Optimal Transport and General Relativity

Joaquín Sánchez

2020

Why?

General relativity has been around for quite a bit and optimal transport is mathematical theory of fast development in the last couple of decades. Very recently a couple of papers [6] and [8] have been addressing the question of doing optimal transport in spacetime. The very first question the reader must be thinking is **why?**.

- Why are we suddenly looking at general relativity through the eyes of optimal transportation theory?
- Why do we expect this technique to have any success?
- What kind of things do we expect to obtain from this approach?
- What would be the goal?

Here's why

- There have been recent developments [3], [5], [7] on understanding the underlying geometry of a manifold through optimal transport. Optimal transport has shown successful to describe geometry from a coordinate-free analytic perspective. Further, these developments have generalized the concept of Ricci curvature to (non-smooth) metric measure spaces. Einstein's equation determines the geometry of spacetime through the Ricci tensor, so if we can understand the Ricci tensor and more general versions of it, we can get insights on connections between physical theories that could have been overlooked by the geometric tools.
- Optimal transport has been successful in generalizing concepts for Riemannian manifolds and describing hidden connections, one expects the same to happen in Lorentzian manifolds.

- There is more than one way to look at a physics law, if we can describe them in another way, maybe they will hint connections. We expect optimal transport to give us reinterpretations of physical laws in general relativity that could enlighten new ideas.
- Understand some ideas from general relativity through the eyes of optimal mass transportation. The second law of thermodynamics can be understood in terms of optimal transportation, what else can be understood in this way?

In [4] they summarize the general idea of casualty between probability measures as follows:

Each infinitesimal part of the probability distribution should travel along a future-directed causal curve.

So the idea is to develop a mathematical model of optimal transport in Lorentzian manifolds.

Key Idea

The second law of thermodynamics states that the total entropy of an isolated system never decreases, this can be interpreted as the gasses having a preferred direction in time (the direction is the one minimizing entropy as time increases). Optimal transport has shown to be efficient to encompass that idea. General relativity deals with a preferred direction in time ($+$ or $-$ depending the convention), so optimal transport could help to reformulate the theory.

Structure of this report

Since we want to develop the ideas of optimal transport theory, I will have 2 different types of objects through this report:

1. *Usual* Optimal Transport objects (denoted U.O.T. at the end)
2. Their general relativity counterparts.

So I will write:

Theorem U.O.T. *To describe a theorem in usual optimal transport*

and without the term U.O.T. to describe the new relativistic setting.

In this report, the idea is to understand the main theorems and proofs on [6]

Contents

1 Introduction to optimal transport

1.1 What is Optimal Transport?

In this section I make a small survey through some topics of optimal transportation, this is done for the unfamiliar reader to get some sense on the topic before we deal into details. **Details** and more specfic formulations **will be addressed in subsequent sections**. This is a short review of some relevant ideas.

In 1781, french mathematician Gaspard Monge raised a problem that would intrigue the world of mathematics:

Given two subsets $U, V \subseteq \mathbb{R}^3$ with the same volume, find a volume preserving map between them, such that it minimizes some cost fixed function $c(x, y)$.

Little did everyone know this 'apparently easy' formulation would lead to centuries of work for many great scientists from a wide variety of backgrounds.

1.2 Monge's Problem

The before mentioned problem can be formulated in terms of measure spaces:
For measure spaces, (X, S_X, μ) and (Y, S_Y, ν) we aim to find:

$$\inf_{T_{\#}\mu=\nu} \left\{ \int_X c(x, Tx) d\mu \right\},$$

Where $T_{\#}\mu = \nu$ denotes that T is in fact what we will call a transport function: a *push forward* from μ to ν. This means that for any set A in S_Y, one has

$$T : X \to Y \text{ y } \mu(T^{-1}(A)) = \nu(A).$$

By this last equation one can use the change of variable formula for measures to formulate the problem in the space Y in terms of the inverse transport function T^{-1}, but in general this makes no difference.
The most interesting part of such problem is that even though it's simple formulation it turned out not to be so easy to solve. Almost 200 years passed without significant breakthroughs to Monge problem until V.N. Sudakov was able to find and prove existence conditions for the mapping T. **Why?** Mainly because the condition defining T is highly non-linear. Meaning that tools from linear analysis would not provide solutions.

2 Kantorovich's Problem

The real breakthrough came in the late twentieth century, when Leonid Kantorovich realized that he could formulate a different problem, easier to analyze but very related to the one of Monge.

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \left\{ \int_{X\times Y} c(x, y) d\gamma \right\}.$$

Where $\Gamma(\mu, \nu) = \{\gamma \in M_1(X \times Y, S) : \pi_x(\gamma) = \mu, \pi_y(\gamma) = \nu\}$.
Here, M_1 denotes the set of all probability measures in the product space $X \times Y$ and π_x denotes the projection of the measure in the first coordinate.

How are these problems related? It turns out that Kantorovich's problem is a relaxation of Monge's.

Theorem U.O.T. *Kantorovich's problem is a true relaxation of Monge's.*

Being **a true relaxation of a problem** means that whenever you have an element in the first problem, you get an element in the second problem; therefore the infimum on the second one must be smaller.

Proof. Let T such that $T_{\#}\mu = \nu$ for every set A in the product σ-algebra define

$$\gamma_T(A) = \mu(\{x \in X : (x, Tx) \in A\})$$

Then $\gamma_T \in \Gamma(\mu, \nu)$.

2.1 Convergence of measures and topologies in the space of measures

The breakthrough by Kantorovich relies on being able to put Monge's in a more manageable setting; namely, the space of probability measures on the product space. By this observation it is now obvious that we need to use the fundamental properties of the space of probability measures.
Now the problem has changed into analyzing how measures interact. We need to understand how measures interact with each other.

Theorem U.O.T. *The set of probability measures inherits properties like completeness and compactness from the underlying metric space.*

This is a very well known result and by using the **variation norm** allows us to use Riesz's theorem to identify the dual space of the continuous functions of compact support with the set of positive Radon measures, which is a necessary condition to understand Kantorovich's Duality formula (next section).
While the second distance is highly related to iso-perimetric type inequalities which is of course of great interest to us. So for a metric space (X, d), $p \in [1, \infty)$and two probability measures μ, ν on X with finite order p we define the Wasserstein distance between μ and ν:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \left\{ \int_{X \times X} d(x, y)^p d\gamma \right\} \right)^{\frac{1}{p}}.$$

It is evident how this definition arises in the context of optimal transport, but the next result may appear utterly surprising:

Theorem U.O.T. *The Wasserstein distance metrizes weak convergence of probability measures.*

2.2 Kantorovich's Duality formula

Why is the Kantorovich problem understood to be 'easier' in some way to the one by Monge? Mainly because it was nice enough to give conditions for solutions.

Theorem U.O.T. *Let X and Y compact spaces.*
Let μ be a probability measure on X and ν a probability measure on Y.
Let $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ be non-negative.
Let $\Gamma(\mu, \nu)$ be the set of Borel measures on $X \times Y$ such that their projections on X and Y are μ and ν respectively
Let Φ_c be the set of measurable functions $(\phi, \psi) \in L_1(\mu) \times L_1(\nu)$ such that

$$\phi(x) + \psi(y) \leq c(x, y)$$

μ-a.e. in x y ν-a.e. in y.

Then

$$\inf_{\pi\in\Gamma(\mu,\nu)}\left\{\int c(x,y)d\pi\right\} = \sup_{(\phi,\psi)\in\Phi_c}\left\{\int \phi d\mu + \int \psi d\nu\right\}$$

Moreover, Kantorovich's problem admits a minimizer

2.3 Real case and the quadratic cost function

In the case were the space $X = \mathbb{R}^n$, and the cost function: $c(x,y) = \dfrac{||x-y||^2}{2}$.
Recall the definition of the **sub-differential** $\partial\phi$:

$$y \in \partial\phi(x) \Leftrightarrow \forall z \in \mathbb{R}^n \phi(z) \geq \phi(x) + \langle y, z-x\rangle$$

Theorem U.O.T. *Optimal plans are supported in subdifferentials Let μ and ν be probability measures and $\pi \in \Gamma(\mu,\nu)$, then if π is optimal for c it must be supported in the subdifferential of a proper lower semi-continuous convex function.*

Theorem U.O.T. *Moreover if μ doesn't give mass to small sets (say a.c. with respect to Lebesgue measure) then there **exists** a convex function ϕ on $\mathbb{R}^n$ such that*

$$\nabla\phi_{\#}\mu = \nu$$

This result and the fact that Kantorovich's Duality formula is proved by means of convex analysis advices us of the real need of understanding forms of convexity to solve transport problems.

2.4 Time dependent formulation, displacement interpolation and displacement of convexity

We have defined the transport problem in a time independent setting. To analyze isoperimetric type inequalities we will need to reformulate the problem, so now for each x we obtain a trajectory $T_t(x)$ where t varies from 0 to 1 and of course it has associated a cost of displacement $C(T_t(x))$.
We need regularity in the trajectories so it is assumed that $t \to T_t(x)$ is continuous and piecewise C^1 for μ a.e. on x. So naturally the time dependent problem is formulated:

$$\inf\left\{\int_X C(T_t(x))d\mu; T_0 = Id, T_{1\#}\mu = \nu\right\}$$

, where $Id(x) = x$ is just the identity operator. For probability measures μ and ν on $\mathbb{R}^n$ that do not give mass to small sets, the collection of measures given by

$$[\mu,\nu]_t = [tId + (1-t)\nabla\phi]_{\#}\mu$$

is called the displacement interpolation of μ into ν, it can be seen as the most natural linear interpolation (the geodesic).
A subset of the absolutely continuous (w.r.t Lebesgue) measures on $\mathbb{R}^n$ is said to be **displament convex** if whenever μ and ν are in the set $[\mu,\nu]_t$ is a.c. and lies in the set for all $t \in [0.1]$.
Meaning $\mathbf{S} \subseteq P_{ac}(\mathbb{R}^n)$ is displacement convex iff $\mu, \nu \in S \Rightarrow [\mu,\nu]_t \in \mathbf{S}\ \forall t \in [0,1]$ Similarly a functional F defined on a displacement convex set is said to be displacement convex if whenever $\rho_t = [\mu,\nu]_t$ is a displacement interpolation, the function $t \to F(\rho_t)$ is convex.

Theorem U.O.T. *Criteria for displacement convexity Let* $\mathbf{S}$ *be a displacement convex subset of* $P_{ac}(\mathbb{R}^n)$*. Let* $U : \mathbf{S} \to \mathbb{R} \cup \{\infty\}$ *then if*

$$U(0) = 0 \text{ ***and*** } r \to r^n U(r^{-n})$$

is convex and non-increasing for positive r *then* U *is displacement convex on* $\mathbf{S}$.

2.5 Conclusions and key ideas from optimal transport

We have two measures, μ and ν and we want to transport μ onto ν. We want to find a function T such that $T_{\#}\mu = \nu$ and minimizes the total cost.

- There might not be such T transporting μ onto ν. (We cannot split mass, so our probabilities must be nice and not give mass to small sets)
- If the cost function c does not satisfy nice properties, there is no hope for optimizers. (We require c to be convex and lower semi-continuous so that it is indeed a cost function)
- There could be more than one optimizer
- The (infinite) linear problem posed by Kantorovich's formula is much easier to solve
- The interpolation on time will have a linear structure between transportation maps but not between measures
- Optimal transference plans are supported in subgradients of convex functions because such sets preserve "cyclical-monotoniciy", a condition necessary for optimality that states that if we permute the points and use the same measure, we would increase the total cost of transport.

Analyzing the structure of the space of probabilities space $\mathcal{P}(M)$ will lead to conclusions on M. This is the main idea behind optimal transport.

3 Framework for optimal transport in general relativity

We say (M^n, g) is a spacetime if it is a smooth, connected, Hausdorff, time-oriented Lorentzian manifold with signature $(+, -, -, \cdots, -)$.
Here the signature convention has been done opposite to the one we used throughout the lectures.
So we will call a tangent vector $v \in T_x M$ timelike if $v^a g_{ab} v^b > 0$, spacelike if $v^a g_{ab} v^b < 0$ and null in the remaining case. We say that a vector is causal if it non-space-like. The time orientation determines if the vector is future or past-directed according to the time-orientation of the manifold.
In order to work our way to optimal transport, we have to set our minds on what it is that we want to minimize. So we must determine our cost function. As shown in the introduction, a very natural cost function in the Euclidean case is the quadratic distance, nevertheless now we must recognize two fundamental ideas:

- Not all points are accesible from each other.
- Nature follows the principle of least action (minimizing Lagrangians)

Definition 3.1. (Convex Lagrangian and action of a curve)
For $0 < q \leq 1$, we define the q-Lagrangian on the tangent bundle of M is given by

$$L(v, x; q) = \begin{cases} -\dfrac{(v^a g_{ab}(x) v^b)^{q/2}}{q} & \text{if } v \text{ is future-directed} \\ \infty & \text{otherwise} \end{cases}$$

For a curve $\sigma \in \mathcal{C}^1([0,1], M)$ continuous curves mapping $[0,1]$ to the manifold M, we define it's **action**:

$$A(\sigma, q) = \int_0^1 L(\sigma'(s), \sigma(s); q) ds$$

So as the convex Lagrangian will be involved in a process of minimization, only future-directed curvves will be relevant. Now we formulate the formal version of a principle of least action.

Definition 3.2. (q-distance) We define the q-distance between two points on the manifold x, y as the total minimal action required from going to y from x along a future-directed curve:

$$\ell(x, y; q) = -\inf\{A(\sigma, q) : \sigma \in \mathcal{C}^1, \sigma(0) = x, \sigma(1) = y\}$$

And define also $\ell(x,y) = \ell(x,y;1)$, ℓ is called the time-separation function.
If $q < 1$ and σ achieves the infimum we call the curve an affinely-parametrized proper-time maximizing geodesic segment.

Observation 3.1. Another way to define $\ell(x,y)$ is to notice that the quantity $(q\ell(x,y;q))^{1/q}$ does not depend on q.
This fact is surprising because at first look $\ell(x,y;q)$ resembles a q-norm, nevertheless when we take the infimum over all such curves, I still do not have a good argument for it.

Note that the lagrangian L is infinite only if y does not belong to the causal future of x. So we get a characterization of causality in terms of ℓ.

$$\ell(x,y) \geq 0 \text{ if and only if } y \text{ lies in the causal future of } x$$

And similarly without the equality

$$\ell(x,y) \geq 0 \text{ if and only if } y \text{ lies in the chronological future of } x$$

One can take this as definitions and continue from there.
Note that one can easily check that $\ell(x,y) \geq \ell(x,z) + \ell(y,z)$ by just concatenating the curve joining x and z and the one joining z and y.

Definition 3.3. (Causal future and causal past) The set $J^+(x) = \{z : \ell(x,z) \geq 0\}$ is called the causal future of x.

The set $J^+(y) = \{z : \ell(z,y) \geq 0\}$ is called the causal past of y.

Definition 3.4. (Global Hyperbolicity) A spacetime (M^n, g) is said to be globally hyperbolic if

- There are no closed causal curves
- $J^+(x) \cap J^-(y)$ is compact.

Assumption 3.1. (Global hyperbolicity) From now on we will assume (M^n, g) is globally hyperbolic, this is to ensure that we can achieve minimizers in the optimal transport problem.

Let $\mathcal{P}(M)$ be the set of Borel-probability measures on M.
Let $\mathcal{P}_c(M) = \{\mu \in \mathcal{P}(M) : supp(\mu) \text{ is compact}\}$ where $supp(\mu) = \overline{\{x \in M : \mu(N_x) > 0\}}$ where N_x is any open neighborhood of x.
In optimal transport, most properties of the underlying metric space (X,d) are inherited by the probabilities space if we give it the correct norm:

Definition 3.5. (Wasserstein distance) In a metric space (X,d) we define de p-th Wasserstein distance between μ and ν as the minimal cost induced by the Kantorovich problem, namely:

$$W_p(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y)^p d\pi(x,y) \right)^{1/p}$$

where again $\Pi(\mu,\nu)$ is the set of probability measures on $X \times X$ having as projections into coordinates μ and ν, i.e. $\pi(A\times X) = \mu(A)$ and $\pi(X \times B) = \nu(B)$. Such probabilities are also called couplings between μ and ν

Definition 3.6. (q-Lorentz (Wasserstein) distance in GR) If (M^n, g) is a globally hyperbolic spacetime as before, the q-Lorentz distance between $\mu,\nu \in \mathcal{P}(M)$ is defined as

$$\ell_q(\mu,\nu) = \left(\sup_{\pi\in\Pi_*(\mu,\nu)} \int_{M\times M} \ell(x,y)^q d\pi(x,y) \right)^{1/q}$$

where Π_* is the set of all probability measures on $M \times M$ having as projections into coordinates μ and ν **with** $supp(\pi) \subseteq \ell^{-1}[0,\infty]$

Observation 3.2. Note how the added restriction on the support of π enforces that the only points on which we can have positive measure are the ones on which we can travel. The condition

$$supp(\pi) \subseteq \ell^{-1}[0,\infty]]$$

is the only real diference between the definitions, the sup and inf are only different because of the sign convention.

One interprets this quantity as follows: We look at the events that μ represents (those that are plausible under μ) and we travel to the events represented by ν. ℓ_1 is the maximum expected time it would take us.
We want to show that this quantity is in fact like a distance, and hence we **will** later prove that it satisfies a reverse triangle inequality in section 4.3:

$$\ell_q(\mu,\nu) \geq \ell_q(\mu,\gamma) + \ell_q(\gamma,\nu)$$

Note that to avoid indeterminations, we settle the convention: $\infty - \infty = -\infty$

4 The idea of q-geodesics

We want to give $\mathcal{P}(M)$ a good structure, we already know the concept of geodesic in M, so we must have an idea of the concept in $\mathcal{P}(M)$, for this goal we extend the idea of straight lines to $\mathcal{P}(M)$.

Definition 4.1. (q-Geodesics on $\mathcal{P}(M)$) A collection of probability measures $\{\mu_t\}_{t\in[0,1]}$ is called a q-geodesic if whenever $t \geq s$,

$$\ell_q(\mu_s,\mu_t) = (t-s)\ell_q(\mu_0,\mu_1)$$

This is interpreted as the probability measures being in the correct path, minimizing up to every time as in a dynamical programming principle.
We know aim to understand q-geodesics. Do they exist? Are they unique? Do they inherit properties from endpoints?

One of the most important concepts in Riemannian geometry is the cut locus, the set of points for which we stop having a unique minimizing geodesic, the set where the exponential function stops being injective. In some sense, if we want to use either the exponential function or minimizing geodesiccs we must stay out of the cut locus, let us define a similar concept.

Definition 4.2. (Singular set of ℓ)
In our globally hyperbolic spacetime (M^n, g) we say $(x, y) \in sing(\ell)$ unless $\ell(x, y) > 0$ and x, y lie in the relative interior of some affinely parametrized proper-time maximizing geodesic segment. (See definition 3.2)

Intuitively, the only way for (x, y) not to be in the singular set is if we can travel from x to y and also there is a proper-time optimal way to do it.
Recall that a function f is called lower semicontinous if

$$\liminf_{x \to x_0} f(x) \geq f(x_0)$$

and it is called upper semi-continuous if

$$\limsup_{x \to x_0} f(x) \leq f(x_0)$$

It is clear form the definition of the q-Lorentz distance (3.6) that our cost function for optimal transport is going to be $\ell(x, y)$. Hence, as addressed in the introduction, we must know continuity/semi-continuity/convexity/regularity properties of ℓ to even hope to have optimal transportation.

4.1 Properties of ℓ

Proposition 4.1. *(Continuity and regularity properties of ℓ) In our globally hyperbolic spacetime (M^n, g)*

- *ℓ is upper-semicontinuous of $M \times M$*
- *ℓ is continuous on $\ell^{-1}[0, \infty]$*
- *ℓ is smooth outside of $sing(l)$*

Sketch of proof. Note that the first two statements follow directly if we know continuity of $d : (x, y) \to \max\{\ell(x, y), 0\}$. Let us give a sketch, following [2], assume by contradiction that it is not upper-semicontinuous at (p, q), then there exists a sequence (p_n, q_n) such that

$p_n \to p$ and $q_n \to q$ but the distance between p_n and q_n is always ϵ bigger than the distance between p and q, for some fixed $\epsilon > 0$. By definition of d, there exists a future-directed curve σ_n such that $L(\sigma_n', \sigma_n, 1) \geq d(p_n, q_n) + \epsilon/2$. But $\{\sigma_n\}$ can be shown to have a limiting future directed curve σ which will join p and q, computing the value of $L(\sigma', \sigma, 1)$ gives the contradiction.
The last point of the proposition is more difficult and will be discussed later.

□

Proposition 4.2. *Unique optimal midpoints outside the singular set.*
Let $(x, y) \in M \times M \setminus sing(\ell)$, *then for* $0 \leq s \leq 1$ *there exists a unique* $z_s \in M$ *such that*

$$\ell(x, z_s) = s\ell(x, y),\ \ell(z_s, y) = (1 - s)\ell(x, y)$$

Further, z_s *depends smoothly on the triple* (x, y, s)

Proof. If (x, y) are not in the singular set, there is a proper-time minimizing geodesic segment σ, such that x and y lie in the relative interior of σ and y is in the chronological future of x ($\ell(x, y) > 0$). Let $z_s(x, y)$ be the unique proper-time parametrized action-minimizing geodesic joining x and y, so $z_s(x, y) = \exp_x(sv)$ for some $x \in T_xM$. The fact that z_s is a segment of a "geodesic" means it will satisfy the desired equation while the smoothness follows from the fact that outside the singular set the exponential acts diffeomorphically. □

So far we obtained a function $z_s(x, y)$ that is smooth and unique as long as $(x, y) \notin sing(l)$, we would like to extend this definition to all the places relevant to us, namely to the set $\{\ell > 0\}$. The problem is that in the singular set we have many options as there is no uniqueness of proper-time action-minimizing geodesics, meaning that the exponential function fails to be injective.

Fact 4.1. *There is an extension of* z_s *from* $M \times M \setminus sing(\ell)$ *to* $\{\ell > 0\}$ *that remains measurable.*

I call this a fact so I can avoid the proof as invoking the Kuratowski-Ryll-Nardzewski measurable selection theorem seems a little strong to not write and explain how it works. It is actually not the most difficult subject and a proof can be found in Bogachev's classical book on measure theory [1] but it is lengthy and distracting from the point of this report.

Definition 4.3. (Midpoint sets)
For our globally hyperbolic spacetime (M^n, g), given a set $S \subseteq M \times M$ and $s \in [0, 1]$, if $\ell(x, y) > 0$ let

$$Z_s(x, y) = \{z \in M : \ell(x, z) = s\ell(x, y), \ell(z, y) = (1 - s)\ell(x, y)\}$$

$$Z_s(S) = \bigcup_{(x,y)\in S} Z_s(x, y)$$

$$Z(S) = \bigcup_{s\in[0,1]} Z_s(S)$$

That is $Z_s(x, y)$ are all the midpoints as in the last propositio between x and y while $Z_s(S)$ are all the midpoints of a set, finally $Z(S)$ are all points in the way from x to y.
In any other case let us define the midpoint sets to be empty.

Fact 4.2. *If S is precompact (it's closure is compact) so is $Z(S)$.*
If S is compact so are $Z_s(S)$ and $Z(S)$

Recall also the concept of gluing of measures: If we have μ_1, μ_2, μ_3 measures in M a gluing π is a measure on $M \times M \times M$ such that the projection into the coordinates i, j is a coupling of (μ_i, μ_j). Gluing is the opposite concept to the desintegration theorem.
Disintegration is a more general version of Fubini's theorem, it can alse be thought of as Bayes theorem. The idea is to "break" a probability measure into 2, and integrate first with respect to one variable and after with respect to another.

Theorem U.O.T. *(Disintegration Theorem)*
Let (Ω_1, μ_1) and (Ω_2, μ_2) be probability spaces. Assume that Ω_2 is a complete, separable metric space. Suppose we have a function $g : \Omega_2 \to \Omega_1$ and suppose that $g_\# \mu_1 = \mu_2$. Let $\mathbb{P} := (g, Id)_\#$, then there exists a collection of probability measures $\mathbb{P}_x$ such that for any bounded measurable functions $h : \Omega_1 \to \mathbb{R}, f : \Omega_2 \to \mathbb{R}$

$$\int f(y)h(g(y))d\mu_2(y) = \int \left(\int f(y)d\mathbb{P}_x(y) \right) h(x)d\mu_1(x)$$

The idea of the desintegration theorem is that we can break integration with respect to μ_2 on integrating first with respect to $\mathbb{P}_x$ and then $d\mu(x)$. A reader familiar with probability or statistics recognizes $\mathbb{P}_x$ as the regular conditional probability given x.
We want to look at the probability of an event involving (x, y), we can break it into looking at x fixed and varying y and then summing over all possible values of x. Now that we have some knowledge on the midpoints we can prove the triangle inequality mentioned in the introduction.

Proposition 4.3. *(ℓ_q reverse triangle inequality)*
If $0 < q < 1$, let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(M)$ with $\ell_q(\mu_2, \mu_3) \neq -\infty \neq \ell_q(\mu_1, \mu_2)$ then

$$\ell_q(\mu_1, \mu_3) \geq \ell_q(\mu_1, \mu_2) + \ell_q(\mu_2, \mu_3)$$

Proposition 4.4. *(Equality cases) In the context of the above proposition:*

- $\mu_1(X_1) = 1 = \mu_3(X_3), \mu_2(Z(X_1 \times X_3) \neq 1$, *and the right hand-side is finte imply that the inequality is strict.*

- *if $\ell_q(\mu_1, \mu_3)$ is finite and equality holds, there is a q.optimal gluing $\pi \in \mathcal{P}(M \times M \times M$ such that every (x, y, z) in supp(π) is an s-midpoint:*

$$\ell(x, z) = s\ell(x, y), \ell(z, y) = (1 - s)\ell(x, y)$$

  for $s = \ell_q(\mu_1, \mu_2)/\ell_q(\mu_1, \mu_3)$.

Proof. Assume that the terms on the right hand-side are not $-\infty$ because if so, the inequality is trivial. Given $\epsilon > 0$ by definition of ℓ_q as the supremum of total costs, there exist $\pi_{1,2} \in \pi_*(\mu_1, \mu_2)$ and $\pi_{2,3} \in \pi_*(\mu_2, \mu_3)$ that are nearly optimal in the sense that

$$\left(\int_{M\times M} \ell(x, y)^q d\pi_{1,2}\right)^{1/q} \geq \min\{\ell_q(\mu_1, \mu_2) - \epsilon, 1/\epsilon\}$$

$$\left(\int_{M\times M} \ell(x, y)^q d\pi_{2,3}\right)^{1/q} \geq \min\{\ell_q(\mu_2, \mu_3) - \epsilon, 1/\epsilon\}$$

Note that $\ell_q(\mu_i, \mu_j)$ could potentially be very large, that's why we restrict to being bigger than the minimum with$1/\epsilon$. Note that this condition doesn't make a difference when $\ell_q(\mu_i, \mu_j)$ is finite because there exists ϵ such that $1/\epsilon$ surpasses that value and we keep the definition of the supremum as always. By the desintegration theorem, let us write:

$$d\pi_{1,2}(x, y) = (d\pi^y_{1,2}(x))d\mu_2(y), d\pi_{2,3}(x, y) = (d\pi^y_{2,3}(x))d\mu_2(y)$$

Meaning that we desintegrate both with respect to μ_2, and I have written a parenthesis just evoke the fact that we integrate inside with respect to x and then with respect to y, that's why y is a paremeter in $\pi^y_{1,2}$.
We proceed to glue, recall that by duality, it is enough for defining a measure to define is integration against continuous bounded functions, so let us define $\pi \in \mathcal{P}(M \times M\times)$ by

$$\int_{M\times M\times M} \phi(x, y, z)d\pi(x, y, z) := \int_M \left(\int_{M\times M} \phi(x, y, z)d\pi^y_{1,2}(x)d\pi^y_{2,3}(z)\right) d\mu_2(y)$$

Note that by definition of π we obtain that it's projection onto $(1, 3)$ is a coupling of μ_1, μ_3 that we call $\pi_{1,3}$
As $\pi_{1,3}$ is a coupling of μ_1, μ_3, by definition of ℓ_q we have:

$$\ell_q(\mu_1, \mu_3) \geq \left(\int_{M\times M} \ell(x, z)^q d\pi_{1,3}(x, z)\right)^{1/q} = \left(\int_{M\times M} \ell(x, z)^q d\pi(x, y, z)\right)^{1/q}$$

Where in the equality we replaced $\pi_{2,3}$ by π because the function does not depend on the middle variable y and μ_2 is a probability measure.
We know apply the reverse triangle inequality for ℓ, so we get

$$\ell_q(\mu_1,\mu_3) \geq \left(\int_{M\times M} (\ell(x,y)+\ell(y,z))^q d\pi(x,y,z)\right)^{1/q}$$

But because $0 < q < 1$ we can use Minkowski's inequality (reversed)

$$\ell_q(\mu_1,\mu_3) \geq \left(\int_{M\times M} \ell(x,y)^q d\pi(x,y,z)\right)^{1/q} + \left(\int_{M\times M} \ell(y,z))^q d\pi(x,y,z)\right)^{1/q}$$

But because of how we glued, as the first integral does not depend on z and the second one on x we have:

$$\ell_q(\mu_1,\mu_3) \geq \left(\int_{M\times M} \ell(x,y)^q d\pi_{1,2}\right)^{1/q} + \left(\int_{M\times M} \ell(x,y)^q d\pi_{2,3}\right)^{1/q}$$

and finally we get:

$$\ell_q(\mu_1,\mu_3) \geq \min\{\ell_q(\mu_1,\mu_2) + \ell_q(\mu_2,\mu_3) - 2\epsilon, 1/\epsilon - \epsilon\}$$

Were the last term is obtained by simple cases:

- If both lower bounds in the hypothesis are obtained in the left side of the minimum, the bound is trivial:

$$\ell_q(\mu_1,\mu_2) - \epsilon + \ell_q(\mu_2,\mu_3) - \epsilon = \ell_q(\mu_1,\mu_2) + \ell_q(\mu_2,\mu_3) - 2\epsilon$$

- If one of the minimums is atained in the right handside and the other one in the left, w.l.o.g. for the one on the left let us μ_1, μ_2, then the bound is

$$\ell_q(\mu_1,\mu_2) - \epsilon + 1/\epsilon, \geq 1/\epsilon - \epsilon$$

  since $\ell_q(\mu_1,\mu_2) > 0$ as it is not $-\infty$.

- Both lower bounds are attained in the right handside term, our lower bound is: $2/\epsilon$ which is bigger than $1/\epsilon - \epsilon$ as long as $\epsilon < 1$.

Finishing the proof of the first part. The other parts follow from the structure of middle points sets $Z_s(S)$ which are able to mantain compactness and precompactness. But the idea is that to get equality we must have equality in all steps and that is the definition of $Z_s(S)$ □

4.2 Existence of q-geodesics

We now address the concept of existence of q-geodesics as defined in definition 4.1.

Theorem 4.1. *(Interpolation theorem) Let $\mu, \nu \in \mathcal{P}(M)$ and suppose that ℓ_q is attained by some $\pi \in \Pi(\mu, \nu)$ such that $\ell > 0$ a.e. with respect to π.*
The collection of probabilities given by $\mu_s = \overline{z}_{s\#}\pi$ defines a q-geodesic where $\overline{z}_s$ is the measurable extension of z_s given by fact 4.1
$(\overline{z}_s \times \overline{z}_t)_{\#}\pi$ is ℓ_q-optimal.
Uniqueness of the geodesic follows if π is the unique probability achieving $\ell_q(\mu, \nu)$ and μ, ν are compactly supported with $\pi(sing(\ell)) = 0$.

Proof. If π attains $\ell_q(\mu, \nu)$, we use $\overline{z}_s$ to define μ_s as in the hypothesis of the theorem, then by definition of ℓ_q and definition of the push-forward:

$$\ell_q(\mu_s, \mu_t)^q \geq \int \ell(\overline{z}_s(x, y), \overline{z}_t(x, y))^q d\pi(x, y) = (t - s)^q \int \ell(x, y)^q d\pi(x, y)$$

Where the second equality is just by definition of $\overline{z}_s$ (recall Proposition 4.2). But the last term is the value of ℓ_q by optimality of π
Now we use this lower bounds for $(s, t) = (0, t_1), (t_1, t_2), (t_2, 1)$ and the fact that $0 < q < 1$ and we get

$$\ell_q(\mu_0, \mu_{t_1}) + \ell_q(\mu_{t_1}, \mu_{t_2}) + \ell_q(\mu_{t_2}, \mu_1) \geq \ell_q(\mu_0, \mu_1)$$

But the reverse triangle inequality gives the other inequality.

□

Exploring the properties for the Lorentz distance l, one of the main results of [6] is the what we wrote in the last part of Proposition 4.1:

Definition 4.4. (Semi-convexity) Let $\overline{g}$ be a Riemannian metric on M, we say that a function $F : U \to M$, where $U \subseteq M$ is open, is semi-convex if there exists a constant C such that

$$\liminf_{v \to 0} \frac{F(\exp_x^{\overline{g}}(v)) + F(\exp_x^{\overline{g}}(-v)) - 2F(x)}{|v|_{\overline{g}}^2} \geq C$$

for every $x \in U$.

Recall the definition of convexity in $\mathbb{R}$, we compare the value of a function at a linearly interpolated point to the linear interpolation of the values. In Riemannian manifolds we must not use linear interpolations but we use the exponential function. We look at small displacements on both directions along geodesics and compare those to the value of the function while dividing by the size of the displacement. It looks like a two-sided Frechet-derivative along geodesics.

Theorem 4.2. *(Smoothness of ℓ)*
The Lorentz-distance ℓ is smooth on the complement of the closed set $sing(\ell)$ and locally Lipschitz and locally semi-convex on $\{\ell > 0\}$.
However, the superdifferential of $\ell(\cdot, y)$ is empty at x if $\ell(x, y) = 0$ if $x \neq y$.
If $x = y$ the supergradients lie in $\{p \in T^*_x M | H(p; 1) = 0\}$ where H is the Hamiltonian, defined similarly to the Lagrangian but for past-directed curves instead.

This result is fundamental to be able to talk about optimal transport functions in the Monge problem and to try to obtain regularity or conditions on the Monge-Ampere equation. The caveat is that there is no remedy for the singular set, in fact one can show that for $(x, y) \in \{l > 0\} \cap sing(\ell)$ semiconvexity fails. (Theorem 3.5 in [6]). Making optimal transport problems in the cut locus completely infeasible.

Idea. The proof is quite technical but the main idea can be summarized as follows: Take the Lorentzian parallel transport and define a variation of an action minimizinf geodesic as the exponential of the Lorentzian parallel transport. Meaning that we do parallel transport and then flow a little in a geodesic fashion, one compares the action of this new curve to the original using Synge's second variational formula. For details check [6] □

As said before, the idea is to lift the structure of ℓ to ℓ_q, it turns out a lot of properties can be lifted from ℓ to ℓ_q.

Corollary 1. (Lifting the structure)

- Smoothness, Lipschitz continuity and semi-convexity are inherited by ℓ_q from ℓ in the respective sets.
- $(x, y) \notin sing(\ell)$ implies that

$$\frac{\partial^2}{\partial^j_x \partial y^i} \ell^q(x, y) \neq 0$$

The second condition is known as ℓ^q being twisted, the twisting condition together with locally bounded derivatives ensures the existence of solutions to the Monge problem. This corollary means that we will optimally transport mass outside the singular set.

5 Kantorovich Duality

As mentioned in the introduction, the fast development of optimal transport is due to Kantorovich's realization that the highly-non-linear problem posed by Monge could be realized by an infinite dimensional linear problem. Hence, we aim to establish the same result in globally hyperbolic spacetimes.
Here, the biggest difference appears, this difference will be discussed in the last section and involves a new concept:

Definition 5.1. (q-separation)
For $0 < q \leq 1$, we say $\mu, \nu \in \mathcal{P}_c(M)$ are q-separated by $\pi \in \Pi(\mu,\nu)$ and a lower-semicontinuous function $u : supp(\mu) \to \mathbb{R} \cup \{\infty\}$ and $v : supp(\nu) \to \mathbb{R} \cup \{\infty\}$ if

1. $$u(x) + v(y) \geq \frac{\ell(x,y)^q}{q} \qquad \text{for all } (x,y) \in supp(\mu \times \nu)$$

2. $$supp(\mu) \subseteq \left\{(x,y) \in supp(\mu \times \nu) : u(x) + v(y) = \frac{\ell(x,y)^q}{q}\right\}$$

3. $$\left\{(x,y) \in supp(\mu \times \nu) : u(x) + v(y) = \frac{\ell(x,y)^q}{q}\right\} \subseteq \{\ell > 0\}$$

we call S the set on the left hand-side of the last equation.

As stated in the introduction, in usual O.T. we have

Theorem U.O.T 1. *(Kantorovich duality) Let X and Y compact spaces.*
Let μ be a probability measure on X and ν a probability measure on Y.
Let $c : X \times Y \to \mathbb{R} \cup \{\infty\}$ be non-negative.
Let $\Gamma(\mu,\nu)$ be the set of Borel measures on $X \times Y$ such that their projections on X and Y are μ and ν respectively
Let Φ_c be the set of measurable functions $(\phi,\psi) \in L_1(\mu) \times L_1(\nu)$ such that

$$\phi(x) + \psi(y) \leq c(x,y)$$

μ-a.e. in x y ν-a.e. in y.

Then

$$\inf_{\pi \in \Gamma(\mu,\nu)} \left\{ \int c(x,y) d\pi \right\} = \sup_{(\phi,\psi) \in \Phi_c} \left\{ \int \phi d\mu + \int \psi d\nu \right\}$$

Hence as we know the Lorentz distance acts like our cost function we expect a formula like:

$$\frac{\ell_q(\mu,\nu)^q}{q} = \inf_{(\phi,\psi) \in \Phi_{\ell^q/q}} \left\{ \int_M \phi d\mu + \int_M \psi d\nu \right\} \tag{1}$$

where of course dividing by q didn't change the behaviour of the supremum (as it appears also in $\Phi_{\ell^q/q}$.
Now, the problem here with this formula is that the infimum can easily be $-\infty$ because of the condition of the Lagrangian on curves that are not future-directed, this is the reason why the paper introduces the concept of q-separation.

5.1 The Fenchel-Young transform and c-cyclically monotone sets

In this section two of the most important concepts in optimal transport are presented.

- The Fenchel-Young transform/Legendre transform/convex transform/ convex conjugate One of the main arguments in optimal transport is to be able to improve a dual pair (ϕ, ψ) to increase the dual cost (namely the supremum in Kantorovich's formula) remaining in the feasible set. The idea is to get the best possible c-concave function.

  Definition 5.2. For a given function $f : \mathbb{R}^n \to \mathbb{R}$ we define the Legendre transform, or convex conjugate f^* as follows:

  $$f^*(x) = \sup_{y \in \mathbb{R}^n} \{x \cdot y - f(y)\}$$

  It is clear by f^* being a supremum that for every $y \in \mathbb{R}^*$, the so-called Fenchel's inequality:

  $$x \cdot y \leq f(x) + f^*(y)$$

  Even though this formula looks trivial, it yields very non-trivial results, like Young's inequality. A more general definition defines f^* in the topological dual of our original space and replaces $x \cdot y$ by the dual pairing $\langle y, x \rangle$.

  Definition 5.3. (c-transform) For a given function $f : \mathbb{R}^n \to \mathbb{R}$ we define the c-Legendre transform, or c-conjugate f^c as follows:

  $$f^c(x) = \sup_{y \in \mathbb{R}^n} \{c(x, y) - f(y)\}$$

  Definition 5.4. (Restricted c-transform) For a given function $f : \mathbb{R}^n \to \mathbb{R}$ we define the restricted to $\mathbf{Y}$, c-Legendre transform, or c-conjugate f^c as follows:

  $$f^c(x) = \sup_{y \in \mathbf{Y}} \{c(x, y) - f(y)\}$$

  The idea is to consider a pair of functions $(\phi, \psi) \in \Phi_c$ and note that $(\phi^c, \psi^c) \in \Phi_c$ with improving cost by Fenchel's inequality. This technique, crucial to finding maximizers in Kantorovich's formula is sometimes referred to as the **double convexification trick**. For the context of Lorentzian manifolds that we are addressing, let us make a shorter notation for the ℓ^q/q-convex conjugate:

  $$v_q := v^{\ell^q/q}$$

- The second amazing concept from optimal transport is c-cyclically monotonocity. Intuitively, a set is c-cyclically monotone if there is no possible way to reorder points on that set such that the total cost of those points is decreased.

  Definition 5.5. A set S is called c-cyclically monotone if for any finite collection of points in S, $\{(x_i, y_i)\}_{i=1}^n$ one has

  $$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{\sigma(i)})$$

  for any permutation σ of $\{1, \ldots, N\}$.

  This concept is fundamental to optimal transport as one can intuitively see that optimal transference plans, are supported in c-cyclically monotone sets: If they were not, we would choose a permutation of the points σ and define a new plan $\tilde{\pi} = \pi + \epsilon \mathbf{1}_{\sigma \neq Id} - \epsilon \mathbf{1}_{\sigma = Id}$ this would decrease total cost, contradicting optimality of π for sufficiently small ϵ. What we did was we "changed" the measure only where we can decrease cost by permuting. This argument is not rigorous, and should be understood as intuitive only. It turns out, from convex analysis we know that gradients of convex functions are c-cyclically monotone, this is the connection why optimal transport maps must be gradients of convex functions (Brenier's Theorem).

Armed with these two powerfull concepts of optimal transport, let us look into Kantorovich's duality proposed in [6]

Theorem 5.1. *(Duality under q-separation)*
Again let $0 < q < 1$ If $\mu, \nu \in \mathcal{P}_c(M)$ are q-separated by the triple (π, u, v) then

1. *$(u, v) = (v^q, u^q)$ on supp$(\mu \times v)$.*
2. *S (as in definition 5.1) is compact and ℓ^q-**cyclically monotone***
3. *(u, v) minimize equation 1 while π maximizes ℓ^q.*

Proof. Let v^q be the restricted to *supp*(ν) ℓ^q/q-convex conjugate of v and similarly the one for u and μ. The supremum on v^q is attained because we know that ℓ is upper semicontinuous and so is $-v$, because v is lower-semi-contuinous by assumption of q-separation, and we know that upper semicontinuous functions attain maximums on compact sets (and *supp*(ν) is compact as we assumed $\nu \in \mathcal{P}_c(M)$. The definition of q-separation implies that $u \geq v^q$ by taking the supremum in the first bullet of 5.1. As X is also compact, so is S as uppersemicontinous image of a compact set (our space is Hausdorff). By compactness of S, let $(x, y) \in S$, then y maximizes $v^q(x)$ meaning that $v^q(x) = \ell^q(x, y)/q - v(y)$. But $S \subseteq \{\ell > 0\}$ implies that $\ell^q(x, y)$ is finite, because y lies in the chronological future of x, but

being in S means that $u(x) = v^q(x)$. The second one is exactly the same.
Part 2. Let σ be a permutation of the first N numbers, then for any $\{(x_i, y_i)\}_{i\leq N}$ we can compute:

$$\frac{1}{q}\sum_{i=1}^{N} \ell(x,y)^q = \sum_{i=1}^{N} u(x_i) + v(y_i) = \sum_{i=1}^{N} u(x_i) + v(y_{\sigma(i)}) \geq \frac{1}{q}\sum_{i=1}^{N} \ell^q(x_i, y_{\sigma(i)})$$

where the first equality is the definition of S, the second one because for the sum the order does not matter and the inequality by the first condition of q-separation.
Part 3 is a direct consequence of the definition of S and the fact that π vanishes outside S, hence equation 1 is established.

□

The condition of q-separation seems a little strong, but the next proposition gives an intuitive sufficient condition

Proposition 5.1. *(Existence of q-separation) Let $\mu, \nu \in \mathcal{P}_c(M)$, if $supp(\mu\times\nu) \subseteq M\backslash\{\ell \leq 0\}$ then μ and ν are q-separated.*

Proof. Because ℓ is continuous on $\{\ell > 0\}$, and $supp(\mu \times \nu)$ is compact, a minimizer exists by usual optimal transport techniques (tightness of Π which is too long to write). This minimizer and the duality formula give the desired triple for q-separation.

□

6 Characterizing optimal maps and McCann's theorem in general relativity

The aim of this section is to understand how different the characterization of solutions for Monge's problem is in Lorentzian manifolds, before I state both theorems (Riemannian and Lorentzian) we must have certain concepts. I won't do the proofs of this section as they require to many new concepts, but the ideas will be clear.

Proposition 6.1. *(Continuous inverse of the exponential) Fix $0 < q < 1$ and let $\mu_0, \mu_1 \in \mathcal{P}_c(M)$ be q-separated. Let X_i be the support of μ_i. Let $s \in (0,1)$. There exists a map $W : Z_s(S) \subseteq M \to S$ such that if μ_s lies on a q-geodesic, then $W_{\#}\mu_s$ attains $\ell_q(\mu_0, \mu_1)$ Further, $\overline{z}_s \circ W = Id$ on $\overline{z}_s(S)$.*

The proof of this proposition requires too many new concepts, but we can understand that under q-separation we can invert the exponential function via the use of midpoints.
The continuous map W maps $Z_s(S)$ into $S \subseteq M \times M$ so it has two components, namely we write

$$W = (U_s(z), V_s(z)$$

where s is fixed in $(0,1)$ and z is a midpoint.
For (u,v) as in 5.1, let us define some related functions

$$(\overline{u},\overline{v}) = (1-t)^q(u\circ U_t - t^{1-q}\frac{\ell^q}{q}\circ(U_t\times Id), v)$$

Essentially, what this functio is doing is pushing q-interpolating between the inverse of the exponential map and the flow along an optimal proper-time minimizing segment. Let us first recall McCann's interpolation theorem in usual optimal transport:

Theorem U.O.T 2. *(McCann's geodesic interpolation for Riemannian manifolds) Let M be a smooth connected complete Riemannian manifold, with an associated volume. Let $\mu,\nu\in\mathcal{P}_c(M)$, let $c(x,y)=d(x,y)$ the geodesic distance on M. If μ is absolutely continuous with respect to the volume on M, the Kantorovich problem has a unique solution given by*

$$\pi = (Id, T)_{\#}\mu$$

where $T(x)=\exp_x(-\nabla\phi)$, where $\nabla\phi$ is the gradient of the convex function ϕ. Further; the collection $\mu_s=[sId+(1-s)T]_{\#}\mu$ satisfies

$$W_2^2(\mu_s,\mu_r) = (s-r)W_s^2(\mu,\nu)$$

where W_2 is as in definition 3.5.

Theorem 6.1. *(McCann's interpolation under q-separation) Fix $0<q<1$, if $\mu,\nu\in\mathcal{P}^{ac}(M)$ are q-separated, write $X\times Y$ for the support of $\mu\times\nu$, then the Kantorovich problem has a unique solution given by*

$$\pi = (Id, F)_{\#}\mu$$

where $F(x)=\exp_x(DH(D\overline{u},x;q)$, H is the Hamiltonian and D denotes differentiation. Further, $\overline{u}$ is the restricted to supp(ν) ℓ^q/q-convex conjugate of u.

Even though these Theorems look exactly alike, I have lied with notation. $\overline{u}$ is involved in the derivative of H and the formula for $\overline{u}$ is significantly more complicated. Meaning that just formulating F is more complicated than T, due to the fact that ℓ^q involves the introduction of q-interpolations. As in the last section.

Proof. Suppose that $F_{\#}\mu=\nu$, let us sketch why π is the maximizer of the Kantorovich problem, the construction of F ensures that

$$\frac{\ell^q(x,F(x))}{q} = \overline{u}(x)+\overline{u}^q(F(x))$$

If we integrate this expression and use that F pushes μ onto ν we obtain

$$\frac{1}{q}\int\ell^q(x,y)d\pi(x,y) = \int\overline{u}(x)d\mu(x)+\int\overline{u}^q(y)d\nu(y)$$

from which the reader can easily conclude by the duality condition proved in the last section. □

7 The strong energy condition and entropy convexity

In this last section we state one of the main theorems in [6] and relate it to concepts that we already know in general relativity (**finally**!). This new characterizations are the reason why this area is developing and the main goal of the treatment. Even though there is a usual optimal transport analogue for this theorem (what I believe motivated McCann to develop this paper) I will not present it here, as I want to give this theorem the importance for it's own.

Definition 7.1. (Entropy)
For a twice differentiable function V on M and $\mu \in \mathcal{P}^{ac}(M)$ with Radon-Nykodin derivative ρ with respect to the volule $dm = e^{-V} dvol_g$ we define the entropy of μ:

$$E_V(\mu) = \int_M \rho(x)\log(\rho(x))e^{-V(x)}dvol_g(x)$$

whenever this integral is well-defined. If it is not, define it to be $-\infty$.
If $V = 0$, it is called the Boltzmann-Shannon entropy.
For our described q-geodesics μ_s, we write

$$e(s) := E_V(\mu_s)$$

Theorem 7.1. *(Displacement of relative entropy)*
Fix $0 < q < 1$ and let $V \in \mathcal{C}^2(M)$ on our globally hyperbolic spacetime. Let $s \in [0,1]$ and define F_s as in Theorem 6.1, and $\mu_s = F_{s\#}\mu_0$, if $e(0)$ and $e(1)$ are finite, then $e(s)$ is continuous and semiconvex on $[0.1]$ and continuously differentiable on $(0,1)$ with

$$e'(s) = \int_M DV_{F_s(x)}F_s'(x) - TrB_s(x)d\mu_0(x)$$

$$e''(s) = \int_M TrB_s^2(x) + (Ricc + D^2V)_{F_s(x)}(F_s'(x), F_s'(x)d\mu_0(x)$$

where $B_s(x) = A_s'(x)A_s^{-1}(x)$ where $A_s = \overline{D}F_s(x)$

REMARK The derivatives of e are distributional, meaning that they are defined with respect to integration against test functions, not in the classic sense. Here the term $\overline{D}$ refers to the approximate derivative:

Definition 7.2. (Approximate derivative)
A function $F : M \to N$ is said to be approximately differentiable a ta point x if there exists a map $\overline{F}$ differentiable at x such that:

$$\lim_{r\to 0} \frac{vol(\{x \in B_r(x)|F(x) \neq \overline{F}(x)\}}{vol(B_r(x))} = 0$$

In this case we call $\overline{D}F := D\overline{F}$ the approximate derivative of F

Sketch of ideas in the proof of THM 7.1. Suppose that the approximate jacobian of F exists and is smooth: $JF := |det(\overline{D}F)|$, then the transport condition $F_{s\#}\mu_0 = \mu_s$ can be written in terms of their densities (as they are absolutely continuous) as a simple change of variables called the Monge-Ampere equation:

$$\rho_s(F_s(x))JF_s(x) = \rho_0(x)$$

And we can compute the entropy:

$$e(s) = \int_M \log(\rho_s(y)) + V(y)d\mu_s$$

The change of variables $y \to F_s(x)$ and using that F_s pushes μ_0 to μ_s

$$e(s) = \int_M \log(\rho_s(F_s(x))) + V(F_s(x))d\mu_0(x)$$

We know use Monge-Ampere equation to replace $\rho_s(F_s(x))$, to get

$$e(s) = \int_M \log(\rho_0(x)) - \log(|JF_s(x)|) + V(F_s(x))d\mu_0(x)$$

□

The next steps are to find a second difference representation of the function inside the integral, and differentiate.

Finally we get to the last definition, which is one of the motivations of this research area. In many works, it has been shown that an appropriate inequality involving the dimension of the manifold and the curvature is enough to ensure nice geometric properties. This field of study defines the concept of the curvature-dimension condition. Optimal transport has been incredible to generalize this idea to non-smooth settings, namely abstract measure metric spaces. The work of [5], [9], [3], [7] are the best references. The condition is perfected in this setting:

Definition 7.3. (K, N, q)-convexity
For $K \in \mathbb{R}$ and $N > 0$ a function e is said to be (K, N)-convex if e is upper semi-continuous, $\{s : e(s) < \infty\}$ us connected and either $e^{-1}(-\infty)$ contains the interior of $\{s : e(s) < \infty\}$ or is empty. If it is empty:

$$e''(s) - \frac{1}{N}(e'(s))^2 \geq K$$

In our globally hyperbolic space time, a functional $E : \mathcal{P}(M) \to \mathbb{R}$ is said to (K, N, q)-convex if for each pair (μ_0, μ_1) there is a q-geodesic from μ_0 to μ_1 on which $E(\mu_s)$ is $(K\ell_q(\mu_0, \mu_1)^2, N)$-convex.

Definition 7.4. (Bakry-Emery curvature-dimension tensor) We define the following tensor:

$$Ricc_{ab}^{(N,V)} = Ricc_{ab} + \nabla_a \nabla_b V - \frac{1}{N-n} \nabla_a V \nabla_b V$$

is the famous Bakry-Emery curvature tensor.

Corollary 2. (Entropic convexity from timelike lower Ricci curvature bounds) In our globally hyperbolic spacetime (M^n, g), let V be twice differentiable on M and $N > n$, If

$$Ricc^{(N,V)}(v, v) \geq K|v|^2 \geq 0$$

for every timelike direction v, then for each $0 < q < 1$ the entropy $E_V(\mu)$ is (K, N, q)-convex.

7.1 Conclusions

Corollary 2 states a connection between two areas of mathematics: general relativity via the strong-energy condition (and hence lower bound on Einstein's equation) and optimal transport and curvature-dimension tensors. This is the first approach, note that once we defined de Bakry-Emery tensor, we can pose it in non-smooth settings, and so we can pose Einstein's equations in more general spaces, maybe obtaining "sub-solutions" that may hint to new knowledge. Note that corollary 2 also relates the strong energy condition with entropy!. This connection is new and could give interesting interpretations on the physical side. Let me attempt one: The strong energy condition can be understood as a condition limiting the amount of availability of energy in terms of it's acceleration. The concept of entropy can give a lot more interpretations and could possible link general relativity to other physical theories.

References

[1] Vladimir I. Bogachev , *Measure Theory*, Springer, 2007.

[2] John K. Beem, Paul Ehrlich, Kevin Easley, *Global Lorentzian Geometry*, CRC Press,2nd edition, 1996.

[3] John Lott and Cedric VIllani,*Ricci curvature for metric-measure spacesvia optimal transport*, Annals of Mathematics,169(2009), 903–991

[4] Michal Eckstein, Tomasz Miller, *Causality for nonlocal phenomena*, arXiv:1510.06386

[5] Karl-Theodor Sturm, *On the geometry of metric measure spaces*, Acta Math. ,Volume 196, Number 1 (2006), 65-131.

[6] Robert J McCann, *Displacement convexity of Boltzmann's entropy characterizes the strong energy condition from general relativity*, arXiv1808.1536v2.

[7] RObert J McCann and Peter Topping, *Ricci flow, entropy, and optimal transportation* , with. (formely titled Diffusion is a 2-Wasserstein contraction on any manifold evolving by reverse Ricci flow) Amer. J. Math. 132 (2010) 711-730

[8] Andrea Mondino, Stefan Suhr, *An optimal transport formulation of the Einstein equations of general relativity*, arXiv:1810.13309

[9] Matthias Erbar, Kazumasa Kuwada, Karl-Theodor Sturm, *On the Equivalence of the Entropic Curvature-Dimension Condition and Bochner's Inequality on Metric Measure, Spaces*, arXiv:1303.4382