e. $\pi(J^+) = 1$. If on the other hand there exists a causal coupling of μ and ν then

$$\left| \int c_L d\pi' \right| \leq \left| \int \tau d\mu \right| + \left| \int \tau d\nu \right| < \infty$$

for every causal coupling π' . It follows that $|C_L(\mu, \nu)| < \infty$.

Theorem 2.6. Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be Polish spaces and $\mathcal{J} \subseteq \mathcal{X} \times \mathcal{Y}$ closed. Further let $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Then the following are equivalent:

- (i) μ and ν are \mathcal{J} -related.
- (ii) $\nu(\mathcal{J}^+(A)) \geq \mu(A)$ and $\mu(\mathcal{J}^-(B)) \geq \nu(B)$ for all measurable $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$.

Recently, Eckstein and Miller [11] proved a similar statement for causally simple Lorentzian spacetimes. Therein the authors study the existence of causal couplings for different causality assumptions via causal function, a relaxed notion of time function.

After addressing the existence problem of optimal couplings, attention turns towards the structure of the optimal couplings. Recall that a set $A \subseteq M \times M$ is c_L -cyclically monotone if

$$\sum c_L(p_i, q_i) \leq \sum c_L(p_i, q_{\sigma(i)})$$

for all $\{(p_i, q_i)\}_{1 \leq i \leq n} \subseteq A$ and all $\sigma \in S(n)$.

Define

$$\mathcal{P}_\tau^+(M) := \{(\mu, \nu) \mid \mu, \nu \in \mathcal{P}_\tau(M) \text{ are } J^+ \text{-related}\}.$$

Proposition 2.7. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$.*

(i) *One has*

$$C_L(\mu, \nu) = \sup \left(\int_M \varphi(q) d\nu(q) - \int_M \psi(p) d\mu(p) \right)$$

where the supremum is taken over the functions $\psi \in L^1(\mu), \varphi \in L^1(\nu)$ with $\varphi(q) - \psi(p) \leq c_L(p, q)$.

(ii) *Every optimal coupling π is concentrated on a c_L -cyclic monotone Borel subset of $M \times M$.*

Proof. Consider the modified cost function $c'_L(p, q) := c_L(p, q) + \tau(q) - \tau(p)$. Since $0 \leq L(v) + d\tau(v)$, it follows that $c'_L \geq 0$. Now the claim follows from [2, Thms. 3.1, 3.2]. \square

Denote by Γ the set of \mathcal{A} -minimizers $\gamma: [0, 1] \rightarrow M$ such that

$$d\tau(\dot{\gamma}) \equiv \tau(\gamma(1)) - \tau(\gamma(0)).$$

Set $\text{ev}: \Gamma \times [0, 1] \rightarrow M$, $(\gamma, t) \mapsto \gamma(t)$ and $\text{ev}_t := \text{ev}(\cdot, t)$. For $(p, q) \in J^+$ consider the subspace

$$\Gamma_{p \rightarrow q} := \text{ev}_1^{-1}(q) \cap \text{ev}_0^{-1}(p).$$

Recall the definition of a dynamical optimal coupling from [21]:

Definition 2.8. A *dynamical optimal coupling* is a probability measure Π on Γ such that $\pi := (\text{ev}_0, \text{ev}_1)_\# \Pi$ is an optimal coupling between $\mu := (\text{ev}_0)_\# \Pi$ and $\nu := (\text{ev}_1)_\# \Pi$.

Proposition 2.9. *For every $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$ there exists a dynamical optimal coupling Π for μ and ν .*

Define the map $[\partial_t \text{ev}]: \Gamma \times [0, 1] \rightarrow PTM$, $(\gamma, t) \mapsto [\dot{\gamma}(t)] \in PTM_{\gamma(t)}$ where PTM denotes the projective tangent bundle. For the canonical projection $P: PTM \rightarrow M$ one has $\text{ev} = P \circ [\partial_t \text{ev}]$. Denote by $\text{supp } \mu$ the support of the measure μ .

Theorem 2.10. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$ with $\text{supp } \mu \cap \text{supp } \nu = \emptyset$. Then every dynamical optimal coupling Π of μ and ν has the following property: The canonical projection P restricted to the image of $T := [\partial_t \text{ev}](\text{supp } \Pi \times]0, 1[)$ is injective. Further the inverse $(P|_T)^{-1}$ is locally Hölder continuous with exponent $\frac{1}{2}$.*

The following example shows the optimality of the Hölder continuity in Theorem 2.10.

Example 2.11. Consider the Minkowski space $(\mathbb{R}^3, \langle \cdot, \cdot \rangle_1)$, where $\langle \cdot, \cdot \rangle_1 = dx^2 + dy^2 - dz^2$ for the natural coordinates $\{x, y, z\}$ on \mathbb{R}^3 . Set

$$C := \{v \in T\mathbb{R}^3 \mid \langle v, v \rangle_1 \leq 0, dz(v) \geq 0\}$$

with the Lorentz–Finsler metric $L|_C(v) = -\sqrt{|\langle v, v \rangle_1|}$.

Next let $\Phi: \mathbb{R} \times (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^3$ be the map $(x, \varphi) \mapsto (x + \cos \varphi, \sin \varphi, 1)$ for $0 < \varepsilon < \frac{\pi}{2}$. The map Φ is an embedding and for x fixed the curve $\varphi \mapsto \Phi(x, \varphi)$ parameterizes $\partial J^+(x, 0, 0) \cap \{z = 1\}$ near $(x + 1, 0, 1)$. For $\varphi \in (-\varepsilon, \varepsilon)$ the Φ -preimage of

$$\{z = 1\} \cap \partial J^+((x - \cos \varphi, -\sin \varphi, -1))$$

near $\Phi(x, \varphi) = (x + \cos \varphi, \sin \varphi, 1)$ is described by a smooth function

$$j_{(x, \varphi)}: (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$$

with $j_{(x, \varphi)}(\varphi) = x$, $j'_{(x, \varphi)}(\varphi) = 0$ and $j''_{(x, \varphi)}(\varphi) > 0$. Choose a constant $j''_{(x, \varphi)}(\varphi) < C < \infty$ and consider the function $w: [0, \varepsilon) \rightarrow \mathbb{R}$, $w(\varphi) = C \cdot \varphi^2$.

By diminishing ε if necessary, the fact that

$$\{x \leq w(\varphi)\} \times (-\varepsilon, \varepsilon) \subset \Phi^{-1}(J^+(w(\varphi) - \cos \varphi, -\sin \varphi, -1))$$

and the choice of $C \in \mathbb{R}$ imply that

$$\Phi^{-1}(J^+(w(\varphi) - \cos \varphi, -\sin \varphi, -1)) \cap \text{graph}(w) \subset (-\infty, w(\varphi)] \times [0, \varphi]$$

for every $\varphi \in [0, \varepsilon)$. By restricting ε further, one can in fact assume that

$$(1) \quad \Phi^{-1}(J^+(w(\varphi) - \cos \varphi, -\sin \varphi, -1)) \cap \text{graph}(w) = \text{graph}(w|_{[0, \varphi]}).$$

Now consider the 1-dimensional Lebesgue measure μ on the interval $\text{im}(w) = [0, C \cdot \varepsilon^2)$ normalized to 1. Define two maps $\psi_{0,1}: [0, C \cdot \varepsilon^2) \rightarrow \mathbb{R}^3$ by setting

$$\begin{aligned} \psi_0(x) &= (x - \cos w^{-1}(x), -\sin w^{-1}(x), -1), \\ \psi_1(x) &= (x + \cos w^{-1}(x), \sin w^{-1}(x), 1). \end{aligned}$$

Denote by $\mu_0 := (\psi_0)_\# \mu$ and $\mu_1 := (\psi_1)_\# \mu$. Since

$$(2) \quad \psi_1(x) = \psi_0(x) + 2(\cos w^{-1}(x), \sin w^{-1}(x), 1),$$

μ_0 and μ_1 are J^+ -related. Due to (1) one knows that $(\psi_0)_\# \mu|_{[0, x]}$ is coupled by any causal coupling to $(\psi_1)_\# \mu|_{[0, x]}$ for all $x \in [0, C \cdot \varepsilon^2)$. Therefore up to changes on a neglectable set the only possible causal coupling is induced by (2). Thus every dynamical coupling Π is concentrated on the curves

$$\gamma_x: t \mapsto \psi_0(x) + 2t(\cos w^{-1}(x), \sin w^{-1}(x), 1),$$

$x \in [0, C \cdot \varepsilon^2)$ and their monotone reparameterizations.

Then the evaluations are

$$\text{ev}\left(\gamma_x, \frac{1}{2}\right) = (x, 0, 0)$$

and

$$[\partial_t \text{ev}] \left(\gamma_x, \frac{1}{2} \right) = [(\cos w^{-1}(x), \sin w^{-1}(x), 1)].$$

The map $(P|_T)^{-1}$ is given by $(x, 0, 0) \mapsto [(\cos w^{-1}(x), \sin w^{-1}(x), 1)]$ and is therefore only $\frac{1}{2}$ -Hölder.

The map $(P|_T)^{-1}$ in Theorem 2.10 is Lipschitz for $m = 2$, i.e. if M is a surface. This is a well known fact for positive definite Lagrangians relying on the fact that trajectories (i) solve a differential equation with smooth coefficients and (ii) have codimension 1 in a surface. These facts carry over readily to this case.

Theorem 2.12. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$ with disjoint supports. Further let K be a compact subset of $\text{int } \mathcal{C}$, the open interior of \mathcal{C} . Then the canonical projection P restricted to the image of $[\partial_t \text{ev}](\text{supp } \Pi \times]0, 1]) \cap K$ is injective and its inverse is Lipschitz for every dynamical optimal coupling Π .*

A set $X \subset M$ is (\mathcal{C} -)achronal if every timelike curve $\eta: I \rightarrow M$ intersects X at most once. Using a splitting, one sees that X can be written as the graph of a function f_X over a subset of N . With the same proof as for [19, Prop. 14.25], one sees that f_X is locally Lipschitz with respect to the metric induced by h . Now one can use a Lipschitz-continuous extension of f_X to N to say that X is the subset of a locally Lipschitz hypersurface.

A locally Lipschitz hypersurface X has a tangent space almost everywhere, and with the induced Riemannian metric one can thus define a Lebesgue measure \mathcal{L}_X on X . A measure concentrated on X is *absolutely continuous with respect to the Lebesgue measure* if it is absolutely continuous with respect to \mathcal{L}_X . Note that the definition is independent of the chosen Riemannian metric since for any pair $\mathcal{L}_X, \mathcal{L}'_X$ of Lebesgue measures induced by Riemannian metrics the measure \mathcal{L}_X is absolutely continuous with respect to \mathcal{L}'_X and vice versa.

Call a hypersurface Y *locally uniformly spacelike* if for one (hence every) splitting there exists a locally Lipschitz continuous function $f_Y: N \rightarrow \mathbb{R}$ with Y being the graph of f_Y and for all compact $K \subseteq M$ there exists $\varepsilon > 0$, such that the Hausdorff distance between $TY_y \cap T^1M$ and $\mathcal{C}^1 := \mathcal{C} \cap T^1M$ is bounded below by ε for all $y \in K \cap \Gamma$ such that TY_y exists. T^1M denotes the unit tangent bundle of h . With these notions the following generalization of [6, Thm. 4.3] can be given.

Theorem 2.13. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$. Assume that μ and ν are concentrated on a locally uniformly spacelike hypersurface A and an achronal set B , respectively. Further assume that μ is absolutely continuous with respect to the Lebesgue measure on A . Then there exists a unique optimal coupling π and a Borel map $F: M \rightarrow M$ such that $\pi = (\text{id}, F)_\# \mu$.*

Uniqueness fails if both A and B are allowed to be achronal only. Consider for example subsets $A, B \subset \partial J^-(p)$ in Minkowski space for some $p \in \mathbb{R}^m$. For suitable choices of A and B not every optimal coupling is supported on a

graph. More precisely every causal coupling has vanishing cost, but not every causal coupling is supported on a graph.

Existence fails if B is not assumed to be achronal. An example is given by μ defined as the 1-dimensional Lebesgue measure on $[0, 1] \times \{0\}$ in the 2-dimensional Minkowski space and ν a nontrivial superposition of the 1-dimensional Lebesgue measures on $[1, 2] \times \{1\}$ and $[2, 3] \times \{2\}$. The only possible causal coupling and therefore optimal one splits every point in $[0, 1] \times \{0\}$ into two parts with weights depending on the superposition. Since the superposition is nontrivial, the coupling cannot be induced by a graph.

Theorem 2.14. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$. Assume that μ is absolutely continuous with respect to the Lebesgue measure on M and ν is concentrated on an achronal set B . Then there exists a unique optimal coupling π and a Borel map $F: M \rightarrow M$ such that $\pi = (\text{id}, F)_\# \mu$.*

Theorem 2.13 corresponds to the classical Monge problem which from the spacetime perspective deals with the problem of coupling two measures concentrated on different level sets of a splitting τ (hence time function) and μ being absolutely continuous with respect to the Lebesgue measure on that level set. Theorem 2.14 on the other hand is a version where the initial measure is distributed in space and time, i.e. from the classical point of view a family of measures.

Remark 2.15. In the spirit of the present approach all results are formulated with as little reference to the splitting τ as possible. Note that τ enters the assumptions of the main results only through an integrability condition, i.e. “ $\tau \in L^1(\mu) \cap L^1(\nu)$ ”. This is automatically satisfied for compactly supported measures. That is, in this special case all results are indeed independent of the splitting.

3. THE PROOFS

3.1. Causal structures. The existence of causal structures is implicitly stated in [16, p. 1534] and [17, p. 583]. The argument is standard material and known for Lorentzian metrics. For completeness it is briefly outlined here.

Let $\mathbb{L}: TM \rightarrow \mathbb{R}$ be a continuous function positive homogenous of degree 2 and smooth on $TM \setminus T^0M$ such that the second fiber derivative is nondegenerate with index $m - 1$. By [16, Prop. 2] the number k of connected components of $TM_p \cap \{\mathbb{L} > 0\}$ is independent of $p \in M$. Thus every point $p \in M$ has a neighborhood U such that the fiber bundle $\pi_{TM}^{-1}(U) \cap \{\mathbb{L} > 0\} \rightarrow U$ is isomorphic to $\bigsqcup_{i=1}^k U \times C_i \rightarrow U$ where C_i denotes the forward time cone in the Minkowski m -space \mathbb{R}_1^m . Let $\{U_l\}_{l \in \mathbb{N}}$ be a locally finite open covering of M such that

$$\pi_{TM}^{-1}(U_l) \cap \{\mathbb{L} > 0\} \cong \bigsqcup_{i=1}^k (U_l \times C_i).$$

Take the disjoint union

$$\mathcal{M} := \bigsqcup_l (U_l \times \{1, \dots, k\})$$

and define “ \sim ” to be the equivalence relation generated by $(p, r) \sim (q, s)$ if $p = q$ and $\{p\} \times C_r$ and $\{q\} \times C_s$ are mapped to the same connected component of $\{\mathbb{L} > 0\} \cap TM_p$ by their respective trivializations. Now one shows that the set

$$M^{\mathbb{L}} := \mathcal{M} / \sim$$

is a smooth manifold and the map $\pi^{\mathbb{L}}: M^{\mathbb{L}} \rightarrow M$, $[(p, r)] \mapsto p$ is a finite covering. See [19, Chap. 7] for the case of Lorentzian manifolds.

Consider the pullback $\overline{\mathbb{L}} := (\pi^{\mathbb{L}})^* \mathbb{L}$. Then for every $[(p, r)] \in M^{\mathbb{L}}$ and every connected component of $TM_{[(p, r)]}^{\mathbb{L}} \cap \{\overline{\mathbb{L}} > 0\}$ there exists a vector field $X \in \Gamma(TM^{\mathbb{L}})$ with

$$X_{[(q, s)]} \in TM_{[(q, s)]}^{\mathbb{L}} \cap \{\overline{\mathbb{L}} > 0\}$$

for all $[(q, s)] \in M^{\mathbb{L}}$, i.e. it belongs to a causal structure. This causal structure is unique since the components of $TM_{[(p, r)]}^{\mathbb{L}} \cap \{\overline{\mathbb{L}} > 0\}$ are strictly convex.

3.2. Proof of Theorem 2.6. (i) \Rightarrow (ii): Let $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a coupling of μ and ν with $\pi(\mathcal{J}) = 1$. For any set $B \subseteq \mathcal{Y}$ one has

$$\pi_{\mathcal{Y}}^{-1}(B) \cap \mathcal{J} \subseteq \pi_{\mathcal{X}}^{-1}(\mathcal{J}^-(B)).$$

Since

$$\nu(B) = \pi(\pi_{\mathcal{Y}}^{-1}(B)) = \pi(\pi_{\mathcal{Y}}^{-1}(B) \cap \mathcal{J})$$

and

$$\mu(\mathcal{J}^-(B)) = \pi(\pi_{\mathcal{X}}^{-1}(\mathcal{J}^-(B)))$$

for $B \subseteq \mathcal{Y}$ measurable, the claim follows. The other inclusion is analogous.

(ii) \Rightarrow (i): For this part of the proof one needs two lemmata.

Lemma 3.3. *Assume that μ and ν satisfy the condition in Theorem 2.6 (ii).*

If there exists a measurable set $A \subseteq \mathcal{X}$ such that $\mu(A) = \nu(\mathcal{J}^+(A)) \in (0, 1)$, then the pairs

$$(\mu_A, \nu_A) := \left(\frac{1}{\mu(A)} \mu|_A, \frac{1}{\mu(A)} \nu|_{\mathcal{J}^+(A)} \right)$$

and

$$(\mu_{A^c}, \nu_{A^c}) := \left(\frac{1}{\mu(A^c)} \mu|_{A^c}, \frac{1}{\mu(A^c)} \nu|_{\mathcal{J}^+(A)^c} \right)$$

satisfy the condition in Theorem 2.6 (ii).

If $\nu(B) = \mu(\mathcal{J}^-(B)) \in (0, 1)$ for a measurable set $B \subseteq \mathcal{Y}$, then the pairs

$$(\mu_B, \nu_B) := \left(\frac{1}{\nu(B)} \mu|_{\mathcal{J}^-(B)}, \frac{1}{\nu(B)} \nu|_B \right)$$

and

$$(\mu_{B^c}, \nu_{B^c}) := \left(\frac{1}{\nu(B^c)} \mu|_{\mathcal{J}^-(B)^c}, \frac{1}{\nu(B^c)} \nu|_{B^c} \right)$$

satisfy the condition in Theorem 2.6 (ii).

Proof. It suffices to consider the first case. The second case follows by exchange of \mathcal{X} and \mathcal{Y} . So assume $\mu(A) = \nu(\mathcal{J}^+(A)) \in (0, 1)$ for some measurable set $A \subseteq \mathcal{X}$. First note that all four measures μ_A, ν_A, μ_{A^c} and ν_{A^c} are well-defined probability measures by the assumption. One has

$$\begin{aligned} \mu_A(B) &= \frac{1}{\mu(A)}\mu(B \cap A) \\ &\leq \frac{1}{\mu(A)}\nu(\mathcal{J}^+(B \cap A)) \\ &= \frac{1}{\mu(A)}\nu(\mathcal{J}^+(B) \cap \mathcal{J}^+(A)) = \nu_A(\mathcal{J}^+(B)) \end{aligned}$$

which shows $\mu_A(B) \leq \nu_A(\mathcal{J}^+(B))$.

Next note that $\mu(A^c) = \nu(\mathcal{J}^+(A)^c)$. Assume that there exists a measurable set $C \subseteq \mathcal{X}$ with $\nu_{A^c}(\mathcal{J}^+(C)) < \mu_{A^c}(C)$, i.e.

$$\nu(\mathcal{J}^+(C) \cap \mathcal{J}^+(A)^c) = \nu|_{\mathcal{J}^+(A)^c}(\mathcal{J}^+(C)) < \mu|_{A^c}(C) = \mu(C \cap A^c).$$

Then a contradiction follows from

$$\begin{aligned} \mu(C \cup A) &= \mu(C \cap A^c) + \mu(A) \\ &> \nu(\mathcal{J}^+(C) \cap \mathcal{J}^+(A)^c) + \nu(\mathcal{J}^+(A)) = \nu(\mathcal{J}^+(C \cup A)) \end{aligned}$$

since $\mathcal{J}^+(C) \cup \mathcal{J}^+(A) = \mathcal{J}^+(C \cup A)$. Therefore one has

$$\mu_{A^c}(C) \leq \nu_{A^c}(\mathcal{J}^+(C))$$

for all measurable $C \subseteq \mathcal{X}$. This shows the first set of inequalities.

It remains to show $\mu_A(\mathcal{J}^-(D)) \geq \nu_A(D)$ and $\mu_{A^c}(\mathcal{J}^-(D)) \geq \nu_{A^c}(D)$ for $D \subseteq \mathcal{Y}$ measurable. If $\mu_A(\mathcal{J}^-(D)) < \nu_A(D)$, one has

$$\mu_A(\mathcal{J}^-(D)^c) = 1 - \mu_A(\mathcal{J}^-(D)) > 1 - \nu_A(D) \geq \nu_A(\mathcal{J}^+(\mathcal{J}^-(D)^c))$$

since $\mathcal{J}^+(\mathcal{J}^-(D)^c)$ and D are disjoint. This contradicts the first part. The inequality $\mu_{A^c}(\mathcal{J}^-(D)) \geq \nu_{A^c}(D)$ follows analogously. \square

Lemma 3.4. *Let $n \in \mathbb{N}$. Consider the product $\{1, \dots, n\} \times \{1, \dots, n\}$ with the canonical projections π_1, π_2 onto the first and second factor, respectively. Let $\mathcal{K} \subseteq \{1, \dots, n\} \times \{1, \dots, n\}$ have the property that*

$$\sharp\pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) \geq \sharp A \quad \text{and} \quad \sharp\pi_2(\pi_1^{-1}(A) \cap \mathcal{K}) \geq \sharp A$$

for all $A \subseteq \{1, \dots, n\}$. Then \mathcal{K} contains the graph of a permutation $\sigma \in S(n)$.

Proof. The proof is carried out by induction over n . If $n = 1$, the claim is trivial since $\mathcal{K} = \{1\} \times \{1\}$.

Now assume that the claim has been shown for numbers less than n . First assume that

$$\sharp\pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) > \sharp A \quad \text{and} \quad \sharp\pi_2(\pi_1^{-1}(A) \cap \mathcal{K}) > \sharp A$$

for all nonempty proper subsets A . Choose $1 \leq j \leq n$ with $(n, j) \in \mathcal{K}$. By renumbering, one can assume $j = n$. Now consider

$$\mathcal{J} := \mathcal{K} \cap \{1, \dots, n-1\} \times \{1, \dots, n-1\}.$$

Since

$$\# \pi_1(\pi_2^{-1}(A) \cap \mathcal{J}) \geq \# \pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) - 1 \geq \# A$$

and vice versa for all $A \subseteq \{1, \dots, n-1\}$, one obtains from the induction hypothesis a permutation $o \in S(n-1)$ whose graph is contained in \mathcal{J} . The permutation o extends to a permutation $\sigma \in S(n)$ whose graph is a subset of \mathcal{K} by setting $\sigma(n) := n$ and $\sigma|_{\{1, \dots, n-1\}} \equiv o$.

If there exists a nonempty proper subset A of $\{1, \dots, n\}$ with

$$\# \pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) = \# A \quad \text{or} \quad \# \pi_2(\pi_1^{-1}(A) \cap \mathcal{K}) = \# A,$$

one reduces the problem to constructing two separate permutations on A and A^c . Thus again the induction hypothesis gives separate permutations on A and A^c which together form a permutation σ whose graph is contained in \mathcal{K} .

One only needs to consider the case $\# \pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) = \# A$. The other case follows by exchanging the order. Further by renumbering one can assume that $A = \pi_1(\pi_2^{-1}(A) \cap \mathcal{K})$. Set $\mathcal{K}_A := \mathcal{K} \cap A \times A$, $\mathcal{K}_{A^c} := \mathcal{K} \cap A^c \times A^c$. The goal is to show that \mathcal{K}_A and \mathcal{K}_{A^c} satisfy the assumptions of the lemma.

It is clear that

$$\# \pi_1(\pi_2^{-1}(B) \cap \mathcal{K}_A) = \# \pi_1(\pi_2^{-1}(B) \cap \mathcal{K}) \geq \# B$$

for all $B \subseteq A$ since $\pi_1(\pi_2^{-1}(B) \cap \mathcal{K}) \subseteq A$. If however there exists $C \subseteq A^c$ with $\# \pi_1(\pi_2^{-1}(C) \cap \mathcal{K}_{A^c}) < \# C$ then $\# \pi_1(\pi_2^{-1}(A \cup C) \cap \mathcal{K}) < \#(A \cup C)$ which contradicts the initial assumption.

Assume now that there exists a set $D \subseteq A$ with $\# \pi_2(\pi_1^{-1}(D) \cap \mathcal{K}_A) < \# D$. Set $E := A \setminus \pi_2(\pi_1^{-1}(D) \cap \mathcal{K}_A)$. Then D and $\pi_1(\pi_2^{-1}(E) \cap \mathcal{K}_A)$ are disjoint. This can be seen as follows. If $i \in \pi_1(\pi_2^{-1}(E) \cap \mathcal{K}_A)$ then there exists $j \in E$ such that $(i, j) \in \mathcal{K}_A$. If $i \in D$ then for all $(i, j) \in \mathcal{K}_A$ one has $j \in \pi_2(\pi_1^{-1}(D) \cap \mathcal{K}_A)$. Thus the sets are disjoint. It follows that

$$\# E = \# A - \# \pi_2(\pi_1^{-1}(D) \cap \mathcal{K}_A) > \# A - \# D \geq \# \pi_1(\pi_2^{-1}(E) \cap \mathcal{K}_A)$$

which clearly contradicts the first part of the argument. Now the same argument applies to subsets of A^c . \square

Assume first that there exists $n \in \mathbb{N}$ such that

$$\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \quad \text{and} \quad \nu = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}.$$

Identify $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$ with $\{1, \dots, n\}$. Define the set

$$\mathcal{K} := \{(i, j) \mid (x_i, y_j) \in \mathcal{J}\} \subseteq \{1, \dots, n\} \times \{1, \dots, n\}.$$

Denote by π_1 and π_2 the canonical projections from $\{1, \dots, n\} \times \{1, \dots, n\}$ onto the first and second factor, respectively. Since μ and ν are counting measures, the assumptions become

$$\# \pi_1(\pi_2^{-1}(A) \cap \mathcal{K}) \geq \# A \quad \text{and} \quad \# \pi_2(\pi_1^{-1}(A) \cap \mathcal{K}) \geq \# A$$

for all $A \subseteq \{1, \dots, n\}$. Lemma 3.4 now gives a permutation σ whose graph is contained in \mathcal{H} . Reversing the identifications, one obtains a bijective map

$$\sigma': \{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_n\}$$

with $(x_i, \sigma'(x_i)) \in \mathcal{J}$ for all i . Since μ and ν are counting measures, $(\text{id}, \sigma')\# \mu$ is the desired coupling.

The general case follows from this special case by an approximation argument. Choose sequences of locally finite, disjoint and measurable coverings of $\text{supp } \mu$ and $\text{supp } \nu$, respectively. Then one can approximate both measures in the weak-* topology by finite measures whose support is contained in a given neighborhood of the supports of μ and ν . Consider $\mathcal{J}_\varepsilon := \overline{B_\varepsilon(\mathcal{J})}$, the closure of the ε -neighborhood of \mathcal{J} with respect to the metric on $\mathcal{X} \times \mathcal{Y}$, for $\varepsilon > 0$. Then every pair of finite measures $\mu' = \sum b_i \delta_{x_i}$ and $\nu' = \sum c_j \delta_{y_j}$, approximating μ and ν sufficiently well, satisfies the assumptions in Theorem 2.6 (ii) for \mathcal{J}_ε instead of \mathcal{J} . In order to apply the special case it would suffice to have $b_i, c_j \in \mathbb{Q}$. Simply approximating the weights b_i and c_j by rational numbers and retaining the assumptions of Lemma 3.4 will in general only work if

$$(3) \quad \nu'(\mathcal{J}_\varepsilon^+(A)) > \mu'(A) \quad \text{and} \quad \mu'(\mathcal{J}_\varepsilon^-(B)) > \nu'(B)$$

for all measurable $A \subset \mathcal{X}$ and $B \subset \mathcal{Y}$. With Lemma 3.3 one can split μ' and ν' into submeasures until (3) is satisfied and proceed with the submeasures. Since μ' and ν' have finite supports, this division process terminates after finitely many steps. For μ' or ν' supported in a single point it is obvious how to build a coupling in \mathcal{J}_ε .

If (3) is satisfied, the weights can be approximated by rational numbers such that (3) still holds for the perturbed measures. Then by the special case there exists a coupling supported in $\overline{B_\varepsilon(\mathcal{J})}$. By construction the approximations of μ and ν form precompact sets in the weak-* topology. This implies that the set of couplings is precompact in the weak-* topology as well, see [21, Chap. 4]. The claim follows when passing to the limit using that $\mathcal{J} \cap \text{supp } \mu \times \text{supp } \nu$ is closed.

3.5. Dynamical optimal coupling. For the splitting $\tau: M \rightarrow \mathbb{R}$ choose a smooth vector field X_τ on M with $d\tau(X_\tau) \equiv 1$. Then X_τ is considered to be a vector field on $\mathbb{R} \times N$. Define a Lagrange function

$$L_\tau: \mathbb{R} \times TN \rightarrow \mathbb{R} \cup \{\infty\}, \quad L_\tau(t, v) := L(X_\tau(t, \pi_{TN}(v)) + v).$$

Denote by $\mathcal{D}_\tau \subseteq \mathbb{R} \times TN$ the domain of L_τ . The function L_τ is continuous on \mathcal{D}_τ and smooth on $\text{int } \mathcal{D}_\tau$, the interior of \mathcal{D}_τ . Note that $L_\tau|_{\text{int } \mathcal{D}_\tau} < 0$. For $(t, x) \in \mathbb{R} \times N$ set $\mathcal{D}_{(t,x)} := \mathcal{D}_\tau \cap (\{t\} \times TN_x)$. The point $(t, v) \in \mathcal{D}_{(t,x)}$ is identified with the vector $X_\tau(t, x) + v \in \mathcal{C}$.

Denote by $\partial_v^2 L_\tau$ the *second fiber derivative* of L_τ , i.e.

$$(\partial_v^2 L_\tau)_{(t,v)}(w, z) := \left. \frac{d^2}{dr ds} \right|_{r=s=0} L_\tau(t, v + rw + sz).$$

Lemma 3.6. (i) $\mathcal{D}_{(t,x)}$ is a compact strictly convex domain with smooth boundary for all $(t, x) \in \mathbb{R} \times N$.

(ii) For all $K \subseteq \mathbb{R} \times N$ compact there exists $\delta > 0$ such that the second fiber derivative satisfies

$$(\partial_v^2 L_\tau)_{(t,v)} \geq \frac{\delta}{|L_\tau(t, v)|} \cdot \text{id}$$

for all $(t, x) \in K$ and $v \in \text{int } \mathcal{D}_{(t,x)}$.

Proof. (i) Denote by $\mathcal{C}_{(t,x)}^*$ the dual cone of $\mathcal{C}_{(t,x)} = \mathcal{C}_p$ via the identification $(t, x) \cong p$. Since τ is a Lyapunov function for \mathcal{C} , one has $d\tau_{(t,x)} \in \text{int } \mathcal{C}_{(t,x)}^*$. This implies that $\mathcal{D}_{(t,x)}$ is compact since $X_\tau + v \in \mathcal{C}$ yields

$$1 = d\tau(X_\tau + v) \geq |v| - |X_\tau|$$

which bounds the norm of v . It is further smooth since $\partial\mathcal{C}_{(t,x)}$ is smooth away from the zero section and $\ker d\tau \cap \mathcal{C} = \{0\}$. Finally, the strict convexity follows from the fact that at points in $\partial\mathcal{C} \setminus T^0M$ the bilinear form $\partial_v^2 \mathbb{L}|_{T\partial\mathcal{C} \times T\partial\mathcal{C}}$ is semidefinite with kernel equal to the radial direction, i.e. definite on any hyperplane transversal to the radial direction. Here the radial direction at $v \in TM$ is $\frac{d}{ds}|_{s=0}(1+s)v \in T(TM_p)_v \cong TM_x$ and $\partial_v^2 \mathbb{L}$ is defined analogous to $\partial_v^2 L_\tau$.

(ii) Recall the formula for the second derivative of L in the fiber direction:

$$\partial_v^2 L = \frac{1}{2\sqrt{\mathbb{L}}} \left(\frac{1}{2} \frac{\partial_v \mathbb{L} \otimes \partial_v \mathbb{L}}{\mathbb{L}} - \partial_v^2 \mathbb{L} \right)$$

where $\partial_v \mathbb{L}_v(w) := \frac{d}{ds}|_{s=0} \mathbb{L}(v + sw)$. As seen in (i) one has

$$\partial_v^2 \mathbb{L}|_{T\partial\mathcal{D}_{(t,x)} \times T\partial\mathcal{D}_{(t,x)}} < 0$$

for all $(t, x) \in \mathbb{R} \times N$. Thus one can choose $n < \infty$ and $\delta_1 > 0$ such that

$$\left(\frac{n}{2} \partial_v \mathbb{L} \otimes \partial_v \mathbb{L} - \partial_v^2 \mathbb{L} \right) \Big|_{T\mathcal{D}_\tau \times T\mathcal{D}_\tau} > \delta_1 \cdot \text{id}$$

on a neighborhood U of $\partial\mathcal{D}_\tau$ in $\mathbb{R} \times TN$ over K . This implies the claim on the smaller neighborhood $U \cap \{\mathbb{L} < \frac{1}{n}\}$.

For the remaining points outside of $U \cap \{\mathbb{L} < \frac{1}{n}\}$ note that

$$\frac{1}{2} \frac{\partial_v \mathbb{L} \otimes \partial_v \mathbb{L}}{\mathbb{L}} - \partial_v^2 \mathbb{L} \geq 0$$

with kernel equal to the radial direction. Thus one has

$$\frac{1}{2} \frac{\partial_v \mathbb{L} \otimes \partial_v \mathbb{L}}{\mathbb{L}} - \partial_v^2 \mathbb{L} > \delta_2 \cdot \text{id}$$

on $\mathcal{D}_\tau \setminus U \cap \{\mathbb{L} < \frac{1}{n}\}$ over K for a $\delta_2 > 0$. □

Let $V \subset N$ be open with a chart $V \rightarrow \mathbb{R}^{m-1}$ of N . The induced trivialization of $TV \rightarrow T\mathbb{R}^{m-1} \cong \mathbb{R}^{m-1} \times \mathbb{R}^{m-1}$ gives local coordinates $(x, v) \in \mathbb{R}^{m-1} \times \mathbb{R}^{m-1}$

on TN . The Euler–Lagrange equation of the action functional associated to L_τ reads in these coordinates:

$$(4) \quad \frac{\partial L_\tau}{\partial t} + \frac{\partial L_\tau}{\partial x} - \frac{d}{dt} \left(\frac{\partial L_\tau}{\partial v} \right) = 0.$$

The equation defines an explicit ordinary differential equation of second order since $\partial^2 L_\tau / \partial v^2 > 0$ at points in $\text{int } \mathcal{D}_\tau$. It is standard that the solutions to (4) are of the form $t \mapsto \dot{\eta}(t)$ for some curve $\eta: I \rightarrow N$, i.e. the solutions are tangent curves in TN . For $(t, v) \in \text{int } \mathcal{D}_\tau$ denote by $\eta_{(t,v)}: I \rightarrow N$ the unique maximal solution to (4) with $\dot{\eta}_{(t,v)}(0) = v$. The solutions define a local flow

$$\Phi_\tau: U_\tau \rightarrow \text{int } \mathcal{D}_\tau, \quad (s, (t, v)) \mapsto (s + t, \dot{\eta}_{(t,v)}(s))$$

where $U_\tau \subset \mathbb{R} \times \text{int } \mathcal{D}_\tau$ is an open neighborhood of $\{0\} \times \text{int } \mathcal{D}_\tau$.

Proposition 3.7. Φ_τ extends to a smooth local flow on an open neighborhood of $\{0\} \times \mathcal{D}_\tau$, i.e. there exists an open neighborhood U of $\{0\} \times \mathcal{D}_\tau$ in $\mathbb{R} \times \mathbb{R} \times TN$ and $\tilde{\Phi}_\tau: U \rightarrow \mathbb{R} \times TN$ smooth with $\tilde{\Phi}_\tau \equiv \Phi_\tau$ on U_τ . Furthermore, $\tilde{\Phi}_\tau$ is complete on \mathcal{D}_τ with $\tilde{\Phi}_\tau$ -invariant $\text{int } \mathcal{D}_\tau$ and $\partial \mathcal{D}_\tau$. The extension of Φ_τ to \mathcal{D}_τ is unique and will be denoted by Φ_τ again.

For a local trivialization $W \times \mathbb{R}^m$ of TM with coordinates $(p, w) \in W \times \mathbb{R}^m$ the Euler–Lagrange equation of the action functional associated to \mathbb{L} ,

$$(5) \quad \frac{d}{dt} \left(\frac{\partial \mathbb{L}}{\partial w} \right) - \frac{\partial \mathbb{L}}{\partial p} = 0,$$

defines a local flow outside the zero section since $\partial^2 \mathbb{L} / \partial w^2$ is nondegenerate. For $w \in TM \setminus T^0M$ let $\gamma_w: J \rightarrow M$ be the unique maximal solution to (5) with $\dot{\gamma}_w(0) = w$. Denote by $\Phi^\mathbb{L}: \mathbb{U} \subset \mathbb{R} \times TM \rightarrow TM$, $(t, w) \mapsto \dot{\gamma}_w(t)$ the maximal flow defined by (5) extended to the zero section by constant flow lines, see [17]. $\Phi^\mathbb{L}$ is smooth outside the zero section.

A causal curve $\eta: I \rightarrow N$ is a Φ_τ -trajectory if $t \mapsto (t, \dot{\eta}(t))$ solves (4). A curve $\gamma: J \rightarrow M$ is a $\Phi^\mathbb{L}$ -trajectory if $t \mapsto \dot{\gamma}(t)$ solves (5).

Lemma 3.8. A curve $\eta: I \rightarrow N$ with $(t, \dot{\eta}(t)) \in \text{int } \mathcal{D}_\tau$ for all $t \in I$ is a Φ_τ -trajectory if and only if its graph $H: t \mapsto (t, \dot{\eta}(t))$ is a reparameterization of a $\Phi^\mathbb{L}$ -trajectory γ with $\dot{\gamma} \in \text{int } \mathcal{C}$. Especially the trajectories of Φ_τ and $\Phi^\mathbb{L}$ are in one-to-one correspondence via reparameterization.

Proof. Fix a local chart of $W \rightarrow \mathbb{R}^m$ of M and the induced trivialization of $TW \rightarrow T\mathbb{R}^m \cong \mathbb{R}^m \times \mathbb{R}^m$. Let $(p, w) \in \mathbb{R}^m \times \mathbb{R}^m$. Expanding the Euler–Lagrange equation of L on $\text{int } \mathcal{C}$ gives

$$(6) \quad 0 = \frac{d}{dt} \left(\frac{\partial L}{\partial w} \right) - \frac{\partial L}{\partial p} = \frac{d}{dt} \left(\frac{1}{2\sqrt{\mathbb{L}}} \right) \frac{\partial \mathbb{L}}{\partial w} + \frac{1}{2\sqrt{\mathbb{L}}} \left[\frac{d}{dt} \left(\frac{\partial \mathbb{L}}{\partial w} \right) - \frac{\partial \mathbb{L}}{\partial p} \right].$$

Since \mathbb{L} is autonomous, \mathbb{L} is preserved along orbits of the local Euler–Lagrange flow $\Phi^\mathbb{L}$ of \mathbb{L} . This immediately shows that $\text{int } \mathcal{C}$ and $\partial \mathcal{C} \setminus T^0M$ are invariant under $\Phi^\mathbb{L}$. Thus according to (6) an orbit of $\Phi^\mathbb{L}$ in $\text{int } \mathcal{C}$ solves the Euler–Lagrange equation of L . Conversely let $\theta: I \rightarrow M$ solve the Euler–Lagrange

equations of L . Reparameterizing θ to a curve γ such that \mathbb{L} is constant along $\dot{\gamma}$ yields an orbit of $\Phi^{\mathbb{L}}$.

Now one shows that η is a Φ_τ -trajectory if and only if H solves the Euler–Lagrange equation of L . Let $\eta: I \rightarrow N$ be a Φ_τ -trajectory. Consider a smooth variation $\mathcal{H}: I \times (-\varepsilon, \varepsilon) \rightarrow \mathbb{R} \times N$ of H with fixed endpoints. Since \mathcal{H} is smooth, one can assume, by diminishing ε if necessary, that $\partial_t(\tau \circ \mathcal{H}) > 0$ everywhere. Thus one can smoothly reparameterize \mathcal{H} to satisfy $\partial_t(\tau \circ \mathcal{H}) = 1$, i.e. \mathcal{H} consists of graphs of curves $\eta_s: I \rightarrow N$ ($s \in (-\varepsilon, \varepsilon)$). This shows that any sufficiently small variation of H can be reparameterized to be a variation by graphs. The reparameterization does not affect the value of \mathcal{A} on the variation. Note that

$$\int L_\tau(t, \dot{\eta}_s(t)) dt = \int L(\partial_t \mathcal{H}(t, s)) dt.$$

Now if the first variation of η vanishes, the first variation of the graph vanishes as well, i.e. H solves the Euler–Lagrange equations of L . The converse is obvious, i.e. if H solves the Euler–Lagrange equations of L , then the first variation of η vanishes.

Combining both paragraphs gives the first claim. For the second claim one has to use the positive homogeneity of $\Phi^{\mathbb{L}}$, i.e. $\Phi^{\mathbb{L}}(\lambda t, w) = \Phi^{\mathbb{L}}(t, \lambda w)$ for $\lambda > 0$. Thus reparameterizing a $\Phi^{\mathbb{L}}$ -trajectory to a curve γ with $d\tau(\dot{\gamma}) \equiv 1$ gives the same curves for initial values w and λw where $\lambda > 0$. It remains to note that half lines in $\text{int } \mathcal{C}$ are in one-to-one correspondence with points in \mathcal{D}_τ . \square

Recall that $\pi_{TM}: TM \rightarrow M$ denotes the canonical projection.

Proof of Proposition 3.7. Since $\Phi^{\mathbb{L}}$ is a smooth local flow on $TM \setminus T^0M$, every $v \in \partial\mathcal{C} \setminus T^0M$ admits an $\varepsilon(v) > 0$ and a neighborhood W in $TM \setminus T^0M$ such that $d\tau(\Phi^{\mathbb{L}}(t, w)) > 0$ for all $|t| \leq \varepsilon(v)$ and $w \in W$.

Parameterize the trajectories $t \mapsto \pi_{TM}(\Phi^{\mathbb{L}}(t, w))$ to curves γ_w such that

$$d\tau(\dot{\gamma}_w) \equiv 1 \quad \text{and} \quad \dot{\gamma}_w(0) = \frac{w}{d\tau(w)}.$$

Since $\Phi^{\mathbb{L}}(t, \lambda v) = \Phi^{\mathbb{L}}(\lambda t, v)$ for $\lambda > 0$, the curves γ_w and $\gamma_{\lambda w}$ coincide for positive λ . The tangent curves $t \mapsto \dot{\gamma}_w(t)$ define a local flow. This is due to the fact that

$$\dot{\gamma}_w(s+t) = \dot{\gamma}_{\dot{\gamma}_w(s)}(t)$$

for $|s|, |t|$ sufficiently small. Now since $\Phi^{\mathbb{L}}$ is autonomous, these local definitions of the extensions patch together to give a local flow on a neighborhood of $\mathcal{C} \cap \{d\tau = 1\}$ in $\{d\tau = 1\}$ which preserves $\text{int } \mathcal{C} \cap \{d\tau = 1\}$ and $\partial\mathcal{C} \cap \{d\tau = 1\}$. Projecting the flow to $\mathbb{R} \times TN$ gives a smooth extension since the projection coincides with Φ_τ on $\text{int } \mathcal{D}_\tau$ by Lemma 3.8.

It remains to prove the completeness of the extension. But this follows directly from Remark 2.2 since the γ_w are causal for $w \in \mathcal{C}$ and

$$\tau(\gamma_w(t)) - \tau(\gamma_w(s)) = t - s. \quad \square$$

The Riemannian metric h induces a Riemannian metric on all higher tangent bundles $T^{(k)}M$ where $T^{(k)}M := T(T^{(k-1)}M)$ and $T^{(1)}M := TM$. For $a \leq b \in \mathbb{R}$ define the C^k -topology on smooth curves $\gamma: [a, b] \rightarrow M$ via the induced metrics as

$$\text{dist}_k(\gamma, \eta) := \sup\{\text{dist}(\gamma^{(k)}(t), \eta^{(k)}(t)) \mid t \in [a, b]\}.$$

Lemma 3.9. *For all $(p, q) \in J^+$ the set $\Gamma_{p \rightarrow q}$ is nonempty, consists of smooth curves and is compact in the C^k -topology for all k .*

Proof. By Proposition 2.3 there exists an \mathcal{A} -minimizer between p and q and every \mathcal{A} -minimizer solves the Euler–Lagrange equations of \mathbb{L} up to monotone reparameterization. Monotonously reparameterizing an \mathcal{A} -minimizer between p and q to $\gamma: [0, 1] \rightarrow M$ with $d\tau(\dot{\gamma}) \equiv \tau(q) - \tau(p)$ yields $\gamma \in \Gamma_{p \rightarrow q}$, i.e. $\Gamma_{p \rightarrow q}$ is nonempty. Every curve in $\Gamma_{p \rightarrow q}$ solves the Euler–Lagrange equations of L .

$\Gamma_{p \rightarrow q}$ contains only the constant curve if $p = q$, i.e. in this case $\Gamma_{p \rightarrow q}$ is compact in every topology. If on the other hand one has $p \neq q$, an \mathcal{A} -minimizer $\gamma \in \Gamma_{p \rightarrow q}$ induces a Φ_τ -trajectory $\eta: [\tau(p), \tau(q)] \rightarrow N$ via the graph

$$H(t) = (t, \eta(t)) := \gamma\left(\frac{t - \tau(p)}{\tau(q) - \tau(p)}\right)$$

of η by Lemma 3.8. Identify $p \cong (\tau(p), x)$ and $q \cong (\tau(q), y)$ via the splitting $M \cong \mathbb{R} \times N$. The set of Φ_τ -trajectories $\delta: [\tau(p), \tau(q)] \rightarrow N$ between x and y is compact in all C^k -topologies on $C^\infty([\tau(p), \tau(q)], N)$ by Proposition 3.7 since it is part of a smooth flow. Now the compactness of $\Gamma_{p \rightarrow q}$ is obvious. \square

The following results are analogous to results in [21, Chap. 7].

Proposition 3.10. *There exists a Borel map $S: J^+ \rightarrow C^0([0, 1], M)$ such that $S(p, q) \in \Gamma_{p \rightarrow q}$.*

Proof. For every $(p, q) \in J^+$ the set $\Gamma_{p \rightarrow q}$ is nonempty and compact in every C^k -topology by Lemma 3.9, i.e. nonempty and closed. Further the evaluation map $\text{ev}_0 \times \text{ev}_1$ is Lipschitz. This implies that the correspondence (for the definition see [1, p. 4])

$$(\text{ev}_0 \times \text{ev}_1)^{-1}: J^+ \rightarrow \Gamma$$

is weakly measurable in the sense of [1, Def. 18.1]. Now [1, Thm. 8.13] implies that $(\text{ev}_0 \times \text{ev}_1)^{-1}$ has a measurable selection S , i.e.

$$(\text{ev}_0 \times \text{ev}_1) \circ S \equiv \text{id}|_{J^+}. \quad \square$$

Proof of Proposition 2.9. Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$ and let π be an optimal coupling of μ and ν for the cost c_L . Consider $\Pi := S_\# \pi$. Since $(\text{ev}_0, \text{ev}_1) \circ S \equiv \text{id}$, the claim follows from the definition of optimal dynamical couplings. \square

Corollary 3.11. *Let Π be a dynamical optimal coupling between J^+ -related measures μ_0 and μ_1 and $\sigma_1, \sigma_2: \Gamma \rightarrow [0, 1]$ measurable functions with $\sigma_1 \leq \sigma_2$. Then the restriction*

$$\pi_{\sigma_1, \sigma_2} := (\text{ev} \circ (\text{id} \times \sigma_1), \text{ev} \circ (\text{id} \times \sigma_2))_\# \Pi$$

is an optimal coupling of $\mu_{\sigma_1} := (\text{ev} \circ (\text{id} \times \sigma_1))_{\#} \Pi$ and $\mu_{\sigma_2} := (\text{ev} \circ (\text{id} \times \sigma_2))_{\#} \Pi$. If furthermore $(\sigma_1, \sigma_2) \neq (0, 1)$ Π -almost everywhere then π_{σ_1, σ_2} is the unique optimal coupling of μ_{σ_1} and μ_{σ_2} .

Proof. By the triangle inequality for c_L and the parameterization invariance of \mathcal{A} one has

$$C_L(\mu_0, \mu_1) \leq C_L(\mu_0, \mu_{\sigma_1}) + C_L(\mu_{\sigma_1}, \mu_{\sigma_2}) + C_L(\mu_{\sigma_2}, \mu_1)$$

and

$$\int c_L d\pi_{0,1} = \int c_L d\pi_{0,\sigma_1} + \int c_L d\pi_{\sigma_1,\sigma_2} + \int c_L d\pi_{\sigma_2,1}.$$

Since $\int c_L d\pi_{0,1} = C_L(\mu_0, \mu_1)$ and C_L is the minimal cost, the three terms on the right-hand sides must individually coincide. More precisely, if one coupling on the right-hand side, without loss of generality π_{0,σ_1} , is not optimal, one can replace this coupling by an optimal coupling π' with strictly smaller cost. Let π' be an optimal coupling of $(\text{ev}_0)_{\#} \Pi$ and $(\text{ev} \circ (\text{id} \times \sigma_1))_{\#} \Pi$. Then one has

$$\int c_L d\pi_{0,1} > \int c_L d\pi' + \int c_L d\pi_{\sigma_1,\sigma_2} + \int c_L d\pi_{\sigma_2,1}.$$

Gluing the three couplings gives a coupling of μ_0 and μ_1 with strictly smaller cost, a contradiction. The second statement follows directly from the triangle inequality for c_L , see Section 2. \square

Corollary 3.12. *Let $(\mu_0, \mu_1) \in \mathcal{P}_\tau^+(M)$. Further let Π be a dynamical optimal coupling of μ_0 and μ_1 . If Ξ is a measure on Γ , such that $\Xi \leq \Pi$ and $\Xi(\Gamma) > 0$, set*

$$\Xi' := \frac{\Xi}{\Xi(\Gamma)} \quad \text{and} \quad \nu_i := (\text{ev}_i)_{\#} \Xi'$$

for $i = 0, 1$. Then Ξ' is a dynamical optimal coupling between ν_0 and ν_1 .

Proof. The assumption $\Xi \leq \Pi$ implies that $\Pi - \Xi$ is a measure on Γ . Set

$$\pi := (\text{ev}_0, \text{ev}_1)_{\#} \Pi, \quad \pi^0 := (\text{ev}_0, \text{ev}_1)_{\#} \Xi, \quad \pi^1 := (\text{ev}_0, \text{ev}_1)_{\#} (\Pi - \Xi).$$

Then one has $\pi = \pi^0 + \pi^1$ and

$$(7) \quad \int c_L d\pi = \int c_L d\pi^0 + \int c_L d\pi^1.$$

Since the left-hand side of (7) is optimal, so must be the terms on the right-hand side. Here optimality of π^0 means optimality of the cost $\int c_L d\pi'$ among all Borel measures π' on $M \times M$ with $\pi'(M \times M) = \pi^0(M \times M) = \Xi(\Gamma)$ and martingales equal to $\Xi(\Gamma)\nu_0$ and $\Xi(\Gamma)\nu_1$. Optimality of π^1 is defined analogously. Now if π^0 was not optimal, one could replace π^0 by a coupling σ of $(\text{ev}_0)_{\#} \Xi$ and $(\text{ev}_1)_{\#} \Xi$ with strictly smaller cost. $\sigma + \pi^1$ is a coupling of μ_0 and μ_1 by construction, but $\int c_L d\pi > \int c_L d\sigma + \int c_L d\pi^1$, a contradiction. This shows that Ξ' is a dynamical optimal coupling. \square

3.13. Intermediate regularity of dynamical optimal couplings. Recall that $\mathbb{U} \subseteq \mathbb{R} \times TM$ denotes the maximal domain of the Euler–Lagrange flow $\Phi^{\mathbb{L}}$ of \mathbb{L} . Let U be defined by $\{1\} \times U := (\{1\} \times TM) \cap \mathbb{U}$. It is a fiberwise star-shaped neighborhood of the zero section. Define the exponential map $\exp^{\mathbb{L}}: U \rightarrow M \times M$ by

$$\exp^{\mathbb{L}}(v) := (\pi_{TM}(v), \pi_{TM} \circ \Phi^{\mathbb{L}}(1, v)).$$

Proposition 3.14. $\exp^{\mathbb{L}}$ is a C^1 -diffeomorphism on a neighborhood of T^0M in TM onto its image. Further $\exp^{\mathbb{L}}$ is smooth outside T^0M .

Proof. Choose local coordinates $V \rightarrow \mathbb{R}^m$ on M and consider the induced coordinates $TV \rightarrow T\mathbb{R}^m \cong \mathbb{R}^m \times \mathbb{R}^m$ on TM . The chart $V \rightarrow \mathbb{R}^m$ induces coordinates $V \times V \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ on $M \times M$ as well.

Abbreviate $\Phi_t^{\mathbb{L}} := \Phi^{\mathbb{L}}(t, \cdot)$. In order to show continuous differentiability it suffices to show this for $\pi_{TM} \circ \Phi_1^{\mathbb{L}}$ at the zero section, since π_{TM} is everywhere smooth and $\Phi_1^{\mathbb{L}}$ is smooth outside of the zero section.

(i) For all $(W, Z) \in \mathbb{R}^m \times \mathbb{R}^m$ the directional derivatives

$$(p, v) \mapsto \partial_{(W, Z)}(\pi_{TM} \circ \Phi_1^{\mathbb{L}})_{(p, v)}$$

exist on U and

$$d(\exp^{\mathbb{L}})_{0_p}(W, Z) = (W, W + Z)$$

in the above coordinates. The only place to check is the zero section. Let $(W, Z) \in T(TM)_{0_p}$ for $p \in V$. Then one has

$$(8) \quad \frac{1}{t}(\pi_{TM} \circ \Phi_1^{\mathbb{L}}(p + tW, tZ) - \pi_{TM} \circ \Phi_1^{\mathbb{L}}(p, 0)) = \frac{1}{t}(\pi_{TM} \circ \Phi_1^{\mathbb{L}}(p + tW, tZ) - p).$$

For $Z = 0$ one has

$$\frac{1}{t}(\pi_{TM} \circ \Phi_1^{\mathbb{L}}(p + tW, 0) - p) = \frac{1}{t}(p + tW - p) = W.$$

For $Z \neq 0$ it follows that the right-hand side of (8) converges for $t \rightarrow 0$ to

$$\begin{aligned} & d\pi_{TM} \left(\left. \frac{d}{dt} \right|_{t=0} \Phi_t^{\mathbb{L}}(p + tW, Z) \right) \\ &= d\pi_{TM} \left(\left. \frac{d}{dt} \right|_{t=0} \Phi_0^{\mathbb{L}}(p + tW, Z) + \left. \frac{d}{dt} \right|_{t=0} \Phi_t^{\mathbb{L}}(p, Z) \right) \\ &= d\pi_{TM}((W, 0) + (Z, 0)) = W + Z \end{aligned}$$

since $d\pi_{TM}(W, Z) = W$. This proves the claim.

(ii) The directional derivatives $(p, v) \mapsto d(\pi_{TM} \circ \Phi_1^{\mathbb{L}})_{(p, v)}(W, Z)$ are continuous. This implies that $\exp^{\mathbb{L}}$ is C^1 on U by a standard theorem of calculus. The diffeomorphism property follows from the inverse function theorem and (i), since

$$d(\exp^{\mathbb{L}})_{0_p}(W, Z) = (W, W + Z).$$

In order to show that the directional derivatives are continuous let $(p, v) \in TV \cap U$ with $v \neq 0$ and $(W, Z) \in T(TM)_{(p,v)}$. One has

$$\begin{aligned} & \partial_{(W,Z)}(\pi_{TM} \circ \Phi_1^{\mathbb{L}})_{(p,v)} \\ &= \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_1^{\mathbb{L}}(p + tW, v) + \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_1^{\mathbb{L}}(p, v + tZ), \end{aligned}$$

since $\pi_{TM} \circ \Phi_1^{\mathbb{L}}$ is smooth around (p, v) .

Setting $\varepsilon := \sqrt{h(v, v)}$, one has for the first term

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_1^{\mathbb{L}}(p + tW, v) &= \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_\varepsilon^{\mathbb{L}} \left(p + tW, \frac{v}{\varepsilon} \right) \\ &= d\pi_{TM}(d\Phi_\varepsilon^{\mathbb{L}})_{(p,v/\varepsilon)}(W, 0). \end{aligned}$$

The last term converges to W uniformly on compact subsets of M for $\varepsilon \rightarrow 0$, because $\frac{v}{\varepsilon}$ is bounded away from the zero section.

For the second term one has

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_1^{\mathbb{L}}(p, v + tZ) &= \frac{d}{dt} \Big|_{t=0} \pi_{TM} \circ \Phi_\varepsilon^{\mathbb{L}} \left(p, \frac{v + tZ}{\varepsilon} \right) \\ &= d\pi_{TM}(d\Phi_\varepsilon^{\mathbb{L}})_{(p,v/\varepsilon)} \left(0, \frac{Z}{\varepsilon} \right). \end{aligned}$$

The last term equals $Y(\varepsilon)$ where Y is the unique solution of the Jacobi equation of \mathbb{L} along $\gamma: t \mapsto \pi_{TM} \circ \Phi_t^{\mathbb{L}}(p, \frac{v}{\varepsilon})$ with $Y(0) = 0$ and $\dot{Y}(0) = \frac{Z}{\varepsilon}$. Since $\frac{v}{\varepsilon} \neq 0$, one can write

$$\begin{pmatrix} Y \\ \dot{Y} \end{pmatrix} (t) = \exp \left[\int_0^t A_\gamma(\sigma) d\sigma \right] \begin{pmatrix} 0 \\ \frac{Z}{\varepsilon} \end{pmatrix}$$

for a curve $t \rightarrow A_\gamma(t)$ of $2m \times 2m$ matrices. Since $\frac{v}{\varepsilon}$ is bounded away from the zero section, A_γ is uniformly bounded for bounded flow parameters t . Further since the Jacobi equation is an equation of second order, $A_\gamma(t)$ has the form

$$A_\gamma(t) = \begin{pmatrix} 0 & E_n \\ B_\gamma(t) & C_\gamma(t) \end{pmatrix}.$$

Thus

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} Y(\varepsilon) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} d\pi_{TM} \left(\exp \left[\int_0^\varepsilon A_\gamma(\sigma) d\sigma \right] \begin{pmatrix} 0 \\ Z \end{pmatrix} \right) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (E_n 0 + 0 + \varepsilon \cdot E_n Z) = Z. \end{aligned}$$

This shows that the partial derivatives are continuous. □

Denote by \mathcal{A}_τ the action of L_τ , i.e. for $\eta: [s, t] \rightarrow N$ set

$$\mathcal{A}_\tau(\eta) := \int_s^t L_\tau(\sigma, \dot{\eta}(\sigma)) d\sigma \in \mathbb{R} \cup \{\infty\}.$$

Define the sets $J^\pm((s, x))$ and $I^\pm((s, x))$ via the splitting $M \cong \mathbb{R} \times N$.

Lemma 3.15. *Every $(s, x) \in \mathbb{R} \times N$ has a neighborhood $V \subseteq \mathbb{R} \times N$ such that for every $(t, y) \in V \cap J^+((s, x))$ the unique Φ_τ -trajectory $\eta: [s, t] \rightarrow N$ from x to y strictly minimizes \mathcal{A}_τ among all curves $\alpha: [s, t] \rightarrow N$ from x to y .*

Proof. As usual define $\exp_p^\perp := \pi_{TM} \circ \Phi^\perp(1, \cdot)|_{U \cap TM_p}$ at $p \in M$. According to Proposition 3.14 every point $p \in M$ admits a normal neighborhood V , i.e. \exp_p^\perp is a diffeomorphism from a neighborhood of 0_p onto V . Further according to [17] every point $q \in J_V^+(p)$ (recall the definition from Section 2) is connected to p via a unique Φ^\perp -trajectory γ with $\gamma(0) = p$ and $\gamma(1) = q$. This γ strictly minimizes \mathcal{A} among all causal curves in V from p to q up to monotone reparameterizations. Since $\tau(q) - \tau(p)$ bounds the h -length of a causal curve between p and q (see Section 2), every causal curve between p and q is contained in V given p and q belong to a sufficiently small subneighborhood. Thus the Φ^\perp -trajectory γ is strictly minimal among all causal curves in M from p to q up to monotone reparameterizations. Since causal curves are the only curves in M with finite \mathcal{A} -action, the strict minimality up to monotone reparameterizations even holds for all curves in M between p and q .

Since $\mathcal{A}_\tau(\eta) = \mathcal{A}(H)$ for every curve $\eta: [s, t] \rightarrow N$, where $H: [s, t] \rightarrow M$ denotes the graph of η , the local minimality follows for the Φ_τ -trajectories. Strict minimality follows from the fact that the conditions $\tau \circ H(u) = u$ fix the parameterization of $H: [s, t] \rightarrow M$ uniquely. Define $V \subset \mathbb{R} \times N$ via the identification of $\mathbb{R} \times N \cong M$. \square

Remark 3.16. For $(s, x) \in \mathbb{R} \times N$, $(t, y) \in U \cap J^+((s, x))$ as in Lemma 3.15 and $\eta: [s, t] \rightarrow N$ the unique Φ_τ -trajectory from x to y , one has

$$\mathcal{A}_\tau(\gamma) = L(v),$$

where $v = (\exp_{(s,x)}^\perp)^{-1}(t, y)$. Further denote by $\mathcal{S}_s^t(x, y)$ the minimal action of a curve from x to y with respect to \mathcal{A}_τ . Then the previous equality and Lemma 3.15 imply

$$\mathcal{S}_s^t(x, y) = L(v)$$

with v as before. Thus $(t, y) \mapsto \mathcal{S}_s^t(x, y)$ is smooth for $(t, y) \in I_V^+((s, x))$ and V as in Lemma 3.15 as follows from Proposition 3.14.

Proposition 3.17. *Let $\varepsilon > 0$ and let $I \times K \subseteq \mathbb{R} \times N$ be a compact subset. Then there exist $\delta, \kappa > 0$ and $C < \infty$ such that for $a, b, c \in I$ with $b - a, c - b \geq \varepsilon$ and Φ_τ -trajectories $x_i: [a, c] \rightarrow N$, $i = 1, 2$, with $\text{dist}(x_1(b), x_2(b)) \leq \delta$, $x_i(b) \in K$ and*

$$\text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2 \geq C \text{dist}(x_1(b), x_2(b))$$

there exist \mathcal{A}_τ -minimizers $y_i: [a, c] \rightarrow N$ with $y_1(a) = x_1(a)$, $y_1(c) = x_2(c)$, $y_2(a) = x_2(a)$, $y_2(c) = x_1(c)$ and

$$\mathcal{A}_\tau(y_1) + \mathcal{A}_\tau(y_2) - \mathcal{A}_\tau(x_1) - \mathcal{A}_\tau(x_2) \leq -\kappa \text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2.$$

A curve $\eta: [a, b] \rightarrow N$ is *causal* if the graph H is causal in $\mathbb{R} \times N \cong M$.

Lemma 3.18. *Let $I \times K \subseteq \mathbb{R} \times N$ be compact and $\varepsilon \in (0, 1)$. Then there exists $\delta > 0$ such that for all $a, b, c \in I$ with $b - a, c - b \geq \varepsilon$ and causal \mathcal{A}_τ -minimizers $\eta: [a, b] \rightarrow N$, $\gamma: [b, c] \rightarrow N$ with $\eta(b) = \gamma(b)$ and $\dot{\eta}(b) \neq \dot{\gamma}(b)$ one has*

$$\mathcal{S}_a^c(\eta(a), \gamma(c)) - \mathcal{A}_\tau(\eta) - \mathcal{A}_\tau(\gamma) \leq -\frac{\delta}{|\mathcal{S}_a^c(\eta(a), \gamma(c))|} |\dot{\eta}(b) - \dot{\gamma}(b)|^2.$$

Proof. First notice that it suffices to prove the statement for $(b, \dot{\eta}(b)), (b, \dot{\gamma}(b)) \in \text{int } \mathcal{D}_\tau$ since δ is claimed to be independent of γ and η . The assertion then follows for $(b, \dot{\eta}(b)), (b, \dot{\gamma}(b)) \in \mathcal{D}_\tau$ via approximating them by tangent vectors $(b, \dot{\eta}(b)), (b, \dot{\gamma}(b)) \in \text{int } \mathcal{D}_\tau$ and using the continuity of \mathcal{A}_τ and \mathcal{S} .

Choose $\delta > 0$ such that

$$(\partial_v^2 L_\tau)_{(s,v)} \geq \frac{\delta}{|L_\tau(s, v)|} \text{id}$$

for all $(s, v) \in \text{int } \mathcal{D}_\tau$ with $s \in I$ and v based at a point in K according to Lemma 3.6 (ii). Cover $I \times K$ with finitely many neighborhoods $V \subset \mathbb{R} \times N$ according to Lemma 3.15. Choose $\varepsilon > \varepsilon' > 0$ such that for every $(s, x) \in I \times K$ the open set $W := (s - \varepsilon', s + \varepsilon') \times B_{\varepsilon'}(x)$ is contained in at least one V . It suffices to prove the claim for a', b, c' with $b - a' = c' - b = \varepsilon'$ since concatenating an \mathcal{A}_τ -minimizer from $\eta(a')$ to $\gamma(c')$ with the arcs $\eta|_{[a, a']}$ and $\gamma|_{[c', c]}$ only decreases the left-hand side as well as increases the right-hand side. The proof continues to use a and c instead of a' and c' though.

For $(t, y) \in I^+(a, \eta(a)) \cap W$ denote by $Y_{t,y}$ the tangent at t to the unique Φ_τ -trajectory on $[a, \tau(y)]$ from $\eta(a)$ to y . The map $(t, y) \mapsto Y_{t,y}$ is smooth for $(t, y) \in I^+((a, \eta(a)))$ by Lemma 3.8 and Proposition 3.14. The inequality

$$\mathcal{S}_a^t(\eta(a), \gamma(t)) \leq \mathcal{A}_\tau(\gamma|_{[b,t]}) + \mathcal{S}_a^b(\eta(a), \eta(b))$$

for $b \leq t \leq c$ implies

$$\partial_t|_{t=s} \mathcal{S}_a^t(\eta(a), \gamma(t)) \leq L_\tau(s, \dot{\gamma}(s))$$

with equality if and only if $\dot{\gamma}(s) = Y_{s, \gamma(s)}$. One has

$$\partial_t|_{t=s} \mathcal{S}_a^t(\eta(a), \gamma(t)) = (\partial_t|_{t=s} \mathcal{S}_a^t)(\eta(a), \gamma(s)) + (\partial_2 \mathcal{S}_a^s)_{(\eta(a), \gamma(s))}(\dot{\gamma}(s)),$$

i.e. $\partial_t|_{t=s} \mathcal{S}_a^t(\eta(a), \gamma(t))$ is an affine function of $\dot{\gamma}(s) \in TN_{\gamma(s)}$. Thus it must coincide with the tangent to L_τ at $Y_{s, \gamma(s)}$, i.e.

$$\partial_s|_{t=s} \mathcal{S}_a^t(\eta(a), \gamma(t)) = L_\tau(s, Y_{s, \gamma(s)}) + (\partial_v L_\tau)_{Y_{s, \gamma(s)}}(\dot{\gamma}(s) - Y_{s, \gamma(s)}).$$

Then one has

$$\begin{aligned} L_\tau(s, \dot{\gamma}(s)) &\geq L_\tau(s, Y_{s, \gamma(s)}) + \partial_v L_\tau(\dot{\gamma}(s) - Y_{s, \gamma(s)}) \\ &\quad + \frac{\delta}{2|L_{\min}(s)|} |\dot{\gamma}(s) - Y_{s, \gamma(s)}|^2 \\ &= \partial_s [\mathcal{S}_a^s(\eta(a), \gamma(s))] + \frac{\delta}{2|L_{\min}(s)|} |\dot{\gamma}(s) - Y_{s, \gamma(s)}|^2. \end{aligned}$$

Claim. Denote by S a lower bound of $\mathcal{S}_s^t(x, y)$ for $(s, x), (t, y) \in I \times K$ and set

$$C_1 := 2 \exp\left(\frac{1}{\delta} \left(\frac{2S}{\varepsilon}\right)^2\right).$$

Further denote by $L_{\min}(s)$ the minimum of L_τ on the convex hull of $\dot{\gamma}(s)$ and $Y_{s, \gamma(s)}$. Then there exists a subset B of $[0, \varepsilon]$ of measure at least $\frac{\varepsilon}{2}$ such that $|L_{\min}(s)| \leq C_1 |L_{Y_{s, \gamma(s)}}|$ for all $s \in B$.

Proof of the claim. For $v, w \in \text{int } \mathcal{D}_{(s, x)}$ with $v \neq w$ and $(s, x) \in I \times K$ consider the convex hull $\text{conv}\{v, w\}$. Parameterize $\text{conv}\{v, w\}$ by

$$f: \lambda \in [0, |v - w|] \mapsto \left(1 - \frac{\lambda}{|v - w|}\right)v + \frac{\lambda}{|v - w|}w.$$

Next denote by L_{\min} the minimum of $L \circ f$ and by λ_0 with $0 \leq \lambda_0 \leq |v - w|$ the parameter achieving this minimum. Let $Y \in \text{conv}\{v, w\}$ and denote

$$L'_Y := \frac{d}{d\lambda} \Big|_{\lambda=\lambda_1} (L \circ f)(\lambda)$$

where $f(\lambda_1) = Y$. Then one has

$$\begin{aligned} (L'_Y)^2 &= |L'_Y| \cdot |0 - L'_Y| \\ &= |L'_Y| \cdot \left| \int_{\lambda_1}^{\lambda_0} (L \circ f)''(\lambda) d\lambda \right| \\ &\geq |L'_Y| \cdot \delta \left| \int_{\lambda_1}^{\lambda_0} \frac{1}{|L \circ f(\lambda)|} d\lambda \right| \\ &\geq \delta \int_{\lambda_1}^{\lambda_0} \frac{(L \circ f)'(\lambda)}{|L \circ f(\lambda)|} d\lambda = \delta \log \left| \frac{L_{\min}}{L_Y} \right| \end{aligned}$$

where the last two manipulations follow from the convexity of $L \circ f$.

Now assume that $|L_{\min}(s)| \geq C_1 |L_{Y_{s, \gamma(s)}}|$ on a set $B \subseteq [b, c]$ of measure at least $\frac{\varepsilon}{2}$. Then from the first paragraph one has $|L'_{Y_{s, \gamma(s)}}| \geq \sqrt{\delta \log C_1}$ on B . Note that this implies $L'_{Y_{s, \gamma(s)}} \leq 0$ since

$$0 \geq \partial_s \mathcal{S}_a^s(\eta(a), \gamma(s)) = L_\tau(s, Y_{s, \gamma(s)}) + L'_{Y_{s, \gamma(s)}}$$

implies $L'_{Y_{s, \gamma(s)}} \leq -L_\tau(s, Y_{s, \gamma(s)}) \leq -S$ by Remark 3.16. Consequently,

$$\partial_s \mathcal{S}_a^s(\eta(a), \gamma(s)) = L_\tau(s, Y_{s, \gamma(s)}) + L'_{Y_{s, \gamma(s)}} \leq L'_{Y_{s, \gamma(s)}} < -\sqrt{\delta \log C_1}$$

which implies

$$\mathcal{S}_a^c(\eta(a), \gamma(c)) - \mathcal{S}_a^b(\eta(a), \gamma(b)) \leq \int_B \partial_s \mathcal{S}_a^s(\eta(a), \gamma(s)) ds \leq -\frac{\varepsilon}{2} \sqrt{\delta \log C_1}.$$

This constitutes a contradiction to the definition of C_1 . □

The claim thus implies

$$\begin{aligned} \mathcal{S}_a^c(\eta(a), \gamma(c)) - \mathcal{S}_a^b(\eta(a), \gamma(b)) &\leq \mathcal{A}_\tau(\gamma) - \frac{\delta}{2} \int_b^c \frac{|\dot{\gamma}(s) - Y_{s,\gamma(s)}|^2}{|L_{\min}(s)|} ds \\ &\leq \mathcal{A}_\tau(\gamma) - \frac{\delta}{2C_1} \int_B \frac{|\dot{\gamma}(s) - Y_{s,\gamma(s)}|^2}{|L_{Y_{s,\gamma(s)}}|} ds. \end{aligned}$$

Next note that the continuity of Φ_τ and the invariance of $\partial\mathcal{D}_\tau = L_\tau^{-1}(0)$ under Φ_τ implies the existence of a $C_2 < \infty$ depending only on $I \times K$ such that $(s-a)|L_{Y_{s,\gamma(s)}}| \leq C_2 |S_a^s(\eta(a), \gamma(s))|$ for all $t \in [b, c]$. Thus one has

$$\begin{aligned} \mathcal{S}_a^c(\eta(a), \gamma(c)) - \mathcal{S}_a^b(\eta(a), \gamma(b)) &\leq \mathcal{A}_\tau(\gamma) - \delta_1 \int_B \frac{|\dot{\gamma}(s) - Y_{s,\gamma(s)}|^2}{|S_a^s(\eta(a), \gamma(s))|} ds \\ &\leq \mathcal{A}_\tau(\gamma) - \frac{\delta_1}{|S_a^c(\eta(a), \gamma(c))|} \int_B |\dot{\gamma}(s) - Y_{s,\gamma(s)}|^2 ds. \end{aligned}$$

Note that again due to the continuity of Φ_τ there exists $C_3 < \infty$ depending only on $I \times K$ such that for all $t \in [b, c]$ one has $|\dot{\gamma}(t) - Y_{t,\gamma(t)}| \leq C_3 |\dot{\gamma}(b) - Y_{b,\gamma(b)}|$. This follows from the fact that the image of Y is locally invariant under Φ_τ . Thus there exists $\delta_3 > 0$ such that

$$\mathcal{S}_a^c(\eta(a), \gamma(c)) - \mathcal{S}_a^b(\eta(a), \gamma(b)) \leq \mathcal{A}_\tau(\gamma) - \frac{\delta_3}{|S_a^c(\eta(a), \gamma(c))|} |\dot{\gamma}(b) - Y_{b,\gamma(b)}|^2.$$

Finally, notice that $Y_{b,\gamma(b)} = \dot{\eta}(b)$. This finishes the proof. \square

Proof of Proposition 3.17. With the same argument as in the preceding proof one can assume that the curves are contained in a normal neighborhood.

(i) The first step is to show that

$$(c, x_2(c)) \in J^+((a, x_1(a)))$$

under the assumptions in the proposition and for $0 < C < \infty$ sufficiently large. Using the local Lipschitz continuity of

$$(s, x) \mapsto \mathcal{D}_{(s,x)} = \mathcal{D}_\tau \cap (\{s\} \times TN_x)$$

with respect to the Hausdorff distance, which follows directly from local Lipschitz continuity of $p \mapsto \mathcal{C} \cap \{d\tau = 1\}$ with respect to the Hausdorff distance, one sees that there exists $C_0 < \infty$ only depending on $I \times K$ such that one can choose $(b, \dot{\chi}_2(b)) \in \mathcal{D}_{(b,x_1(b))}$ with

$$(9) \quad \text{dist}(\dot{\chi}_2(b), \dot{x}_2(b)) \leq C_0 \text{dist}(\chi_2(b), x_2(b)) = C_0 \text{dist}(x_1(b), x_2(b)).$$

With the smoothness of Φ_τ this then implies

$$\text{dist}(\chi_2(c), x_2(c)) \leq C_1 \text{dist}(x_1(b), x_2(b))$$

for some $C_1 < \infty$ only depending on $I \times K$ and ε . Now the triangle inequality yields

$$\begin{aligned} & \text{dist}((c, x_2(c)), \partial J^+((a, x_1(a)))) \\ & \geq \text{dist}((c, \chi_2(c)), \partial J^+((a, x_1(a)))) - C_1 \text{dist}(x_1(b), x_2(b)). \end{aligned}$$

Recall from Remark 3.16 that

$$\mathcal{S}_a^c(x_1(a), \chi_2(c))^2 = L((\exp_{(a, x_1(a))}^{\mathbb{L}})^{-1}((c, \chi_2(c))))^2$$

and that L^2 is smooth up to and beyond $\partial \mathcal{D}_\tau$. Thus there exists $\delta_0 > 0$, only depending on $I \times K$ and ε , such that

$$\text{dist}((c, \chi_2(c)), \partial J^+((a, x_1(a)))) \geq \delta_0 \mathcal{S}_a^c(x_1(a), \chi_2(c))^2.$$

Now Lemma 3.18 implies $\mathcal{S}_a^c(x_1(a), \chi_2(c))^2 \geq \delta_2 \text{dist}(\dot{x}_1(b), \dot{\chi}_2(b))^2$ as a special case. With (9) one then obtains

$$\mathcal{S}_a^c(x_1(a), \chi_2(c))^2 \geq \delta_3 \text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2$$

for some $\delta_3 > 0$ if C is sufficiently large. Consequently, one has

$$\text{dist}((c, x_2(c)), \partial J^+((a, x_1(a)))) \geq \frac{1}{2} \text{dist}((c, \chi_2(c)), \partial J^+((a, x_1(a))))$$

for $\text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2 \geq C \text{dist}(x_1(b), x_2(b))$ with C sufficiently large. Thus one concludes $(c, x_2(c)) \in J^+((a, x_1(a)))$.

(ii) Remark 3.16 implies that

$$\mathcal{S}_a^c(x_1(a), x_2(c))^2 \geq \delta_4 \text{dist}((c, x_2(c)), J^+((a, x_1(a))))$$

for some $\delta_4 > 0$ depending only on $I \times K$ and ε , since the fiber derivative of L_τ^2 does not vanish anywhere on $\partial \mathcal{D}_\tau$. Thus one has

$$\mathcal{S}_a^c(x_1(a), x_2(c))^2 \geq \frac{\delta_4 \delta_0}{2} \mathcal{S}_a^c(x_1(a), \chi_2(c))^2.$$

With the convexity of \mathcal{S} one then concludes

$$\begin{aligned} (10) \quad & |\mathcal{S}_a^c(x_1(a), x_2(c)) - \mathcal{S}_a^c(x_1(a), \chi_2(c))| \\ & \leq \frac{C_2}{|\mathcal{S}_a^c(x_1(a), \chi_2(c))|} \text{dist}(x_1(b), x_2(b)) \end{aligned}$$

for some $C_2 < \infty$ depending only on $I \times K$ and ε . Finally, one has

$$\begin{aligned} & \mathcal{S}_a^c(x_1(a), x_2(c)) - \mathcal{S}_a^b(x_1(a), x_1(b)) - \mathcal{S}_b^c(x_2(b), x_2(c)) \\ & \leq \mathcal{S}_a^c(x_1(a), \chi_2(c)) - \mathcal{S}_a^b(x_1(a), x_1(b)) - \mathcal{S}_b^c(\chi_2(b), \chi_2(c)) \\ & \quad + |\mathcal{S}_a^c(x_1(a), \chi_2(c)) - \mathcal{S}_a^c(x_1(a), x_2(c))| \\ & \quad + |\mathcal{A}_\tau(x_2|_{[b,c]}) - \mathcal{A}_\tau(\chi_2)|. \end{aligned}$$

The first term on the right-hand side is bounded from above by

$$-\frac{\delta}{|\mathcal{S}_a^c(x_1(a), \chi_2(c))|} \text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2$$

according to Lemma 3.18 and the choice of χ_2 for some $\delta > 0$. The second term is bounded from above by

$$\frac{C_2}{|\mathcal{S}_a^c(x_1(a), \chi_2(c))|} \text{dist}(x_1(b), x_2(b))$$

according to (10). Finally, one has

$$|\mathcal{A}_\tau(x_2|_{[b,c]}) - \mathcal{A}_\tau(\chi_2)| \leq C_3 \text{dist}(x_1(b), x_2(b))$$

according to (9) for some $C_3 < \infty$. Now note that

$$\text{dist}(\dot{x}_1(b), \dot{\chi}_2(b)) \geq \text{dist}(\dot{x}_1(b), \dot{x}_2(b)) - C_0 \text{dist}(x_1(b), x_2(b))$$

by the triangle inequality and (9). This implies that

$$\mathcal{S}_a^c(x_1(a), x_2(c)) - \mathcal{S}_a^b(x_1(a), x_1(b)) - \mathcal{S}_b^c(x_2(b), x_2(c)) \leq -\frac{\delta}{2} \text{dist}(\dot{x}_1(b), \dot{x}_2(b))$$

for $C < \infty$ sufficiently large.

Repeating the arguments for $\mathcal{S}_a^c(x_2(a), \cdot)$, $x_2|_{[a,b]}$ and $x_1|_{[b,c]}$, one obtains

$$\begin{aligned} \mathcal{S}_a^c(x_1(a), x_1(c)) + \mathcal{S}_a^c(x_2(a), x_2(c)) - \mathcal{S}_a^c(x_1(a), x_2(c)) - \mathcal{S}_a^c(x_2(a), x_1(c)) \\ \leq -\delta \text{dist}(\dot{x}_1(b), \dot{x}_2(b))^2 \end{aligned}$$

for $C < \infty$ sufficiently large. The claim is now immediate for $y_1: [a, c] \rightarrow N$ the \mathcal{A}_τ -minimizer from $x_1(a)$ to $x_2(c)$ and $y_2: [a, c] \rightarrow N$ the \mathcal{A}_τ -minimizer from $x_2(a)$ to $x_1(c)$. \square

Proof of Theorem 2.10. Let Π be a dynamical optimal coupling of μ and ν . For $k \in \mathbb{N}$ consider the subcoupling

$$\Pi_k := \Pi|_{\{\gamma | \tau(\gamma(1)) - \tau(\gamma(0)) \geq 1/k\}}.$$

Since the supports of μ and ν are disjoint, one knows that for every compact set $I \times K \subseteq \mathbb{R} \times N \cong M$ there exists k with

$$(\text{ev}_t)_\# \Pi|_{I \times K} \equiv (\text{ev}_t)_\# \Pi_k|_{I \times K}$$

for all $t \in [0, 1]$. Fix $I \times K \subset M$ compact and $k \in \mathbb{N}$ such that $\gamma \in \text{supp } \Pi_k$ for all $\gamma \in \text{supp } \Pi$ with $\gamma \subset I \times K$.

Consider the reparameterization $\eta: [\tau(\gamma(0)), \tau(\gamma(1))] \rightarrow M$ of $\gamma \in \text{supp } \Pi_k$ such that $\tau \circ \eta(s) = s$. Next let $\varepsilon_0 > 0$ be given and consider the restriction of $\gamma \in \text{supp } \Pi_k$ to $[\varepsilon_0, 1 - \varepsilon_0]$. Then there exists $\varepsilon_1 > 0$ only depending on ε_0 such that $|s(t) - \tau(\gamma(0))|, |s(t) - \tau(\gamma(1))| \geq 2\varepsilon_1$ for all $\gamma \in \text{supp } \Pi_k$ and the reparameterization η with $\eta(s(t)) \equiv \gamma(t)$ and $t \in [\varepsilon_0, 1 - \varepsilon_0]$.

Let $(\gamma_1, t_1), (\gamma_2, t_2) \in \text{supp } \Pi_k \times [\varepsilon_0, 1 - \varepsilon_0]$. Denote by

$$\eta_i: [\tau(\gamma_i(0)), \tau(\gamma_i(1))] \rightarrow M$$

the reparameterization of γ_i as in the previous paragraph. Since τ is Lipschitz on $I \times K$ with constant $L < \infty$, i.e.

$$|\tau(\gamma_1(t_1)) - \tau(\gamma_2(t_2))| \leq L \text{dist}(\gamma_1(t_1), \gamma_2(t_2)),$$

one has $|b_2 - b_1| \leq L \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2))$ for $b_i = \tau(\gamma_i(t_i))$ and $i = 1, 2$. For $\operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2))$ smaller than ε_1/L one has thus

$$b_1 - \tau(\gamma_2(0)), \tau(\gamma_2(1)) - b_1 \geq \varepsilon_1.$$

Therefore η_2 is well defined on $[b_1 - \varepsilon_1, b_1 + \varepsilon_1]$. With the smoothness of Φ_τ one concludes that there exists a constant $C_0 < \infty$ depending only on ε_0 such that

$$\operatorname{dist}(\dot{\eta}_2(b_1), \dot{\eta}_2(b_2)) \leq C_0 \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2)).$$

Further the triangle inequality implies that

$$\operatorname{dist}(\dot{\eta}_2(b_1), \dot{\eta}_1(b_1)) \geq \operatorname{dist}(\dot{\eta}_1(b_1), \dot{\eta}_2(b_2)) - C_0 \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2))$$

and

$$\operatorname{dist}(\eta_2(b_1), \eta_1(b_1)) \leq C_1 \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2))$$

for some $C_1 < \infty$ depending only on $I \times K$.

Now choose $\delta, \kappa > 0$ and $C < \infty$ for $\varepsilon := \varepsilon_1$ and $I \times K$ according to Proposition 3.17. For $C_2 < \infty$ sufficiently large assuming that

$$C_2 \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2)) < \operatorname{dist}([\dot{\gamma}_1(t_1)], [\dot{\gamma}_2(t_2)])^2$$

and $\operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2)) \leq \min\{\delta/C_1, \varepsilon_1/L\}$ one has

$$C \operatorname{dist}(\eta_2(b_1), \eta_1(b_1)) < \operatorname{dist}(\dot{\eta}_2(b_1), \dot{\eta}_1(b_1))^2.$$

Then Proposition 3.17 implies that

$$\begin{aligned} c_L(\eta_1(b_1 - \varepsilon), \eta_2(b_1 + \varepsilon)) + c_L(\eta_2(b_1 - \varepsilon), \eta_1(b_1 + \varepsilon)) \\ - c_L(\eta_1(b_1 - \varepsilon), \eta_1(b_1 + \varepsilon)) - c_L(\eta_2(b_1 - \varepsilon), \eta_2(b_1 + \varepsilon)) < 0. \end{aligned}$$

With the triangle inequality for c_L it follows that

$$c_L(\gamma_1(0), \gamma_2(1)) + c_L(\gamma_2(0), \gamma_1(1)) - c_L(\gamma_1(0), \gamma_1(1)) - c_L(\gamma_2(0), \gamma_2(1)) < 0,$$

which clearly contradicts the cyclic monotonicity of the optimal coupling $(\operatorname{ev}_0, \operatorname{ev}_1)_\# \Pi$ of μ and ν , see Proposition 2.7. Thus there exists $D < \infty$ with

$$\operatorname{dist}([\dot{\gamma}_1(t_1)], [\dot{\gamma}_2(t_2)])^2 \leq D \operatorname{dist}(\gamma_1(t_1), \gamma_2(t_2)),$$

showing the injectivity of the projection and the Hölder continuity of the inverse. \square

A C^2 -function $L_T: \mathbb{R} \times TN \rightarrow \mathbb{R}$ is a *Tonelli-Lagrangian*, see [4], if for all $(t, x) \in \mathbb{R} \times N$

- (i) the restriction $L_T|_{\{t\} \times TN_x}$ is convex with positive definite Hessian everywhere,
- (ii) $L_T(t, v)/|v| \rightarrow \infty$ as $|v| \rightarrow \infty$ for $v \in TN_x$, and
- (iii) the Euler-Lagrange flow of L_T is complete.

Proof of Theorem 2.12. Choose a compact set $K' \subseteq \operatorname{int} \mathcal{D}_\tau$ such that $K \subseteq \operatorname{int} K'$. Next construct a Tonelli-Lagrangian $L_T: \mathbb{R} \times TN \rightarrow \mathbb{R}$ with $L_T \geq L_\tau$ and $L_T|_{K'} \equiv L_\tau|_{K'}$. Then every \mathcal{A}_τ -minimizer γ with $\dot{\gamma} \in K'$ is also a minimizer for the action induced by L_T . Now the claim follows from the classical regularity result for Tonelli-Lagrangians, see e.g. [4, Thm. A]. \square

3.19. Proof of Theorems 2.13 and 2.14. First the proof of Theorem 2.13 is given. After that the necessary changes to the argument for the proof of Theorem 2.14 are indicated.

The proof of Theorem 2.13 is essentially carried out via two propositions.

Proposition 3.20. *Let $(\mu, \nu) \in \mathcal{P}_\tau^+(M)$. Assume that μ and ν are concentrated on a locally uniformly spacelike hypersurface A and an achronal set B , respectively. Further assume that μ is absolutely continuous with respect to the Lebesgue measure on A and that $\text{supp } \mu$ is ν -neglectable. Then for every optimal coupling π of μ and ν there exists a set $R \subset M \times M$ of full π -measure such that for all $(x, y_1), (x, y_2) \in R$ there exists an \mathcal{A} -minimizer $\gamma: [0, 1] \rightarrow M$ containing x, y_1, y_2 in its trace.*

The proof of Proposition 3.20 needs the following lemma. Recall that $x \in N$ is a *Lebesgue point* of a set $C \subset N$ if

$$\lim_{\delta \rightarrow 0} \frac{\mathcal{L}^n(C \cap B_\delta(x))}{\mathcal{L}^n(B_\delta(x))} = 1,$$

where $B_\delta(x)$ continues to denote the metric ball of radius $\delta > 0$ around x .

Lemma 3.21. *Let N be a manifold, $\mu, \nu \in \mathcal{P}(N)$, $\pi \in \Pi(\mu, \nu)$ and Σ a σ -compact set such that $\pi(\Sigma) = 1$. Assume that μ is absolutely continuous with respect to the Lebesgue measure on N . Then π is concentrated on a σ -compact set $R(\Sigma)$ such that for all $(x, y) \in R(\Sigma) \subset N \times N$ the point x is a Lebesgue point of $\pi_1(\Sigma \cap (N \times \overline{B_r(y)}))$ for all $r > 0$.*

A version for the case $M = \mathbb{R}^n$ is proved in [9, Lem. 4.3]. The proof carries over mutatis mutandis to the present situation of manifolds.

Proof of Proposition 3.20. Let Π be a dynamical optimal coupling of μ and ν . Then $\pi := (ev_0, ev_1)_\# \Pi$ is an optimal coupling of μ and ν . One can assume that \mathcal{A} -minimizers between points in $\text{supp } \mu$ and $\text{supp } \nu$ are unique up to parameterization. This can be seen as follows. By passing to a dynamical subcoupling Ξ' , according to Corollary 3.12, one can first assume that $\text{supp } \mu$ is compact. The proof continues to use the notation Π for the dynamical optimal coupling. By Corollary 3.11 one can assume that the \mathcal{A} -minimizers between points in $\text{supp } \mu$ and $\text{supp } \nu$ are unique up to parameterization by considering the transport $\pi_{0,\sigma}$ between 0 and $\sigma: \Gamma \rightarrow [0, 1]$ with $0 < \sigma(\gamma)$ sufficiently small. One can choose σ such that $\text{supp } \nu$ is compact.

Note that

$$0 = \nu(\text{supp } \mu) = \pi(\text{supp } \mu \times \text{supp } \nu \cap \Delta)$$

where Δ denotes the diagonal in $M \times M$. Thus Π -almost every \mathcal{A} -minimizer is nonconstant. The assumption that μ is concentrated on a locally uniformly spacelike hypersurface implies that every nonconstant causal curve can intersect $\text{supp } \mu$ at most once. Therefore $\text{supp } \mu$ is $(ev_t)_\# \Pi$ -neglectable for all $t > 0$.

Note that since μ and ν are supported on Lipschitz graphs over N , one can consider both measures to be supported on N without losing the absolute continuity of μ with respect to the Lebesgue measure. Therefore one can apply

Lemma 3.21 to μ and ν seen as measures on N and obtain a set $R \subseteq A \times B$ by revoking the identification via the graphs. Choose a set $R = R(\Sigma) \subseteq A \times B$ according to Lemma 3.21.

Assume that there exist $(x, y_1), (x, y_2) \in R$ such that y_i does not lie on the \mathcal{A} -minimizer between x and y_j for $i \neq j$. Then one knows that the tangents $\dot{\gamma}_i(0)$ to the \mathcal{A} -minimizers $\gamma_i \in \Gamma_{x \rightarrow y_i}$ are not parallel. Choose a diffeomorphism ψ from the unit ball $B_1(0)$ in \mathbb{R}^m to a neighborhood U of x with $\psi(0) = x$. For $\delta > 0$ define $\psi_\delta: B_1(0) \rightarrow U, v \mapsto \psi(\delta v)$.

It is obvious that $L_\delta := \frac{1}{\delta} \psi_\delta^*(L)$ converges for $\delta \rightarrow 0$ to $L|_{\mathcal{C}_x}$ uniformly on compact subsets of $\text{int } \mathcal{C}$ in any C^k -topology. Especially the minimizers of the action induced by L_δ converge uniformly to straight lines in $B_1(0)$.

Next choose sequences $\delta_n, r_n \downarrow 0$ such that

$$(11) \quad \lim_{n \rightarrow \infty} \frac{\mathcal{L}_A(\pi_1(\Sigma \cap (A \times \overline{B_{r_n}(y_2)})) \cap \text{im}(\psi_{\delta_n}))}{\mathcal{L}_A(A \cap \text{im}(\psi_{\delta_n}))} = 1$$

where \mathcal{L}_A denotes the Lebesgue measure on A . Since the distance from x to y_2 can be bounded from below and due to the structure of the ψ_δ 's one concludes that the tangents $\dot{\eta}$ at x converge to $\dot{\gamma}_2(0)$ at x for \mathcal{A} -minimizers $\eta \in \Gamma$ connecting a point in $\text{im } \psi_{\delta_n}$ with a point in $B_{r_n}(y_2)$. Further by (11) one can choose points $(w_n, z_n) \in \text{im } \psi_{\delta_n} \times B_{r_n}(y_2)$ with $(w_n, z_n) \in \text{supp } \pi$,

$$\text{dist}(\psi_{\delta_n}^{-1}(x), \psi_{\delta_n}^{-1}(w_n)) \geq \frac{1}{2} \quad \text{and} \quad \psi_{\delta_n}^{-1}(w_n) \rightarrow a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0)$$

with $a, b \in \mathbb{R}$ and $b < 0$. Thus the $L|_{\mathcal{C}_x}$ -minimizer $t \mapsto t \cdot \dot{\gamma}_1(0)$ and $t \mapsto v + t \cdot \dot{\gamma}_2(0)$ intersect for some positive value of t . A simplified version of Proposition 3.17 now shows that this crossing can be shortened by a nonzero amount. Since the convergence is uniform, a fraction of this shortening survives when passing to L_{δ_n} for n sufficiently large. This now contradicts the cyclic monotonicity of the optimal coupling. \square

Consider the set I'_B of \mathcal{A} -minimizers $\gamma' \in \Gamma$ which intersect B in more than one point. Note that $\dot{\gamma}' \in \partial \mathcal{C}$ for all $\gamma' \in I'_B$ since B is achronal. Identify M with $\mathbb{R} \times N$ via the splitting τ as in Section 3.13. Define the set I_B to be the set of reparameterizations γ of $\gamma' \in I'_B$ with $\tau \circ \gamma = \text{id}$. Then the curves in I_B correspond one-to-one with Φ_τ -orbits in N . Denote the set of these Φ_τ -orbits by I_B as well.

Proposition 3.22. *If A is a locally uniformly spacelike hypersurface and B is achronal, then the set formed by the intersections of orbits in I_B with A is \mathcal{L}_A -neglectable.*

Assume for the moment that (i) A is uniformly spacelike, (ii) B is pre-compact, and (iii) the distance between the first and the last intersections of \mathcal{A} -minimizers with B is uniformly bounded from below. Let $(y_1, y_2) \in J^+ \cap (B \times B)$ and let $\gamma \in I_B$ be an \mathcal{A} -minimizer between y_1 and y_2 . Choose $\delta > 0$ such that $\tau(B_\delta(y_1))$ and $\tau(B_\delta(y_2))$ are disjoint. Choose $b \in \mathbb{R}$ between

$\tau(B_\delta(y_1))$ and $\tau(B_\delta(y_2))$. Denote by S_B the set of intersections of curves $\gamma \in I_B$ with $\{b\} \times N$ and let $\mathcal{L}_{\{b\} \times N}$ denote the Lebesgue measure on $\{b\} \times N$.

Lemma 3.23. S_B is an $\mathcal{L}_{\{b\} \times N}$ -neglectable set.

Proof. Consider $\eta \in \Gamma$ with endpoints in $B_\delta(y_1) \cap B$ and $B_\delta(y_2) \cap B$. Denote by z the intersection of η with $\{b\} \times N$. Choose a convex neighborhood U around z disjoint from $B_\delta(y_1) \cap B$ and $B_\delta(y_2) \cap B$. Denote by η_α and η_ω the initial and the terminal point on η in U , respectively. Then one has

$$S_B \cap U \subseteq J^-(\eta_\omega)^c \cap J^+(\eta_\alpha)^c = (J^-(\eta_\omega) \cup J^+(\eta_\alpha))^c.$$

With the same argument as in the proof of Proposition 3.20 one can assume, after possibly restricting U , that $(\{b\} \times N) \cap J^+(\eta_\alpha)$ and $(\{b\} \times N) \cap J^-(\eta_\omega)$ are strictly convex sets. Thus there exists $r > 0$ such that for every point $z_0 \in S_B$ there exist two points $z_1, z_2 \in \{b\} \times N$ with

$$B_r(z_1) \subset (\{b\} \times N) \cap J^+(\eta_\alpha), \quad B_r(z_2) \subset (\{b\} \times N) \cap J^-(\eta_\omega)$$

and

$$B_r(z_1) \cap B_r(z_2) = \{z_0\}.$$

Therefore, for every $\varepsilon > 0$, S_B can be covered by at most ε^n disjoint sets with volume less than ε^{n+1} . This shows that S_B is $\mathcal{L}_{\{b\} \times N}$ -neglectable. \square

Lemma 3.24. The map $S_B \rightarrow TN \cong \{b\} \times TN$ mapping $z \in S_B$ to the tangent vector in \mathcal{D}_z of an \mathcal{A}_τ -minimizer in I_B intersecting z is well defined and Lipschitz.

Proof. Let γ_1 be an \mathcal{A} -minimizer between $y_1, y_2 \in B$, let γ_2 be an \mathcal{A} -minimizer between $y_3, y_4 \in B$, and let both meet at an intermediate point z with different tangent vectors. Then $y_2 \in I^+(y_3)$ and $y_4 \in I^+(y_1)$, which induces a contradiction to the achronality of B . Thus the map is well defined.

Now let $x, z \in S_B$ and let $\gamma_x, \gamma_z \in I_B$ contain respectively x and z in their traces. Choose $y_1, y_2 \in B$ such that γ_z connects y_1 and y_2 . Then one has $x \in I^+(y_1)^c \cap I^-(y_2)^c$ by the achronality of B . Therefore

$$\text{dist}(x, I^+(y_1)), \text{dist}(x, I^-(y_2)) \leq C_1 \text{dist}(x, z)^2$$

for some $C_1 < \infty$ depending only on $\text{supp } \mu \cup \text{supp } \nu$.

Choose $w \in \partial J^+(y_1) \cap (\{b\} \times N)$ the nearest point to x . Then there exist $C_2 < \infty$ and $(b, \dot{\chi}(b)) \in \mathcal{D}_w$ with

$$(12) \quad \text{dist}((b, \dot{\chi}(b)), \dot{\gamma}_x(b)) \leq C_2 \text{dist}(w, z).$$

Recall that one has assumed that the distance between the intersections of \mathcal{A} -minimizers with B is bounded from below. Therefore there exists $\varepsilon > 0$ such that

$$\max \tau|_{B_\delta(y_1)} < b - \varepsilon < b + \varepsilon < \min \tau|_{B_\delta(y_2)}.$$

Then by Lemma 3.18 there exists $\delta_1 > 0$ only depending on $\text{supp } \mu \cup \text{supp } \nu$ such that

$$c_L(\gamma_z(b - \varepsilon), (b + \varepsilon, \chi(b + \varepsilon)))^2 \geq \delta_1 \text{dist}(\dot{\chi}(b), Y_w)^2$$

where Y_w denotes the tangent to the unique \mathcal{A}_τ -minimizer $\eta_w: [b - \varepsilon, b] \rightarrow N$ whose graph connects $\gamma_z(b - \varepsilon)$ and w . Further since $c_L(\gamma_z(b - \varepsilon), \cdot)^2$ is Lipschitz up to the boundary of its domain, there exists $\delta_2 > 0$ with

$$\text{dist}((b + \varepsilon, \chi(b + \varepsilon)), \partial J^+(\gamma_z(b - \varepsilon))) \geq \delta_2 c_L^2(\gamma_z(b - \varepsilon), (b + \varepsilon, \chi(b + \varepsilon))).$$

By the triangle inequality and (12) one has

$$\begin{aligned} \text{dist}(\dot{\chi}(b), Y_w) &\geq \text{dist}(Y_w, \dot{\gamma}_x(b)) - C_2 \text{dist}(w, x) \\ &\geq \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b)) - C_3 \text{dist}(w, z) - C_2 \text{dist}(w, x) \end{aligned}$$

where the last inequality follows from the Lipschitz continuity of the vector field $u \mapsto Y_u$. Since $\text{dist}(w, x) \leq \text{dist}(z, x)$, one concludes

$$\begin{aligned} \text{dist}(\dot{\chi}(b), Y_w) &\geq \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b)) - (2C_3 + C_2) \text{dist}(z, x) \\ &\geq \frac{1}{2} \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b)). \end{aligned}$$

if $\text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b)) \geq 2(2C_3 + C_2) \text{dist}(z, x)$. Now the triangle inequality and the last estimate imply that

$$\begin{aligned} \text{dist}(\gamma_x(b + \varepsilon), \partial J^+(\gamma_z(b - \varepsilon))) &\geq \text{dist}((b + \varepsilon, \chi(b + \varepsilon)), \partial J^+(\gamma_z(b - \varepsilon))) \\ &\quad - \text{dist}(\gamma_x(b + \varepsilon), (b + \varepsilon, \chi(b + \varepsilon))) \\ &\geq \delta_3 \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b))^2 - \text{dist}(\gamma_x(b + \varepsilon), (b + \varepsilon, \chi(b + \varepsilon))) \end{aligned}$$

for some $\delta_3 > 0$. Next by the continuity of Φ_τ one has

$$\begin{aligned} \text{dist}(\gamma_x(b + \varepsilon), \partial J^+(\gamma_z(b - \varepsilon))) &\geq \delta_3 \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b))^2 - C_4 \text{dist}(w, x) \\ &\geq \delta_4 \text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b))^2 - C_4 C_1 \text{dist}(z, x)^2 \end{aligned}$$

for some $C_4 < \infty$. Now if

$$\text{dist}(\dot{\gamma}_z(b), \dot{\gamma}_x(b)) \geq \max\left\{2(2C_3 + C_2), \sqrt{\frac{C_4 C_1}{\delta_4}}\right\} \text{dist}(z, x),$$

one concludes $\gamma_x(b + \varepsilon) \in I^+(\gamma_z(b - \varepsilon))$. This in turn implies that the endpoint of γ_x in B is contained in $I^+(y_2)$, clearly a contradiction to the achronality of B . \square

Proof of Proposition 3.22. Since a countable union of neglectable sets is neglectable, one makes a few simplifying assumptions. One assumes that (i) A is uniformly spacelike, i.e. the distance of $TA \cap T^1M$ from \mathcal{C}^1 is bounded away from 0, (ii) B is precompact, and (iii) the distance between the first and the last intersections of \mathcal{A} -minimizers with B is uniformly bounded from below.

By Lemma 3.23 the set S_B is $\mathcal{L}_{\{b\} \times N}$ -neglectable. Further by Lemma 3.24 the map that assigns to each intersection point the tangent of the corresponding \mathcal{A}_τ -minimizer is Lipschitz. Choose a Lipschitz extension of this map to N according to Kirzbraun's theorem, cp. [20, Thm. 1.31]. Then the unique intersection of \mathcal{A}_τ -minimizers in I_B with A is the image of an $\mathcal{L}_{\{b\} \times N}$ -neglectable set under a Lipschitz map. Therefore it is \mathcal{L}_A -neglectable. \square

Proof of Theorem 2.13. First one shows that any optimal coupling is concentrated on the graph of a map. Any such map is measurable since couplings are Borel measures. Choose a dynamical optimal coupling Π . Denote by Π_Δ the restriction of Π to the set of constant \mathcal{A} -minimizers and $\Pi_C := \Pi - \Pi_\Delta$. Further set $\mu_\Delta := (\text{ev}_0)_\# \Pi_\Delta$ and $\mu_C := (\text{ev}_0)_\# \Pi_C$. Note that by construction one has $\mu = \mu_\Delta + \mu_C$.

First one shows that $\text{supp } \mu_\Delta \cap \text{supp } \mu_C$ is an \mathcal{L}_A -neglectable set. To this end note that $\text{supp } \mu_\Delta \subseteq \text{supp } \nu$ since μ_Δ is induced by constant curves. Now if $x \in \text{supp } \mu_\Delta \cap \text{supp } \mu_C$, then x is contained in $\text{supp } \nu$ and there exists $y \in \text{supp } \nu \cap J^+(x) \setminus \{x\}$. So x lies on an \mathcal{A} -minimizer that intersects the support of ν at least twice. The set consisting of such points was shown in Proposition 3.22 to be \mathcal{L}_A -neglectable which implies the initial claim.

Assume for the moment that $\pi_\Delta := (\text{ev}_0, \text{ev}_1)_\# \Pi_\Delta$ and $\pi_C := (\text{ev}_0, \text{ev}_1)_\# \Pi_C$ are separately concentrated on a graph. Then π is concentrated on the union of these graphs since $\pi = \pi_\Delta + \pi_C$. The overlap of these graphs lies in $\pi_1^{-1}(\text{supp } \mu_\Delta \cap \text{supp } \mu_C)$. Since $\text{supp } \mu_\Delta \cap \text{supp } \mu_C$ is \mathcal{L}_A -neglectable, it is also μ -neglectable and therefore $\pi_1^{-1}(\text{supp } \mu_\Delta \cap \text{supp } \mu_C)$ is π -neglectable. Thus π is concentrated on a graph.

Therefore one has to show that π_Δ and π_C are concentrated separately on a graph. This claim is trivial for π_Δ since π_Δ is concentrated on the diagonal of $M \times M$.

For π_C note that by construction $\pi_C(\Delta) = 0$. Since $\pi_C(\Delta) \geq \nu(\text{supp } \mu)$, one can apply Proposition 3.20 to the situation of μ_C and $\nu_C := (\text{ev}_1)_\# \Pi_C$ with the coupling π_C . Assume first that there exists a set $S \subset M$ with $\mu_C(S) > 0$ such that for every $x \in S$ there exist $y_1, y_2 \in \text{supp } \nu_C$ with $y_1 \neq y_2$, $(x, y_i) \in \text{supp } \pi_C$ and no \mathcal{A} -minimizer from x to y_i meets y_j for $i \neq j$. By the martingale property of π_C one has $\pi_C(\pi_1^{-1}(S)) = \mu_C(S) > 0$. Now for the set R constructed in Proposition 3.20 one has $R \cap \pi_1^{-1}(S) \neq \emptyset$. But this contradicts the property of R given in Proposition 3.20. Therefore the set of points transported into two different directions is μ_C -neglectable.

It remains to show that the set transported along one \mathcal{A} -minimizer, but to at least two points in B is \mathcal{L}_A -neglectable. But this is the content of Proposition 3.22 since μ_C is absolutely continuous with respect to \mathcal{L}_A . This follows directly from the assumption that μ is absolutely continuous with respect to \mathcal{L}_A .

Uniqueness of the optimal coupling follows from the observation that if two optimal couplings exist, any convex combination of both is optimal as well. But any nontrivial convex combination of two couplings, concentrated on separate graphs, is not concentrated on a graph unless they coincide. \square

The proof of Theorem 2.14 differs only in minor details from that of Theorem 2.13. These modifications are indicated in the following.

Proposition 3.25. *Let $\mu, \nu \in \mathcal{P}(M)$ be as in the assumptions of Theorem 2.14 and assume that $\text{supp } \mu$ is ν -neglectable. Then for every optimal coupling π of μ*

and ν there exists a set R of full π -measure such that for all $(x, y_1), (x, y_2) \in R$ there exists an \mathcal{A}_τ -minimizer γ containing x, y_1, y_2 in its trace.

Proof. Choose a dynamical optimal coupling Π between μ and ν . Like in Proposition 3.20 one can assume that \mathcal{A} -minimizers between $\text{supp } \mu$ and $\text{supp } \nu$ are nonconstant and unique up to parameterization. To see this, first consider dynamical subcouplings Ξ' instead of Π , according to Corollary 3.12, for the restriction of μ to $\overline{B_r(p)} \cap B_\varepsilon(\text{supp } \nu)^c$ for $p \in \text{supp } \mu$, $\varepsilon > 0$ and $r < \text{inj}(L)/2$, where $\text{inj}(L)$ denotes the injectivity radius of L on a sufficiently large compact subset of M . The proof continues to use the notation Π for the dynamical optimal coupling.

Since the distance between $\text{supp } \mu$ and $\text{supp } \nu$ is positive, Π -almost all \mathcal{A} -minimizers are nonconstant. Consequently, one can choose a measurable function $\sigma: \Gamma \rightarrow (0, 1]$ such that $\gamma(\sigma(\gamma)) \in (\text{supp } \mu)^c \cap B_{\text{inj}(L)}(\gamma(0))$ for Π -almost all $\gamma \in \Gamma$. The resulting restriction is optimal according to Corollary 3.11. By construction one knows that $\text{supp } \mu$ is $(\text{ev} \circ (\text{id} \times \sigma))_\# \Pi$ -neglectable.

Now one applies Lemma 3.21 to μ and ν to obtain the set $R \subset M \times M$. The remainder of the argument is absolutely analogous. \square

Recall that I_B denotes the set of \mathcal{A} -minimizers $\gamma: I \rightarrow M$ with $\tau \circ \gamma = \text{id}$ which intersect B in more than one point.

Proposition 3.26. *The set formed by the traces of orbits in I_B is \mathcal{L}_M -neglectable.*

Proof. As before one can assume that B is precompact and the distance between two intersections of an \mathcal{A} -minimizer with B is uniformly bounded from below. Then as above Lemmata 3.23 and 3.24 apply to the present case as well with the same notation. Choose a Lipschitz extension of the Lipschitz map obtained in Lemma 3.24 to M . Then the union of the traces of orbits in I_B is the image under the locally Lipschitz map of evaluation of an $\mathcal{L}_1 \times \mathcal{L}_{\{b\}} \times N$ -neglectable set. Therefore it is \mathcal{L}_M -neglectable. \square

The proof of Theorem 2.14 follows word-by-word the proof of Theorem 2.13 except for obvious changes.

Acknowledgements. The author would like to thank Victor Bangert for suggesting the problem of optimal transportation in the context of Lorentzian geometry and Albert Fathi for encouraging the pursuit of the project. The author would further like to thank Patrick Bernard for providing the opportunity to carry out the ideas for this article and making numerous suggestions which helped to shape the exposition of the present results. Valentine Roos and Rodolfo Ríos-Zertuche are kindly thanked for many helpful discussions in the process of this research.

REFERENCES

- [1] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, third edition, Springer, Berlin, 2006. MR2378491

- [2] L. Ambrosio and A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in *Optimal transportation and applications* (Martina Franca, 2001), 123–160, Lecture Notes in Math., 1813, Springer, Berlin, 2003. MR2006307
- [3] P. Bernard and B. Buffoni, Weak KAM pairs and Monge-Kantorovich duality, in *Asymptotic analysis and singularities—elliptic and parabolic PDEs and related problems*, 397–420, Adv. Stud. Pure Math., 47-2, Math. Soc. Japan, Tokyo, 2007. MR2387248
- [4] P. Bernard and B. Buffoni, Optimal mass transportation and Mather theory, *J. Eur. Math. Soc. (JEMS)* **9** (2007), no. 1, 85–121. MR2283105
- [5] P. Bernard and S. Suhr, Lyapounov Functions of closed cone fields: From Conley theory to time functions, *Comm. Math. Phys.* **359** (2018), no. 2, 467–498. MR3783554
- [6] J. Bertrand and M. Puel, The optimal mass transport problem for relativistic costs, *Calc. Var. Partial Differential Equations* **46** (2013), no. 1-2, 353–374. MR3016512
- [7] Y. Brenier, Extended Monge-Kantorovich theory, in *Optimal transportation and applications* (Martina Franca, 2001), 91–121, Lecture Notes in Math., 1813, Springer, Berlin, 2003. MR2006306
- [8] Y. Brenier, U. Frisch, M. Henon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskii, Reconstruction of the early universe as a convex optimization problem, *Mon. Not. Roy. Astron. Soc.* **346** (2003), 501–524.
- [9] T. Champion and L. De Pascale, The Monge problem for strictly convex norms in \mathbb{R}^d , *J. Eur. Math. Soc. (JEMS)* **12** (2010), no. 6, 1355–1369. MR2734345
- [10] M. Eckstein and T. Miller, Causal evolution of wave packets, *Phys. Rev. A* **95** (2017), 032106.
- [11] M. Eckstein and T. Miller, Causality for nonlocal phenomena, *Ann. Henri Poincaré* **18** (2017), no. 9, 3049–3096. MR3685983
- [12] A. Fathi and A. Figalli, Optimal transportation on non-compact manifolds, *Israel J. Math.* **175** (2010), 1–59. MR2607536
- [13] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, A reconstruction of the initial conditions of the Universe by optimal mass transportation, *Nature* **417** (2002), 260–262.
- [14] M. A. Javaloyes and M. Sánchez, Finsler metrics and relativistic spacetimes, *Int. J. Geom. Methods Mod. Phys.* **11** (2014), no. 9, 1460032, 15 pp. MR3270295
- [15] T. Miller, Polish spaces of causal curves, *J. Geom. Phys.* **116** (2017), 295–315. MR3623662
- [16] E. Minguzzi, Light cones in Finsler spacetime, *Comm. Math. Phys.* **334** (2015), no. 3, 1529–1551. MR3312442
- [17] E. Minguzzi, Convex neighborhoods for Lipschitz connections and sprays, *Monatsh. Math.* **177** (2015), no. 4, 569–625. MR3371365
- [18] E. Minguzzi, An equivalence of Finslerian relativistic theories, *Rep. Math. Phys.* **77** (2016), no. 1, 45–55. MR3461797
- [19] B. O’Neill, *Semi-Riemannian geometry*, Pure Appl. Math., 103, Academic Press, Inc., New York, 1983. MR0719023
- [20] J. T. Schwartz, *Nonlinear functional analysis*, Gordon and Breach Science Publishers, New York, 1969. MR0433481
- [21] C. Villani, *Optimal transport*, Grundlehren Math. Wiss., 338, Springer, Berlin, 2009. MR2459454

Received October 12, 2016; accepted March 25, 2018

Stefan Suhr
 Fakultät für Mathematik,
 Ruhr-Universität Bochum,
 Universitätsstraße 150, 44780 Bochum, Germany
 E-mail: stefan.suhr@rub.de