EXISTENCE AND STRUCTURE OF SOLUTIONS OF STEINER PROBLEMS IN OPTIMAL TRANSPORT

by
Jonathan Dahl

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March, 2010

Abstract

We study the Steiner problem of finding a minimal spanning network in the setting of a space of probability measures with metric defined by cost of optimal transport between measures. Existence of a solution is shown for the Wasserstein space $P_p(\mathcal{X})$ over any base space $\mathcal{X}$ which is a separable, locally compact Hadamard space. Structural results are given for the case $p = 2$ under further restrictions on $\mathcal{X}$.

READERS: Chikako Mese (Advisor) and Joel Spruck.

Acknowledgments

I would first like to thank my advisor, Professor Chikako Mese. Her guidance, patience, and enthusiasm were invaluable over the past few years. This thesis would not exist without her help.

I would also like to thank Professor William Goldman and Professor Richard Schwartz for introducing me to the beauty of geometry. Furthermore, I would not have nearly as successful as I was in studying geometry further without the endless stream of useful insights from Professor William Minicozzi and Professor Joel Spruck, for which I am eternally grateful.

I am also indebted to my fellow graduate students for many helpful discussions over the years, in particular John Baber, Christine Breiner, Stephen Kleene, Joel Kramer, Longzhi Lin, Duncan Sinclair, and Patrick Zulkowski.

I would like to thank my family for being supportive and understanding during my graduate career.

I dedicate this dissertation to my wife, Erin, whose loving support and faith in me make the impossible seem possible.

Contents

Chapter 1

Introduction

In this dissertation, we investigate the geometry of optimal transport problems by finding hypotheses under which the Steiner problem of finding a minimal network between a given set of points can be solved in the metric space of probability measures on a base space. Here the distance between two probability measures is defined by the minimal transport cost between them for some fixed infinitesimal cost function. Our main result is that under certain geometric assumptions on the base space, the Steiner problem for the space of probability measures is always solvable.

The problem of optimal transport, originally proposed by Monge, has a long history of investigation and application ([Vil09] is an extensive reference). Roughly stated, the problem involves one who has an initial configuration of mass and would like to transport it to a terminal configuration of mass, doing so at least cost. For instance, one might have a set of water towers and a region of drought that one would like to relieve as quickly as possible. Abstractly, this becomes a constrained optimization problem in a space $\mathcal{P}$ of probability measures over the base space. Unfortunately, the mere existence of a solution is difficult to come by, due to the non-linear nature of the problem. It was over two hundred years before Kantorovich [Kan04b, Kan04a]

Material in this thesis will be published in *Steiner problems in optimal transport*, Transactions of the American Mathematical Society, American Mathematical Society, Providence, RI.

provided serious progress by formulating and solving a weak version of the problem. We will focus on this Monge-Kantorovich problem, defined in detail below.

Returning to our drought problem, suppose that the drought and even the construction of the water towers have yet to occur. The question becomes where to build the water towers to best prepare for possible droughts. If there are multiple possible droughts one wishes to protect against, but one can only afford enough water towers to combat a single drought at a time, one wishes to find a configuration of water towers which is nicely balanced amongst the possible droughts. We will investigate this by solving Steiner-type problems in the space $\mathcal{P}$ of probabilities. A Steiner problem is a search for a length minimizing network, usually satisfying some boundary conditions, in a metric space.

As Steiner solutions can be considered generalized geodesics [MY06], we cannot reasonably hope to solve the classical Steiner problem if $\mathcal{P}$ is not a geodesic space. Therefore, we will also define and solve for weak solutions of a Monge-Kantorovich-Steiner problem. The main idea in our definition of the weak problem is that, in a geodesic space, the edges of a Steiner solution are always geodesic segments, so the problem only sees distances of a finite point configuration. The weak problem is then to minimize the sum of these distances in place of the sum of the lengths of the connecting paths. One may then discuss weak solutions of the problem in any metric space.

Steiner problems are traditionally solved via local compactness arguments; however, as we cannot expect local compactness from our space $\mathcal{P}$, we will instead need to argue using the geometry of the base space. In particular, we show:

Main Theorem. *Suppose $\mathcal{X}$ is a separable, locally compact Hadamard space or a compact complete metric space. Then for any $p > 1$, the parameterized and general versions of the Steiner problem are solvable in $(P_p(\mathcal{X}), W_p)$ for arbitrary boundary data.*

Here $(P_p(\mathcal{X}), W_p)$ is the p-Wasserstein space, whose definition and basic properties are recalled in Section 2.3. The argument actually gives a technically more general result, listed precisely as Theorem 3.3.1 in Section 3.3.

We prove the Main Theorem by first considering the weak Monge-Kantorovich-Steiner problem only for boundary data of compact support. The boundary data is then supported on a convex compact subset H of the base space. By examining the orthogonal projection onto H, we see that the problem reduces to finding a solution with measures having support restricted to remain in H. We may then apply Prokhorov's Theorem to show that the direct method of minimization gives existence of a solution. The Main Theorem will then follow by a suitable approximation argument.

We conclude by using the geometry of Wasserstein spaces of order 2 to study the structure of the Steiner solutions. The standard variational argument for showing solutions of planar Steiner problems have vertex degrees of at most 3 is partially generalized to Steiner problems in $(P_2(\mathbb{R}^n), W_2)$. We show that for boundary data of compact support, vertices of degree greater than 3 in a Steiner solution cannot be absolutely continuous with respect to Lebesgue measure. This structural result relies heavily on the nonnegative curvature of $(P_2(\mathbb{R}^n), W_2)$, shown independently by Sturm [Stu06], and Lott and Villani [LV09]. Both of these proofs fail to generalize even to Wasserstein metrics W_p of order $p \neq 2$, so we include Appendix A as a possible foundation for future work on lower curvature bounds in optimal transport problems with more general cost functions and other infinite-dimensional settings. We then discuss the strong degree to which $(P_2(\mathcal{X}), W_2)$ lacks an upper bound on curvature. Finally, we investigate the special case of Gaussian measures as boundary data for Steiner problems in $(P_2(\mathbb{R}^n), W_2)$.

Chapter 2

Background

We begin by providing some necessary background on Alexandrov curvature, Steiner problems, and optimal transport problems. We are then in a position where we can formally define the problems which will the focus of the work: the Monge-Kantorovich-Steiner problem and the Monge-Steiner problem.

2.1 Alexandrov curvature

We first recall some notions of metric geometry, beginning with length.

Definition 2.1.1. *Let $(\mathcal{X}, d)$ be a metric space and $\gamma : [a, b] \to \mathcal{X}$ be a continuous map. $l(\gamma)$, the length of γ, is the supremum of*

$$\sum_{i=1}^{k} d(\gamma(t_i), \gamma(t_{i-1}))$$

over all k and choices of

$$a = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_k = b.$$

If $l(\gamma) < \infty$, γ is called rectifiable.

Definition 2.1.2. *Let $(\mathcal{X}, d)$ be a metric space. Let*

$$\mathrm{Path}(x, y) = \{\gamma : [a, b] \to \mathcal{X} \textit{ continuous}; \gamma(a) = x, \gamma(b) = y\}.$$

The intrinsic metric d_l is given by

$$d_l(x, y) = \inf_{\gamma \in \mathrm{Path}(x,y)} l(\gamma).$$

$\mathcal{X}$ is called a length space if $d = d_l$.

The intrinsic metric produces a length space as $(d_l)_l = d_l$.

Definition 2.1.3. *A shortest path in a length space is a rectifiable curve such that the distance between two points on the curve is equal to the length of the corresponding segment of the curve. A geodesic is a continuous path such that each point in the domain has a neighborhood on which the path is a shortest path. A length space is called a geodesic space if every two points are joined by a shortest path.*

We also need a notion of angle for metric spaces.

Definition 2.1.4. *Given three distinct points x, y, and z in a metric space $\mathcal{Y}$, the comparison angle $\tilde{\angle} xyz$ is defined as the corresponding angle in the triangle in $\mathbb{R}^2$ of sides $d(x, y), d(x, z), d(y, z)$.*

Explicitly,

$$\tilde{\angle} xyz = \arccos \frac{d^2(x, y) - d^2(x, z) + d^2(y, z)}{2d(x, y)d(y, z)}.$$

Definition 2.1.5. *If $\beta : [0, \varepsilon) \to \mathcal{Y}$ and $\gamma : [0, \varepsilon) \to \mathcal{Y}$ are two paths in a metric space $\mathcal{Y}$ with $\beta(0) = \gamma(0) = p$, then we define the angle*

$$\angle(\alpha, \beta) = \lim_{s,t \to 0} \tilde{\angle}(\alpha(s), p, \beta(t))$$

whenever the limit exists.

We may now define a notion of curvature for metric spaces.

Definition 2.1.6. *A length space $\mathcal{Y}$ is said to have nonnegative (nonpositive) Alexandrov curvature if it has a covering by neighborhoods $\{V_i\}$ such that for any two shortest paths $\beta : [0, \varepsilon) \to V_i$ and $\gamma : [0, \varepsilon) \to V_i$ with $\beta(0) = p$ and $\gamma(0) = p$,*

$$\tilde{\angle}(\alpha(s), p, \beta(t))$$

is nonincreasing (nondecreasing) in both s and t.

In particular, the angle between a pair of geodesics starting at a common point always exists in a space of nonnegative or nonpositive Alexandrov curvature. Nonnegative (nonpositive) Alexandrov curvature is equivalent to nonnegative (nonpositive) sectional curvature in the setting of Riemannian manifolds (Section 6.5 of [BBI01]).

There are several equivalent definitions of nonnegative Alexandrov curvature. For instance, it is shown in Proposition 10.1.1 in [BBI01] that for a locally compact length space $\mathcal{Y}$, nonnegative Alexandrov curvature is equivalent to having a covering by neighborhoods $\{V_i\}$ such that for any four distinct points $a, b, c, d \in V_i$, the quadruple condition

$$\tilde{\angle}bac + \tilde{\angle}cad + \tilde{\angle}dab \leq 2\pi$$

is satisfied. It should be noted that local compactness is not actually necessary here; however, we have local compactness everywhere we will be using the quadruple condition.

The theory of nonnegative Alexandrov curvature is significantly more developed under the additional assumption of finite Hausdorff dimension. We recall:

Definition 2.1.7. *Let $(\mathcal{X}, d)$ be a metric space. The δ-dimensional Hausdorff mea-*

sure of $\mathcal{X}$ is

$$\lim_{\varepsilon\to 0}\left(\inf_{\{S_i\}}\sum_{i=1}^{\infty}\frac{(\Gamma(1/2))^{\delta}}{\Gamma\left(1+\frac{\delta}{2}\right)}(\operatorname{diam} S_i)^{\delta}\right),$$

where the infimum is taken over all countable coverings $\{S_i\}$ of $\mathcal{X}$ with $\operatorname{diam} S_i < \varepsilon$ for all i. $\dim_H(\mathcal{X})$, the Hausdorff dimension of $\mathcal{X}$, is the infimum of the set of δ such that the δ-dimensional Hausdorff measure of $\mathcal{X}$ is zero.

Here

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}\,dt$$

is the usual Gamma function and the normalizing constant

$$\frac{(\Gamma(1/2))^{\delta}}{\Gamma\left(1+\frac{\delta}{2}\right)}$$

ensures that n-dimensional Hausdorff measure generalizes n-dimensional Lebesgue measure in Euclidean domains.

We now present a brief survey of some of the results known for finite-dimensional spaces of nonnegative Alexandrov curvature. First, we have dimensional homogeneity:

Theorem 2.1.8 (Theorem 10.6.1 of [BBI01])**.** *Let $\mathcal{X}$ be a space of nonnegative Alexandrov curvature. If $\dim_H(\mathcal{X})$ is finite, then $\dim_H(\mathcal{X})$ is an integer and $\dim_H(U) = \dim_H(\mathcal{X})$ for any open set $U \subset \mathcal{X}$.*

Finite-dimensional spaces of nonnegative Alexandrov curvature are well-behaved on a dense set.

Theorem 2.1.9 (Corollary 10.8.23 of [BBI01])**.** *Let $\mathcal{X}$ be an n-dimensional space of nonnegative Alexandrov curvature. $\mathcal{X}$ is then locally compact and contains an open dense set which is an n-dimensional manifold.*

We have good control on the space of finite-dimensional spaces of nonnegative Alexandrov curvature. We recall a few definitions to make this precise.

Definition 2.1.10. *A correspondence between two sets $\mathcal{X}$ and $\mathcal{Y}$ is a set $R \subset \mathcal{X} \times \mathcal{Y}$ such that the projections onto $\mathcal{X}$ and $\mathcal{Y}$ are both surjective.*

Definition 2.1.11. *The distortion of a correspondence R between metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ is defined by*

$$\operatorname{dis} R = \sup_{(x_1,y_1),(x_2,y_2)\in R} |d_{\mathcal{X}}(x_1, x_2) - d_{\mathcal{Y}}(y_1, y_2)|.$$

Definition 2.1.12. *The Gromov-Hausdorff distance between two metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ is defined by*

$$d_{GH}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \inf_R \operatorname{dis} R,$$

where the infimum is taken over all correspondences R between $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$. The pointed Gromov-Hausdorff distance between two pointed metric spaces $(\mathcal{X}, d_{\mathcal{X}}, p)$ and $(\mathcal{Y}, d_{\mathcal{Y}}, q)$ is defined by

$$d_{GH,p,q}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \inf_R \operatorname{dis} R,$$

where the infimum is taken over all correspondences R between $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ such that $(p, q) \in R$.

We may now state:

Theorem 2.1.13 (Theorem 8.5 in [BGP92])**.** *The class of space of nonnegative Alexandrov curvature with Hausdorff dimension $\leq n$ and diameter $\leq D$ is compact with respect to Gromov-Hausdorff convergence.*

In addition, we have several generalizations of results in Riemannian geometry to Alexandrov geometry, such as Milka's generalization [Mil67] of Toponogov's splitting theorem [Top64],

Theorem 2.1.14. *If a finite-dimensional space of nonnegative Alexandrov curvature $\mathcal{X}$ contains a line γ, i.e., $\gamma : (-\infty, \infty) \to \mathcal{X}$ and $\gamma|_{[a,b]}$ is a shortest path for any a and b, then $\mathcal{X}$ splits isometrically as $\mathbb{R} \times \mathcal{Y}$ for some space of nonnegative Alexandrov curvature $\mathcal{Y}$.*

and Perelman's generalization [Per91] of the Soul Theorem of Cheeger and Gromoll [CG72],

Theorem 2.1.15. *Let $\mathcal{X}$ be an n-dimensional space of nonnegative Alexandrov curvature. Then there is a compact convex subset $S \subset \mathcal{X}$, called the soul of $\mathcal{X}$, such that S is a deformation retract of $\mathcal{X}$.*

A major technical tool for finite-dimensional spaces of nonnegative Alexandrov curvature is the notion of an (n, ε)-strainer.

Definition 2.1.16. *A point $p \in \mathcal{X}$ is called an (n, ε)-strained point if there exist n pairs of points (a_i, b_i), called an (n, ε)-strainer, such that $\tilde{\angle} a_i p b_i > \pi - \varepsilon$, $\tilde{\angle} a_i p a_j > \frac{\pi}{2} - \varepsilon$, $\tilde{\angle} a_i p b_j > \frac{\pi}{2} - \varepsilon$, and $\tilde{\angle} b_i p b_j > \frac{\pi}{2} - \varepsilon$.*

A strainer may be thought of as an approximate orthogonal frame. The maximal n for which $\mathcal{X}$ has an (n, ε)-strained point is $n = \dim_H(\mathcal{X})$.

Definition 2.1.17. *If $n = \dim_H(\mathcal{X})$ and p is an (n, ε)-strained point for all $\varepsilon > 0$, p is called a regular point.*

In a small enough neighborhood of a regular point p, $\mathcal{X}$ looks like $\mathbb{R}^n$. We require a slightly different notion of distance between metric spaces to state this precisely.

Definition 2.1.18. *The Lipschitz distance between two metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ is given by*

$$d_L(\mathcal{X}, \mathcal{Y}) = \inf_{g : \mathcal{X} \to \mathcal{Y}} \log(\max\{\operatorname{dil} g, \operatorname{dil}(g^{-1})\}),$$

where the infimum is taken over all bi-Lipschitz homeomorphisms $g : \mathcal{X} \to \mathcal{Y}$ and the dilation is given by

$$\operatorname{dil} g = \sup_{x_1, x_2 \in \mathcal{X}} \frac{d_{\mathcal{Y}}(g(x_1), g(x_2))}{d_{\mathcal{X}}(x_1, x_2)}.$$

The local relation between a neighborhood of a regular point p and $\mathbb{R}^n$ is given by:

Lemma 2.1.19 (Theorem 10.9.16 of [BBI01]). *For every integer $n \geq 1$ and every $\varepsilon > 0$, there is a $\delta > 0$ such that if p is an (n, δ)-strained point in an n-dimensional Alexandrov space, then p has a neighborhood which is ε away in the Lipschitz distance from a region in $\mathbb{R}^n$. In particular, if p is regular, ε may be chosen arbitrarily small.*

Regular points are a powerful tool due to the following related form of Theorem 2.1.9.

Theorem 2.1.20 (Corollary 10.9.13 of [BBI01]). *If $\mathcal{X}$ be an n-dimensional space of nonnegative Alexandrov curvature, then the set of regular points is dense.*

We will also make use of the following equivalent definition of nonpositive Alexandrov curvature.

Definition 2.1.21. *A space of nonpositive Alexandrov curvature is a length space which can be covered by a family of open sets $\{V_i\}$ such that for each V_i:*

1. *There exists a shortest path in V_i connecting any two points in V_i.*

2. *For any $a, b, c \in V_i$ with each pair of points connected by a shortest path and any point d in the shortest path ac, let $\Delta\tilde{a}\tilde{b}\tilde{c}$ be the comparison triangle for Δabc in $\mathbb{R}^2$, i.e. $|ab| = |\tilde{a}\tilde{b}|$, $|ac| = |\tilde{a}\tilde{c}|$ and $|bc| = |\tilde{b}\tilde{c}|$, and let $\tilde{d}$ be the point in $\tilde{a}\tilde{c}$ such that $|ad| = |\tilde{a}\tilde{d}|$. Then $|bd| \leq |\tilde{b}\tilde{d}|$. (Intuitively, Δabc is skinnier than $\Delta\tilde{a}\tilde{b}\tilde{c}$.)*

The geometry of a space of nonpositive Alexandrov curvature better controlled under the assumption of simple connectivity.

Definition 2.1.22. *A Hadamard space (or complete* $CAT(0)$ *space) is a simply connected complete space of nonpositive Alexandrov curvature.*

Hadamard spaces $\mathcal{X}$ are important because the curvature conditions hold for all triangles in $\mathcal{X}$, not just small triangles (Theorem 9.2.9 in [BBI01]). In our discussion, this allows us to ensure the convexity of distance functions on $\mathcal{X}$, which will give us good control of convex hulls. In particular, $\mathcal{X}$ is locally convex. We will need $\mathcal{X}$ to be separable, so we note that this is always true for $\mathcal{X}$ of finite Hausdorff dimension.

Lemma 2.1.23. *If a metric space* $(\mathcal{X}, d)$ *has finite Hausdorff dimension* δ*, then* $\mathcal{X}$ *is separable.*

Proof. By definition, $\mathcal{X}$ has $(\delta+1)$-dimensional Hausdorff measure zero, so

$$\lim_{\varepsilon \to 0} \left(\inf_{\{S_i\}} \sum_{i=1}^{\infty} (\operatorname{diam} S_i)^{\delta+1} \right) = 0,$$

where the infimum is taken over all countable coverings $\{S_i\}$ of $\mathcal{X}$ with $\operatorname{diam} S_i < \varepsilon$ for all i. In particular, such countable coverings exist for all $\varepsilon \leq 1/N$ for some large N. For each integer $k \geq 1$, choose such a covering $\{S_{k,i}\}$ with $\operatorname{diam} S_{k,i} < 1/(N+k)$ and points $x_{k,i} \in S_{k,i}$. For every $x \in \mathcal{X}$ and every k, there is some i such that $x \in S_{k,i}$, so $d(x, x_{k,i}) < 1/(N+k)$. Thus the set $\{x_{k,i}; k, i \in \{1, 2, 3, \dots\}\}$ is a countable dense subset of $\mathcal{X}$. $\square$

2.2 Steiner problems

Steiner problems are concerned with finding minimal networks between a fixed set of points in a metric space. More precisely, a network Γ is a continuous map $\phi : G \to \mathcal{X}$ where $\mathcal{X}$ is a metric space and G is a graph, topologized in the standard way, called the parametric graph of Γ. ϕ may be decomposed into a union of curves, allowing one to compute the total length $l(\Gamma) = l(\phi)$ by working on each curve separately. The Steiner problem is to find a network of minimal length in some set of networks. We will focus on two cases.

Definition 2.2.1. *The parameterized Steiner problem for a graph G and a metric space $\mathcal{X}$ is defined as follows: given k vertices $v_1, \ldots, v_k \in G$ and k points $p_1, \ldots, p_k \in \mathcal{X}$, find a network of minimal length in the set of all networks in $\mathcal{X}$ with parametric graph G that send v_i to p_i for $1 \leq i \leq k$.*

Definition 2.2.2. *The general Steiner problem for a metric space $\mathcal{X}$ is defined as follows: given k points $p_1, \ldots, p_k \in \mathcal{X}$, find a network of minimal length in the set of all networks $\phi : G \to \mathcal{X}$ such that G is a connected graph and $p_1, \ldots, p_k$ are contained in the image of the vertex set.*

We call the points $p_1, \ldots, p_k$ the boundary points of the problem. A solution to a Steiner problem is referred to as a Steiner minimal tree. If $\mathcal{X}$ is a complete, locally compact, geodesic space, then the parameterized and general Steiner problems are solvable for any boundary points [IT94]. For the parameterized problem, one simply applies the direct method of the calculus of variations to the set of networks with parametric graph G and all vertices within r of all of the boundary points $p_1, \ldots, p_k$, where r is the diameter of the set $\{p_1, \ldots, p_k\}$. The general case then follows by a simple argument, given in the proof of Theorem 3.2.2. Conversely, since for boundary points p_1, p_2, these Steiner problems are equivalent to the geodesic problem, $\mathcal{X}$ must be a geodesic space for solutions to exist for arbitrary boundary data. We will show by an explicit class of examples, however, that local compactness is not a necessary condition for existence of solutions to arbitrary boundary data.

If we do not allow new vertices, but instead look for a length-minimizing spanning subgraph of the complete graph on our given boundary points, a solution is called a minimal spanning tree. We are minimizing over a finite number of graphs here, so the minimal spanning tree problem may be solved by a simple algorithm. This is important, as the Steiner problem is much more difficult (in fact, even for $\mathbb{R}^2$ it is NP-hard [GGJ77]), and the minimal spanning tree may be seen as a good approximation. We define the Steiner ratio to codify the quality of the approximation of the Steiner problem by the minimal spanning tree problem.

Definition 2.2.3. *The Steiner ratio $\rho(\mathcal{X})$ of a metric space $\mathcal{X}$ is*

$$\inf_M \frac{L_s(M)}{L_a(M)},$$

where M is any finite set of points in $\mathcal{X}$, $L_s(M)$ is the infimum of the lengths of all networks spanning M, and $L_a(M)$ is the length of the minimal spanning tree of M.

In general, the Steiner ratio is in $[1/2, 1]$. It is trivial to see that the Steiner ratio of $\mathbb{R}$ is 1, but calculation of the Steiner ratio of the plane $\mathbb{R}^2$ took a good deal of effort. In 1968, Gilbert and Pollak [GP68] conjectured that the Steiner ratio of the plane is $\sqrt{3}/2$, achieved by the vertices of a regular triangle. It was not until 1990 that Du and Hwang [DH90] positively resolved the conjecture. A generalized Gilbert and Pollak conjecture stated that $\rho(\mathbb{R}^d)$ is achieved by the set M_d of $d+1$ vertices of a regular d-simplex, i.e.,

$$\rho(\mathbb{R}^d) = \frac{L_s(M_d)}{L_a(M_d)}.$$

By studying the Steiner problem for the vertices of a regular simplex, Chung and Gilbert [CG76] were able to provide an upper bound

$$\lim_{d\to\infty} \rho(\mathbb{R}^d) \leq \frac{\sqrt{3/2}}{2^{3/2}-1} < .66984,$$

significantly strengthening the bound

$$\rho(\mathbb{R}^2) = \frac{\sqrt{3}}{2} > .86602.$$

It should be noted that Chung and Gilbert did not determine the explicit value of $L_s(M_d)/L_a(M_d)$ for each d, merely bounding it from above by examining a particular candidate for the Steiner minimal tree. Furthermore, Du and Smith [DS96] have shown the generalized Gilbert and Pollak conjecture

$$\rho(\mathbb{R}^d) = \frac{L_s(M_d)}{L_a(M_d)}$$

to fail in every dimension $d \geq 3$. Explicit bounds for $\rho(\mathbb{R}^d)$ from the work of Du and Smith are only known in low dimensions, so it remains unknown if

$$\lim_{d\to\infty} \rho(\mathbb{R}^d) < \lim_{d\to\infty} \frac{L_s(M_d)}{L_a(M_d)},$$

and the Chung and Gilbert bound on $\lim_{d\to\infty} \rho(\mathbb{R}^d)$ has yet to be improved.

2.3 Optimal transport

Given two spaces $\mathcal{X}, \mathcal{Y}$ and two subsets of probability measures $\mathcal{P} \subset P(\mathcal{X})$ and $\mathcal{Q} \subset P(\mathcal{Y})$, we define the set of transport plans $\Pi(\mathcal{P}, \mathcal{Q})$ as the set of probability measures $\pi \in P(\mathcal{X} \times \mathcal{Y})$ such that $(\mathrm{proj}_{\mathcal{X}})_{\#}\pi \in \mathcal{P}$ and $(\mathrm{proj}_{\mathcal{Y}})_{\#}\pi \in \mathcal{Q}$. $(\mathrm{proj}_{\mathcal{X}})_{\#}\pi$ and $(\mathrm{proj}_{\mathcal{Y}})_{\#}\pi$ are called the marginals of π. Here $f_{\#}\mu$ denotes the push-forward of μ by f, defined by $f_{\#}\mu(A) = \mu(f^{-1}(A))$.

If we suppose that the cost of implementing a transport plan depends only on the structure of the spaces $\mathcal{X}$ and $\mathcal{Y}$, we might suppose that for some cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, the total cost of the transport plan is

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, d\pi(x, y).$$

If for some $\pi_0 \in P(\mathcal{X} \times \mathcal{Y})$ with marginals μ and ν, we have

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi_0(x,y) = \inf_{\pi\in\Pi(\{\mu\},\{\nu\})} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y),$$

then we say that π_0 is an optimal transference plan. In this case

$$\left[\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi_0(x,y)\right]^{\alpha}$$

is called the α-optimal cost between μ and ν. (We include the parameter α so that we can allow Wasserstein distance in this framework.)

Definition 2.3.1. *The Monge-Kantorovich problem for spaces $\mathcal{X}, \mathcal{Y}$ is defined as follows: given a cost function c, and measures $\mu \in P(\mathcal{X})$ and $\nu \in P(\mathcal{Y})$, find an optimal transference plan between μ and ν.*

The Monge-Kantorovich problem is solvable in very general settings.

Theorem 2.3.2 (Theorem 4.1 of [Vil09])**.** *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two Polish probability spaces. Let $a : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and $b : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be upper semicontinuous functions such that $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Suppose that $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous function such that $c(x,y) \geq a(x) + b(y)$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Then there exists an optimal transference plan between μ and ν.*

An important special case is when $\mathcal{X} = \mathcal{Y}$ is a Polish metric space with distance function d and $c = d^p$ for some $p \in [1, \infty)$. $1/p$-optimal cost is then the Wasserstein distance of order p, given by

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

For any arbitrary $x_0 \in \mathcal{X}$, we define

$$P_p(\mathcal{X}) = \left\{ \mu \in P(\mathcal{X}); \int_{\mathcal{X}} d^p(x, x_0)\, d\mu(x) < \infty \right\}.$$

W_p is then a metric on $P_p(\mathcal{X})$. Furthermore, if $\mathcal{X}$ is a complete, separable and locally compact length space and $p > 1$, then $P_p(\mathcal{X})$ is a geodesic space (Corollary 7.22 of [Vil09]).

2.4 Steiner problems in optimal transport

We wish to solve the Steiner problem in a probability space $\mathcal{P} \subset P(\mathcal{X})$. It is useful however to first consider a hybrid Monge-Kantorovich-Steiner problem. In a geodesic space, solutions of Steiner problems map edges to geodesics, so this becomes equivalent to minimizing a certain sum of distances. In a more general metric space, this correspondence need not hold, but one could view a solution which minimizes the sum of distances as a weak solution to the Steiner problem. Accordingly, the α-optimal cost of a network $\phi : G \to P(\mathcal{X})$ is defined as the sum over all edges $\{v_i, v_j\}$ of G of the α-optimal cost between $\phi(v_i)$ and $\phi(v_j)$. The optimal costs are achieved by solutions of the Monge-Kantorovich problem, so we consider the following Monge-Kantorovich-Steiner problems:

Definition 2.4.1. *The parameterized Monge-Kantorovich-Steiner problem for a graph G, a metric space $\mathcal{X}$, $\alpha > 0$ and a cost function c is defined as follows: given k vertices $v_1, \ldots, v_k \in G$ and k probability measures $\mu_1, \ldots, \mu_k \in \mathcal{P} \subset P(\mathcal{X})$, find a network of minimal α-optimal cost in the set of all networks in $\mathcal{P}$ with parametric graph G that send v_i to μ_i for $1 \leq i \leq k$.*

Definition 2.4.2. *The general Monge-Kantorovich-Steiner problem for a metric space $\mathcal{X}$, $\alpha > 0$ and a cost function c is defined as follows: given a subset $\mathcal{P} \subset P(\mathcal{X})$ and*

k probability measures $\mu_1, \ldots, \mu_k \in \mathcal{P}$, find a network of minimal α-optimal cost in the set of all networks $\phi : G \to \mathcal{P}$ such that G is a connected graph and $\mu_1, \ldots, \mu_k$ are contained in the image of the vertex set.

If for some $\pi \in \Pi(P(\mathcal{X}), P(\mathcal{Y}))$ with marginals $\mu \in P(\mathcal{X}), \nu \in P(\mathcal{Y})$ there exists a measurable map $T : \mathcal{X} \to \mathcal{Y}$ such that $\pi = (\mathrm{id}, T)_{\#}\mu$, then π is said to be deterministic, and T is called the transport map. The classical Monge problem is to look for an optimal deterministic transference plan. We may thus consider the following two problems as well:

Definition 2.4.3. *The parameterized Monge-Steiner problem for a graph G, a metric space $\mathcal{X}$, $\alpha > 0$ and a cost function c is defined as follows: given k vertices $v_1, \ldots, v_k \in G$ and k probability measures $\mu_1, \ldots, \mu_k \in \mathcal{P} \subset P(\mathcal{X})$, find a network of minimal α-optimal cost in the set of all networks in $\mathcal{P}$ with parametric graph G that send v_i to μ_i for $1 \leq i \leq k$, and each α-optimal cost is achieved by a deterministic transference plan.*

Definition 2.4.4. *The general Monge-Steiner problem for a metric space $\mathcal{X}$, $\alpha > 0$ and a cost function c is defined as follows: given k probability measures $\mu_1, \ldots, \mu_k \in \mathcal{P} \subset P(\mathcal{X})$, find a network of minimal α-optimal cost in the set of all networks $\phi : G \to \mathcal{P}$ such that G is a connected graph and $\mu_1, \ldots, \mu_k$ are contained in the image of the vertex set, and each α-optimal cost is achieved by a deterministic transference plan.*

Solvability of the Monge problem holds far less generally than solvability of the Monge-Kantorovich problem. For example, a transport map can only send a Dirac mass to another Dirac mass. We will therefore see the strongest results for the Monge-Kantorovich-Steiner problems and the classical Steiner problems.

Chapter 3

Existence of solutions

In this chapter, we investigate the existence of solutions to Monge-Kantorovich-Steiner and Monge-Steiner problems, culminating in the proof of our Main Theorem. The outline of the argument is as follows:

1. First, we show existence of solutions assuming $\mathcal{X}$ is compact.

2. We then show how appropriate geometric conditions on $\mathcal{X}$, mainly a suitable class of contraction self-mappings, allow us to generalize from the previous case to noncompact $\mathcal{X}$ with boundary data of compact support.

3. Finally, we consider a specific approximation of arbitrary boundary data by boundary data of compact support and prove existence in the general case.

3.1 The direct method of minimization

In the existence proof for a solution of the classical Steiner problem on $\mathbb{R}^n$, one shows that any minimizing sequence must eventually remain in a bounded set and applies a compactness argument. Unfortunately, bounded sets are not necessarily precompact in a probability space $P(\mathcal{X})$ endowed with a reasonable metric. We must therefore use another criterion for precompactness, which is given by the following:

Theorem 3.1.1 (Prokhorov's Theorem [Pro56]). *If $\mathcal{X}$ is a Polish space, then a set $\mathcal{P} \subset P(\mathcal{X})$ is precompact for the weak topology if and only if it is tight, i.e. for any $\varepsilon \geq 0$ there exists a compact set $K_\varepsilon \subset \mathcal{X}$ such that $\mu(\mathcal{X} \setminus K_\varepsilon) \leq \varepsilon$ if $\mu \in \mathcal{P}$.*

We now show that the direct method works if a tightness bound is assumed.

Proposition 3.1.2 (cf. Theorem 2.3.2). *Let $\mathcal{X}$ be a Polish space, and suppose $\alpha > 0$ and $c : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$ is a lower semicontinuous cost function. Fix boundary points $\mu_1, \dots, \mu_k \in P(\mathcal{X})$ and let $\lambda^{1,1}, \lambda^{1,2}, \dots, \lambda^{k+l,k+l} \in [0, \infty)$ be fixed edge coefficients. Let $\Phi : (P(\mathcal{X})^l) \to \mathbb{R}$ be the Monge-Kantorovich-Steiner functional*

$$\Phi(\mu_{k+1}, \dots, \mu_{k+l}) = \sum_{i,j=1}^{k+l} \lambda^{i,j} \left[\inf_{\pi \in \Pi(\{\mu_i\},\{\mu_j\})} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\, d\pi(x,y) \right]^\alpha .$$

If $\mathcal{P} \subset P(\mathcal{X})$ is tight, then there exist $\nu_1, \dots, \nu_l \in \overline{\mathcal{P}}$ such that

$$\Phi(\nu_1, \dots, \nu_l) = \inf_{\mathcal{P}^l} \Phi.$$

Proof. Define $F : [\Pi(P(\mathcal{X}), P(\mathcal{X}))]^{(k+l)^2} \to \mathbb{R}$ by

$$F(\pi_{1,1}, \dots, \pi_{k+l,k+l}) = \sum_{i,j=1}^{k+l} \lambda^{i,j} \left[\int_{\mathcal{X}\times\mathcal{X}} c(x,y)\, d\pi_{i,j}(x,y) \right]^\alpha ,$$

and note that

$$\inf_{\mathcal{P}^l} \Phi = \inf_{\mathcal{Q}} F$$

where $\mathcal{Q} \subset [\Pi(P(\mathcal{X}), P(\mathcal{X}))]^{(k+l)^2}$ is the set of $(k+l)^2$-tuples of transference plans with marginals matching fixed points $\mu_1, \dots, \mu_k$ where appropriate, with marginals in $\mathcal{P}$ otherwise and with internally consistent marginals. Since $\mathcal{X}$ is Polish, each set $\{\mu_i\}$ is tight. Also, $\mathcal{P}$ is tight by assumption. Thus $\mathcal{P}, \{\mu_1\}, \dots, \{\mu_k\}$ is a finite collection of precompact sets by Prokhorov's Theorem and $\mathcal{P}' = \mathcal{P} \cup \{\mu_1, \ldots, \mu_k\}$ is precompact and tight. $\mathcal{Q} \subset [\Pi(\mathcal{P}', \mathcal{P}')]^{(k+l)^2}$ by construction. $\Pi(\mathcal{P}', \mathcal{P}')$ is tight, hence precompact, by the following lemma:

Lemma 3.1.3 (Lemma 4.4 of [Vil09])**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $\mathcal{P} \subset P(\mathcal{X})$ and $\mathcal{Q} \subset P(\mathcal{Y})$ be tight subsets. Then the set $\Pi(\mathcal{P}, \mathcal{Q})$ of all transference plans whose marginals lie in $\mathcal{P}$ and $\mathcal{Q}$ respectively, is itself tight in $P(\mathcal{X} \times \mathcal{Y})$.*

Let $\{\pi^n_{i,j}\} \subset \mathcal{Q}$ be an F-minimizing sequence. Since $\{\pi^n_{1,1}\} \subset \Pi(\mathcal{P}', \mathcal{P}')$, by taking a subsequence we may assume $\pi^n_{1,1}$ converges weakly to some $\pi_{1,1} \in \overline{\Pi(\mathcal{P}', \mathcal{P}')}$. Taking a subsequence $(k+l)^2 - 1$ more times, we may even assume that for all i, j, $\pi^n_{i,j}$ converges weakly to some $\pi_{i,j} \in \overline{\Pi(\mathcal{P}', \mathcal{P}')}$. $\mathcal{Q}$ is clearly closed, so $(\pi_{1,1}, \ldots, \pi_{k+l,k+l}) \in \mathcal{Q}$. We cite another lemma to show that $(\pi_{1,1}, \ldots, \pi_{k+l,k+l})$ is an F-minimizer.

Lemma 3.1.4 (Lemma 4.3 of [Vil09])**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous cost function. Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function such that $c \geq h$. Let $(\pi_k)_{k \in \mathbb{N}}$ be a sequence in $P(\mathcal{X} \times \mathcal{Y})$, converging weakly to some $\pi \in P(\mathcal{X} \times \mathcal{Y})$, in such a way that $h \in L^1(\pi_k)$, $h \in L^1(\pi)$, and*

$$\int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi_k \to \int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi.$$

Then

$$\int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \leq \liminf_{k \to \infty} \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi_k.$$

In particular, if $c \geq 0$, then $F : \pi \to \int c\pi$ is lower semicontinuous on $P(\mathcal{X} \times \mathcal{Y})$, equipped with the topology of weak convergence.

Applying Lemma 3.1.4 for $h \equiv 0$,

$$\inf_{\mathcal{Q}} F \leq F(\pi_{1,1}, \ldots, \pi_{k+l,k+l}) \leq \liminf_{n \to \infty} F(\pi^n_{1,1}, \ldots, \pi^n_{k+l,k+l}) = \inf_{\mathcal{Q}} F$$

and $(\pi_{1,1}, \dots, \pi_{k+l,k+l})$ is F-minimizing. Taking marginals of $(\pi_{1,1}, \dots, \pi_{k+l,k+l})$ yields the desired Φ-minimizer. $\square$

3.2 Nonpositive curvature and tightness

Proposition 3.1.2 shows that the direct method for solving the Monge-Kantorovich-Steiner problem works as long as one has an a priori tightness estimate. We will now see how the geometry of the base space can provide this tightness estimate. In order to assure that the Monge-Kantorovich-Steiner problem is aware of our geometric assumptions, we will assume that the cost function is based on the distance function.

Lemma 3.2.1. *Let $\mathcal{X}$ be a separable Hadamard space, $\alpha > 0$ and $c : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$ a lower semicontinuous cost function of the form $c = \varphi \circ d$ where φ is a monotone non-decreasing function and d is the distance in $\mathcal{X}$. Let $\mu_1, \dots, \mu_k \in P(\mathcal{X})$ be fixed boundary points and let $\lambda^{1,1}, \lambda^{1,2}, \dots, \lambda^{k+l,k+l} \in [0, \infty)$ be fixed edge coefficients. Suppose that $\mu_1, \dots, \mu_k$ all have compact support. Let $\Phi : (P(\mathcal{X})^l) \to \mathbb{R}$ be the Monge-Kantorovich-Steiner functional*

$$\Phi(\mu_{k+1}, \dots, \mu_{k+l}) = \sum_{i,j=1}^{k+l} \lambda^{i,j} \left[\inf_{\pi \in \Pi(\{\mu_i\},\{\mu_j\})} \int_{\mathcal{X} \times \mathcal{X}} c(x, y)\, d\pi(x, y) \right]^{\alpha}.$$

Then there exist $\nu_1, \dots, \nu_l \in P(\mathcal{X})$ (with compact support) such that

$$\Phi(\nu_1, \dots, \nu_l) = \inf_{(P(\mathcal{X}))^l} \Phi.$$

Proof. Let K be a large enough compact set so that $\mu_1, \dots, \mu_k \in P(K)$. Since $\mathcal{X}$ is locally convex, $H = \overline{\mathrm{co}\,(K)}$ is compact [AT04].

As shown in Proposition II.2.4 in [BH99], there exists a unique orthogonal projection $\mathrm{proj}_H : \mathcal{X} \to H$ which is a distance non-increasing retraction of $\mathcal{X}$ onto H. So given $\pi \in \Pi(\{\mu_i\}, P(\mathcal{X}))$, we have $(\mathrm{proj}_H \times \mathrm{proj}_H)_\# \pi \in \Pi(\{\mu_i\}, P(H))$ and

$$\int_{\mathcal{X}\times\mathcal{X}} c(x,y)\, d(\mathrm{proj}_H \times \mathrm{proj}_H)_\# \pi(x,y) \leq \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\, d\pi(x,y).$$

Thus $\inf_{(P(\mathcal{X}))^l} \Phi = \inf_{(P(H))^l} \Phi$.

Since H is compact, choosing $K_\varepsilon = H$ yields that $P(H)$ is tight. $P(H)$ is also closed, so Proposition 3.1.2 implies that there exists $\nu_1, \ldots, \nu_l \in P(H)$ such that

$$\Phi(\nu_1, \ldots, \nu_l) = \inf_{(P(\mathcal{X}))^l} \Phi = \inf_{(P(H))^l} \Phi. \qquad \square$$

Theorem 3.2.2. *If $\mathcal{X}$ is a separable Hadamard space or a compact space, then the parameterized Monge-Kantorovich-Steiner problems are solvable for compactly supported boundary data with a lower semicontinuous cost function $c : \mathcal{X} \times \mathcal{X} \to [0, +\infty]$ of the form $c = \varphi \circ d$ where φ is a monotone non-decreasing function and d is the distance in $\mathcal{X}$.*

If $c = d^p$ and $\alpha = 1/p$ for some $p \in [1, \infty)$, then the general Monge-Kantorovich-Steiner problem is solvable as well.

Furthermore, the minimizing configuration in $P(\mathcal{X})$ produced in both cases consists of measures with compact support.

Proof. For the parameterized problem with graph G, label the boundary vertices as $v_1, \ldots, v_k$ and the remaining vertices as $v_{k+1}, \ldots, v_{k+l}$. Let $\lambda^{i,j}$ be $1/2$ if v_i and v_j are connected by an edge in G and 0 otherwise. Proposition 3.1.2 or Lemma 3.2.1 then provides the solution.

For the general problem, we follow the classical argument (see [IT94]). Let $\mathcal{K}$ be the set of finite connected graphs G with k distinguished boundary vertices, let Φ_G denote the corresponding Monge-Kantorovich-Steiner functional for each $G \in \mathcal{K}$. We wish to achieve

$$\inf_{G\in\mathcal{K}} \inf \Phi_G = \inf_{G\in\mathcal{K}} \Phi_G(\nu_1^G, \ldots, \nu_{|G|-k}^G),$$

where $(\nu_1^G, \ldots, \nu_{|G|-k}^G) \in (P(\mathcal{X}))^{|G|-k}$ is the solution of the parameterized problem. (Here $|G|$ denotes the number of vertices of G.)

Let $\mathcal{K}'$ be the subset of $\mathcal{K}$ consisting of trees, and let $\mathcal{K}''$ be the subset of $\mathcal{K}'$ where all vertices not in the distinguished boundary set have degree at least three. Since $\mathcal{K}''$ is a finite set, the infimum over $\mathcal{K}''$ is trivially achieved. It remains to show that

$$\inf_{G\in\mathcal{K}} \inf \Phi_G = \inf_{G\in\mathcal{K}'} \inf \Phi_G = \inf_{G\in\mathcal{K}''} \inf \Phi_G.$$

For any $G \in \mathcal{K}$ and any $(\nu_1, \ldots, \nu_{|G|-k}) \in (P(\mathcal{X}))^{|G|-k}$, we may take a tree $G' \in \mathcal{K}'$ such that G' is a subgraph of G and $\Phi_{G'}(\nu_1, \ldots, \nu_{|G|-k}) \leq \Phi_G(\nu_1, \ldots, \nu_{|G|-k})$ since

$$\lambda^{i,j} W_p(\mu_i, \mu_j)$$

is always non-negative and we are only possibly setting some of the $\lambda^{i,j}$ to zero. So

$$\inf_{G\in\mathcal{K}} \inf \Phi_G = \inf_{G\in\mathcal{K}'} \inf \Phi_G.$$

Similarly, for any $G' \in \mathcal{K}'$, removing interior vertices of degree one will keep us in $\mathcal{K}'$ and will not increase $\Phi_{G'}(\nu_1, \ldots, \nu_{|G|-k})$. We may also replace any interior vertex of degree two by an edge between its neighbors to obtain a graph $G'' \in \mathcal{K}''$. It follows from the triangle inequality for Wasserstein distances that $\Phi_{G''}(\nu_1, \ldots, \nu_{|G|-k}) \leq \Phi_{G'}(\nu_1, \ldots, \nu_{|G|-k})$. Thus

$$\inf_{G\in\mathcal{K}} \inf \Phi_G = \min_{G\in\mathcal{K}''} \Phi_G(\nu_1^G, \ldots, \nu_{|G|-k}^G).$$

The compactness of supports also follows from Lemma 3.2.1. $\square$

We note that in the $c = d^2$ and $\alpha = 1/2$ case, if $\mathcal{X}$ is a Riemannian manifold, G is a star and the boundary data is absolutely continuous, the optimal couplings given by Theorem 2.3.2 may be taken to be deterministic, i.e. the Monge problem is solvable (Theorem 10.40 of [Vil09]). Thus Theorem 3.2.2 gives:

Corollary 3.2.3. *If M is a Riemannian manifold, M is compact or has non-positive sectional curvature, G is a star, $c = d^2$, $\alpha = 1/2$ and each fixed boundary point μ_i is compactly supported and absolutely continuous with respect to the volume measure of M, then the parameterized Monge-Steiner problem is solvable.*

We may say quite a bit more for the classical Steiner problem.

Corollary 3.2.4. *If $\mathcal{X}$ is a compact space and $p > 1$, then the parameterized and general Steiner problems are solvable for arbitrary boundary data on $(P_p(\mathcal{X}), W_p)$. If $\mathcal{X}$ is a separable locally compact Hadamard space and $p > 1$, then the parameterized and general Steiner problems are solvable for compactly supported boundary data on $(P_p(\mathcal{X}), W_p)$.*

Proof. In either case, we have Monge-Kantorovich-Steiner solutions for the cost function d^p with $\alpha = 1/p$. The associated cost functional is W_p, which metrizes $P_p(\mathcal{X})$ as a geodesic space. Thus the adjacent measures may be joined by geodesics, forming a minimal network. □

Note that the geodesics in the Steiner network may be assumed to stay in the set $P_2(H)$ of measures of compact support, as can be seen by considering the geodesic problem as a parametric Steiner problem.

The curvature assumption on $\mathcal{X}$ was only used above to control properties of convex hulls. In particular, the arguments carry through for any complete separable length space $\mathcal{X}$ satisfying:

1. If $K \subset \mathcal{X}$ is compact, then $\overline{\mathrm{co}\,(K)}$ is compact.

2. There exists a compact set $K_0 \subset \mathcal{X}$ such that if $K \subset M$ is a compact set with $K_0 \subset K$, then there exists a distance non-increasing retraction $\operatorname{proj}_H : \mathcal{X} \to H$ where H is compact and $K \subset H$.

As mentioned above, condition 1 holds for any complete metric space which is locally convex, thus condition 1 holds for Riemannian manifolds (by the existence of strongly convex neighborhoods [dC92]) and for Alexandrov spaces of curvature bounded above (Proposition II.1.4 of [BH99]). If $\mathcal{X} = \mathcal{S} \times \mathcal{N}$ where $\mathcal{X}$ has the product metric, $\mathcal{S}$ is compact and $\mathcal{N}$ is a Hadamard space, then it is easy to show that condition 2 holds for $K_0 = \emptyset$ and $H = \mathcal{S} \times \overline{\operatorname{co}(\operatorname{proj}_{\mathcal{N}} K)}$ by setting

$$\operatorname{proj}_H = \operatorname{id} \times \operatorname{proj}_{\overline{\operatorname{co}(\operatorname{proj}_{\mathcal{N}} K)}},$$

where the second component function is the orthogonal projection in $\mathcal{N}$. We therefore have:

Proposition 3.2.5. *If $\mathcal{X} = \mathcal{S} \times \mathcal{N}$ where $\mathcal{X}$ is a complete, separable, locally convex length space endowed with the product metric, $\mathcal{S}$ is compact, $\mathcal{N}$ is a Hadamard space and $p > 1$, then both the parameterized and general versions of the Steiner problem are solvable in $(P_p(\mathcal{X}), W_p)$ for arbitrary boundary data of compact support.*

As a particular case, the hypothesis of Proposition 3.2.5 is satisfied for any Riemannian manifold M which splits isometrically as $M = S \times \mathbb{R}^n$ with S compact. One may think of this condition as being able to apply Toponogov's splitting theorem [Top64] enough times.

3.3 Proof of the Main Theorem

We also assumed compact support for the boundary data in order to say that the supports of the minimizing sequence could be assumed to be compact. For general boundary data $\mu_1, \ldots, \mu_k$, one only has tightness of the set $\{\mu_1, \ldots, \mu_k\}$. We will thus approximate the solution for boundary data $\mu_1, \ldots, \mu_k$ by a sequence of solutions for compact boundary data $\mu_1^n, \ldots, \mu_k^n$ and show that the approximate solutions converge to a solution of the original problem.

Theorem 3.3.1. *Suppose $\mathcal{X} = \mathcal{S} \times \mathcal{N}$ where $\mathcal{X}$ is a complete, separable, locally compact, locally convex length space endowed with the product metric, $\mathcal{S}$ is compact, $\mathcal{N}$ is a Hadamard space and $p > 1$. Then the parameterized and general versions of the Monge-Kantorovich-Steiner problem and the Steiner problem are solvable in $(P_p(\mathcal{X}), W_p)$ for arbitrary boundary data. (For $p = 1$, the Monge-Kantorovich-Steiner problems are solvable.)*

Proof. First consider the G-parameterized problem where each vertex of G is adjacent to the boundary. Let $\mu_1, \ldots, \mu_k \in P_p(\mathcal{X})$ denote the boundary data and let $\Phi : (P_p(\mathcal{X}))^l \to \mathbb{R}$ be the Monge-Kantorovich-Steiner functional. Since $\{\mu_1, \ldots, \mu_k\}$ is tight, we may choose $r_n > 0$ such that $\mu_i(\mathcal{X} \setminus B_{r_n}) \leq 1/n$ for all i, n, where B_{r_n} is the intrinsic closed ball of radius r_n about some fixed base point x_0. In particular, $\mu_i(B_{r_n}) \geq (n-1)/n$. B_{r_n} is compact since $\mathcal{X}$ is complete and locally compact. Define the cutoff measures

$$\mu_i^n = \frac{\mu_i|_{B_{r_n}}}{\mu_i|_{B_{r_n}}(\mathcal{X})} = \frac{\mu_i|_{B_{r_n}}}{\mu_i(B_{r_n})}.$$

Note that for $m \geq 2$, we have the tightness estimate

$$\mu_i^m(\mathcal{X} \setminus B_{r_n}) = \frac{\mu_i|_{B_{r_m}}(\mathcal{X} \setminus B_{r_n})}{\mu_i(B_{r_m})} \leq \frac{\mu_i(\mathcal{X} \setminus B_{r_n})}{\mu_i(B_{r_m})} \leq \frac{m}{m-1}\frac{1}{n} \leq \frac{2}{n}.$$

By Theorem 6.9 in [Vil09], a sequence μ^m converges to μ in the W_p metric if and only if $\mu^m \to \mu$ weakly and

$$\limsup_{m\to\infty} \int d^p(x, x_0)\, d\mu^m(x) \leq \int d^p(x, x_0)\, d\mu(x).$$

Thus the weak convergence of μ_i^m to μ_i and the bound

$$\int d^p(x,x_0)\,d\mu_i^m(x) = W_p^p(\mu_i^m,\delta_{x_0}) \le W_p^p(\mu_i,\delta_{x_0}) = \int d^p(x,x_0)\,d\mu_i(x)$$

imply that $\mu_i^m \to \mu_i$ in W_p.

Let Φ^m denote the Monge-Kantorovich-Steiner functional for the boundary data $\mu_1^m,\dots,\mu_k^m$. Since

$$b := \Phi(\delta_{x_0},\dots,\delta_{x_0}) \ge \Phi^m(\delta_{x_0},\dots,\delta_{x_0}),$$

b gives an upper bound on the infima for Φ and Φ^m.

Let $\nu_1^m,\dots,\nu_l^m$ solve the G-parameterized problem for the compact boundary data $\mu_1^m,\dots,\mu_k^m$. We now show that $\{\nu_1^m,\dots,\nu_l^m\}$ is tight. Let

$$\psi(n) = (b+1)\left(\frac{n}{2}\right)^{1/p} + 2r_n.$$

Suppose that there exists a pair (j,m) such that for some n

$$\nu_j^m(\mathcal{X}\setminus B_{\psi(n)}) > \frac{4}{n}.$$

Choose i such that ν_j^m and μ_i^m are G-adjacent. Since $\mu_i^m(\mathcal{X}\setminus B_{r_n}) < 2/n$, transporting from ν_j^m to μ_i^m must move at least $2/n$ of the mass from outside $B_{\psi(n)}$ to inside B_{r_n}. More precisely, if $\pi \in \Pi(\{\nu_j^m\},\{\mu_i^m\})$ then

$$\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n}) + \pi((\mathcal{X}\setminus B_{\psi(n)})\times(\mathcal{X}\setminus B_{r_n})) = \nu_j^m(\mathcal{X}\setminus B_{\psi(n)}) > \frac{4}{n},$$

and similarly

$$\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n}) + \pi(B_{\psi(n)}\times B_{r_n}) = 1-\mu_i^m(\mathcal{X}\setminus B_{\psi(n)}) > 1-\frac{2}{n}.$$

Adding inequalities we find

$$
\begin{aligned}
1+\frac{2}{n} &< 2\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n})+\pi(B_{\psi(n)}\times B_{r_n}) \\
&\quad + \pi((\mathcal{X}\setminus B_{\psi(n)})\times(\mathcal{X}\setminus B_{r_n})) \\
&\le 2\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n})+\pi(B_{\psi(n)}\times B_{r_n}) \\
&\quad + \pi((\mathcal{X}\setminus B_{\psi(n)})\times(\mathcal{X}\setminus B_{r_n}))+\pi(B_{\psi(n)}\times(\mathcal{X}\setminus B_{r_n})) \\
&= \pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n})+\pi(\mathcal{X}\times\mathcal{X}) \\
&= \pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n})+1
\end{aligned}
$$

since $\pi(\mathcal{X}\times\mathcal{X})=1$, and therefore $\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n})>2/n$. Thus

$$
\begin{aligned}
\int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\,d\pi(x,y) &\ge \int_{(\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n}} d^p(x,y)\,d\pi(x,y) \\
&\ge \int_{(\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n}} (\psi(n)-r_n)^p\,d\pi(x,y) \\
&= (\psi(n)-r_n)^p\pi((\mathcal{X}\setminus B_{\psi(n)})\times B_{r_n}) \\
&> \frac{2}{n}(\psi(n)-r_n)^p,
\end{aligned}
$$

and we see that

$$
W_p(\nu_j^m,\mu_i^m)\ge\left(\frac{2}{n}\right)^{1/p}(\psi(n)-r_n)\ge b+1.
$$

In particular,

$$
\Phi^m(\nu_1^m,\ldots,\nu_l^m)>b,
$$

which contradicts the minimality of Φ^m at $(\nu_1^m,\ldots,\nu_l^m)$. Therefore

$$
\nu_j^m(\mathcal{X}\setminus B_{\psi(n)})\ge\frac{4}{n}
$$

and $\{\nu_1^m,\ldots,\nu_l^m\}$ is tight.

For general graphs G, we may inductively show tightness of $\{\nu_1^m,\ldots,\nu_l^m\}$ for vertices $k+1$ edges away from the boundary by assuming tightness of such sequences for vertices k edges away from the boundary by the above argument. Since G is a finite graph, we cover all vertices in a finite number of iterations of the argument.

Taking subsequences, we may assume that for all j, $\nu_j^m \to \nu_j$ weakly. Since $\mu_i^m \to \mu_i$ weakly as well, we may use Lemma 3.1.3 to take subsequences and assume for all i, j that the optimal transference plans $\pi_{i,j}^m$ converge weakly to some $\hat{\pi}_{i,j}$ with the appropriate marginals.

Let $\varepsilon > 0$. Since $\mu_i^m \to \mu_i$ in W_p for all i, $\exists M$ such that $\forall m \geq M$,

$$\|\Phi^m - \Phi\|_{L^\infty((P_p(\mathcal{X}))^l)} \leq \varepsilon.$$

In particular,

$$\left| \inf_{(P_p(\mathcal{X}))^l} \Phi^m - \inf_{(P_p(\mathcal{X}))^l} \Phi \right| \leq \varepsilon.$$

As in the proof of Proposition 3.1.2, consider $F : [\Pi(P_p(\mathcal{X}), P_p(\mathcal{X}))]^{(k+l)^2} \to \mathbb{R}$ defined by

$$F(\pi_{1,1}, \ldots, \pi_{k+l,k+l}) = \sum_{i,j=1}^{k+l} \lambda^{i,j} \left[\int_{\mathcal{X}\times\mathcal{X}} d^p(x,y)\, d\pi_{i,j}(x,y) \right]^{(1/p)},$$

where again $\lambda^{i,j} = 1/2$ for edges of G and $\lambda^{i,j} = 0$ otherwise. By construction,

$$F(\pi_{1,1}^m, \ldots, \pi_{k+l,k+l}^m) = \Phi^m(\nu_1^m, \ldots, \nu_l^m) = \inf_{(P_p(\mathcal{X}))^l} \Phi^m.$$

By Lemma 3.1.4, F is lower semicontinuous and

$$\begin{aligned} F(\hat{\pi}_{1,1}, \ldots, \hat{\pi}_{k+l,k+l}) &\leq \liminf_{m\to\infty} F(\pi_{1,1}^m, \ldots, \pi_{k+l,k+l}^m) \\ &= \liminf_{m\to\infty} \inf_{(P_p(\mathcal{X}))^l} \Phi^m \\ &\leq \left(\inf_{(P_p(\mathcal{X}))^l} \Phi \right) + \varepsilon. \end{aligned}$$

Sending $\varepsilon \to 0$,

$$F(\hat{\pi}_{1,1}, \ldots, \hat{\pi}_{k+l,k+l}) = \inf_{(P_p(\mathcal{X}))^l} \Phi,$$

so the $\hat{\pi}_{i,j}$ are in fact optimal transference plans and

$$\Phi(\nu_1, \ldots, \nu_l) = F(\hat{\pi}_{1,1}, \ldots, \hat{\pi}_{k+l,k+l}) = \inf_{(P_p(\mathcal{X}))^l} \Phi.$$

The solution of the general problem follows as in the proof of Theorem 3.2.2. $\square$

Chapter 4

Investigations of further structure

We conclude by discussing how further specialization of our problem allows us to recover portions of the classical theory of the Steiner problem for the plane. We begin by partially generalizing the degree bound on vertices in a Steiner minimal tree from $\mathbb{R}^n$ to $(P_2(M), W_2)$, where M is a Riemannian manifold with isometric splitting $M = S \times \mathbb{R}^n$ where S is compact with nonnegative sectional curvature. After a brief motivating excursion into questions of upper curvature bounds for $(P_2(\mathcal{X}), W_2)$, we then consider the special case of Steiner problems in $(P_2(\mathbb{R}^d), W_2)$ with Gaussian boundary data.

4.1 Structure of Steiner trees in $(P_2(M), W_2)$

We now restrict our attention to the general Steiner problem on $(P_2(M), W_2)$ and investigate how the geometry of M can force structure on the parametric graph of a solution.

The following results of Lott and Villani will allow us to work geometrically on our space of probabilities.

Theorem 4.1.1 (Theorem A.8 of [LV09]). *Suppose M is a compact Riemannian* *manifold with nonnegative sectional curvature. Then for all $\mu_0, \ldots, \mu_3 \in P_2(M)$, in the W_2 metric*

$$\tilde{\angle}\mu_1\mu_0\mu_2 + \tilde{\angle}\mu_2\mu_0\mu_3 + \tilde{\angle}\mu_3\mu_0\mu_1 \leq 2\pi.$$

In particular, $(P_2(M), W_2)$ has nonnegative Alexandrov curvature.

By passing to limits in the inequality, we obtain

$$\angle(\gamma_1, \gamma_2) + \angle(\gamma_2, \gamma_3) + \angle(\gamma_3, \gamma_1) \leq 2\pi$$

for geodesics γ_i starting at μ_0.

Theorem 4.1.2 (Proposition A.33 of [LV09])**.** *Suppose M is a compact Riemannian manifold with nonnegative sectional curvature. Then for each absolutely continuous measure $\mu \in P_2(M)$, the tangent cone K_μ of $(P_2(M), W_2)$ at μ is a Hilbert space, under the inner product generated by angles of geodesics in the space of directions.*

We also recall a first variation formula for Alexandrov spaces.

Lemma 4.1.3 (Corollary 4.5.7 of [BBI01])**.** *Let $\mathcal{Y}$ be a complete, locally compact length space of nonnegative Alexandrov curvature, $p \in \mathcal{Y}$ and $\gamma : [0, \varepsilon) \to \mathcal{Y}$ a unit-speed shortest path. Let $l(t) = d(p, \gamma(t))$. Then*

$$\lim_{t \to 0^+} \frac{l(t_i) - l(0)}{t_i} = \min_{\sigma_0} \left[-\cos(\angle \sigma_0 \gamma) \right],$$

where the minimum is taken over all shortest paths from $\gamma(0)$ to p. (In particular, the limit exists and the minimum is achieved.)

We now assume $M = S \times \mathbb{R}^n$ where S is compact with nonnegative sectional curvature, and $\mu_1, \ldots, \mu_k \in P_2(M)$ have compact support. By Proposition 3.2.5, there is a general Steiner solution, which by the proof of Theorem 3.2.2, may be represented by a network Γ whose parametric graph G is a tree and all interior vertices have degree at least three. This is known as the canonical representative. We will now show that if an interior vertex does not have degree three, then the corresponding measure is not absolutely continuous.

First, note that for some compact $K \subset M$ containing the supports of $\mu_1, \ldots, \mu_k$, we have that $\Gamma(G) \subset P_2(H)$ for $H = \mathcal{S} \times \overline{\mathrm{co}\,(\mathrm{proj}_{\mathcal{N}} K)}$ as above. Γ is thus trivially a Steiner solution for the restrained general Steiner problem on $P_2(H)$, where we can apply Theorems 4.1.1 and 4.1.2 and Lemma 4.1.3.

By Theorem 4.1.1, it suffices to show that the angle between any pair of adjacent geodesics γ_1, γ_2 connected at an absolutely continuous $\mu_0 \in P_2(H)$ is at least $2\pi/3$. Suppose that

$$\angle \gamma_1 \mu_0 \gamma_2 < 2\pi/3.$$

Let e_1, e_2 be the edges corresponding to γ_1, γ_2, and let v be the vertex corresponding to μ_0. Split the vertex v into v_1, v_2 and create a new graph G' where v_1 is incident to exactly e_1, e_2 and a new edge e, and v_2 is incident to e and the remaining edges originally incident to v. There is an obvious graph homomorphism $h : G' \to G$ identifying v_1 and v_2. Let $\Gamma' : G' \to P_2(H)$ denote the network $\Gamma \circ h$. Γ' is clearly a (non-canonical) Steiner network, so it is a global and local minimizer for length.

Let N_1, N_2 be the unit vector representatives of γ_1, γ_2 in the tangent cone K_{μ_0} at μ_0. Since K_{μ_0} is an inner product space, there is a (unit-speed) geodesic $\eta : [0, \varepsilon) \to P_2(H)$ with $\eta(0) = \mu_0$ such that the angle between η and $N_1 + N_2$ is arbitrarily small. Let N be the unit vector representative of η and let $l(t)$ be the length of the network Γ'_t given by shifting v_1 to $\eta(t)$. By minimality of Γ',

$$\lim_{t \to 0^+} \frac{l(t_i) - l(0)}{t_i} \geq 0.$$

The edge e maps to $\eta([0, t])$ and thus has length t. The only other lengths changed are the images of e_1, e_2, so by Lemma 4.1.3,

$$\begin{aligned}0 \leq \lim_{t\to 0^+} \frac{l(t_i) - l(0)}{t_i} &= 1 + \min_{\sigma_1}\left[-\cos(\angle\sigma_1\eta)\right] + \min_{\sigma_2}\left[-\cos(\angle\sigma_2\eta)\right] \\ &\leq 1 - \cos(\angle N_1\eta) - \cos(\angle N_2\eta) \\ &= 1 - \langle N_1, N\rangle - \langle N_2, N\rangle \\ &= 1 - \langle N_1 + N_2, N\rangle \\ &= 1 - \|N_1 + N_2\| \cos(\angle(N_1 + N_2), N) < 0,\end{aligned}$$

since $\|N_1 + N_2\| > 1$ and $\cos(\angle(N_1 + N_2), N)$ is arbitrarily close to 1. This contradiction implies

$$\angle\gamma_1\mu_0\gamma_2 \geq 2\pi/3.$$

Summarizing, we have

Theorem 4.1.4. *Suppose M is a Riemannian manifold with isometric splitting $M = S \times \mathbb{R}^n$ where S is compact with nonnegative sectional curvature, and $\mu_1, \ldots, \mu_k \in P_2(M)$ have compact support. Then there is a Steiner solution in $(P_2(M), W_2)$ spanning $\mu_1, \ldots, \mu_k$. Furthermore, this solution has a canonical representative $\Gamma : G \to P_2(M)$ such that*

1. *G is a tree.*
2. *Vertices in G not mapped to $\mu_1, \ldots, \mu_k$ have degree at least three.*
3. *For any vertex v in $G \setminus \Gamma^{-1}(\{\mu_1, \ldots, \mu_k\})$, if $\Gamma(v)$ is absolutely continuous with respect to the volume measure, then the degree of v is three and all pairs of geodesics in $\Gamma(G)$ meeting at $\Gamma(v)$ do so with an angle of $2\pi/3$.*

The method of proof for Theorem 4.1.4 also applies to Steiner trees in locally compact, finite dimensional, nonnegatively curved Alexandrov space; one must simply replace the notion of absolute continuity of a measure with the notion of being a manifold point. This further illustrates the analogy between measures in $P_2(M) \setminus P_2^{ac}(M)$ and singular points in a finite dimensional Alexandrov space mentioned in [LV09].

One may see that the absolute continuity assumption for the vertex v in Theorem 4.1.4 is only used to establish the existence of ε-almost midpoints in the tangent cone K_v. These ε-almost midpoints always exist for finite dimensional Alexandrov spaces of nonnegative curvature, however [Hal00] has an infinite dimensional counterexample.

4.2 Unboundedness of curvature

As nonpositive curvature provides useful convexity properties for making uniqueness arguments for minimizers, one might hope that $(P_2(M), W_2)$ has nonpositive curvature for some class of manifolds M. In particular, one might ask if $P_2(\mathbb{R}^n)$ is Alexandrov flat. For $n \geq 2$, this is shown to be false by the following example of [AGS05]:

Consider $\mathbb{R}^2 \subset \mathbb{R}^n$, and let $\mu_0 = \frac{1}{2}(\delta_{(1,1)} + \delta_{(5,3)})$, $\mu_1 = \frac{1}{2}(\delta_{(-1,1)} + \delta_{(-5,3)})$ and $\mu_2 = \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-4)})$. The (constant speed) W_2-geodesic between μ_0 and μ_1 is given by

$$\mu_t = \frac{1}{2}(\delta_{(1-6t,1+2t)} + \delta_{(5-6t,3-2t)}), \text{ so } \mu_{1/2} = \frac{1}{2}(\delta_{(-2,2)} + \delta_{(2,2)}).$$

One may then compute

$$\begin{aligned}
W_2^2(\mu_0, \mu_1) &= \frac{1}{2}((1-(-6))^2 + (1-3)^2) + \frac{1}{2}((-1-5)^2 + (1-3)^2) = 40, \\
W_2^2(\mu_0, \mu_2) &= \frac{1}{2}((1-0)^2 + (1-(-4))^2) + \frac{1}{2}((5-0)^2 + (3-0)^2) = 30, \\
W_2^2(\mu_1, \mu_2) &= \frac{1}{2}((-1-0)^2 + (1-(-4))^2) + \frac{1}{2}((-5-0)^2 + (3-0)^2) = 30, \\
W_2^2(\mu_{1/2}, \mu_2) &= \frac{1}{2}((-2-0)^2 + (2-0)^2) + \frac{1}{2}((2-0)^2 + (2-(-4))^2) = 24, \\
d^2(\tilde{\mu}_{1/2}, \tilde{\mu}_2) &= (\sqrt{30})^2 - (\sqrt{40}/2)^2 = 20,
\end{aligned}$$

where $\tilde{\mu}_{1/2}, \tilde{\mu}_2$ are the corresponding points on the comparison triangle $\Delta\tilde{\mu}_0\tilde{\mu}_1\tilde{\mu}_2$ in $\mathbb{R}^2$. Since $W_2(\mu_{1/2}, \mu_2) > d(\tilde{\mu}_{1/2}, \tilde{\mu}_2)$, $(P_2(\mathbb{R}^d), W_2)$ cannot have nonpositive Alexandrov curvature.

We will now adapt the above example to show that the Alexandrov curvature of $(P_2(M), W_2)$ is unbounded from above for any Riemannian manifold M of dimension at least two. To state this precisely, we must generalize our definition of Alexandrov space.

Definition 4.2.1. *For $\kappa \in \mathbb{R}$, the model space M^n_κ is defined as*

- $M^n_0 = \mathbb{R}^n$,
- *if $\kappa > 0$, then M^n_κ is the n-dimensional sphere of sectional curvature κ,*
- *if $\kappa < 0$, then M^n_κ is the n-dimensional hyperbolic space of sectional curvature κ.*

Definition 4.2.2. *An Alexandrov space of curvature bounded above (below) by $\kappa \in \mathbb{R}$ is a length space which can be covered by a family of open sets $\{V_i\}$ such that for each V_i:*

1. *There exists a shortest path in V_i connecting any two points in V_i.*
2. *For any $a, b, c \in V_i$ and any point d in the shortest path ac, let $\Delta\tilde{a}\tilde{b}\tilde{c}$ be the comparison triangle for Δabc in the model space M^2_κ, i.e. $|ab| = |\tilde{a}\tilde{b}|$, $|ac| = |\tilde{a}\tilde{c}|$ and $|bc| = |\tilde{b}\tilde{c}|$, and let $\tilde{d}$ be the point in $\tilde{a}\tilde{c}$ such that $|ad| = |\tilde{a}\tilde{d}|$. Then $|bd|$ is less than (greater than) $|\tilde{b}\tilde{d}|$. (Intuitively, Δabc is skinnier (fatter) than $\Delta\tilde{a}\tilde{b}\tilde{c}$.)*

This is consistent with our previous definitions.

Consider $(P_2(\mathbb{R}^d), W_2)$ and $\mu_0, \mu_1, \mu_2, \mu_{1/2}$ as above. It is easy to check that the triangle comparison condition varies continuously in κ, so for some $\varepsilon > 0$, the Alexandrov curvature of $(P_2(\mathbb{R}^d), W_2)$ is not bounded above by ε. Now set for $\lambda > 0$

$$\begin{aligned}
\mu_{0,\lambda} &= \frac{1}{2}(\delta_{(\lambda,\lambda)} + \delta_{(5\lambda,3\lambda)}),\\
\mu_{1,\lambda} &= \frac{1}{2}(\delta_{(-\lambda,\lambda)} + \delta_{(-5\lambda,3\lambda)}),\\
\mu_{2,\lambda} &= \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-4\lambda)}),\\
\mu_{1/2,\lambda} &= \frac{1}{2}(\delta_{(-2\lambda,2\lambda)} + \delta_{(2\lambda,2\lambda)}).
\end{aligned}$$

Then

$$\begin{aligned}
W_2^2(\mu_{0,\lambda}, \mu_{1,\lambda}) &= \frac{1}{2}((\lambda - (-6\lambda))^2 + (\lambda - 3\lambda)^2) + \frac{1}{2}((-\lambda - 5\lambda)^2 + (\lambda - 3\lambda)^2)\\
&= 40\lambda^2,\\
W_2^2(\mu_{0,\lambda}, \mu_{2,\lambda}) &= \frac{1}{2}((\lambda - 0)^2 + (\lambda - (-4\lambda))^2) + \frac{1}{2}((5\lambda - 0)^2 + (3\lambda - 0)^2)\\
&= 30\lambda^2,\\
W_2^2(\mu_{1,\lambda}, \mu_{2,\lambda}) &= \frac{1}{2}((-\lambda - 0)^2 + (\lambda - (-4\lambda))^2) + \frac{1}{2}((-5\lambda - 0)^2 + (3\lambda - 0)^2)\\
&= 30\lambda^2,\\
W_2^2(\mu_{1/2,\lambda}, \mu_{2,\lambda}) &= \frac{1}{2}((-2\lambda - 0)^2 + (2\lambda - 0)^2) + \frac{1}{2}((2\lambda - 0)^2 + (2\lambda - (-4\lambda))^2)\\
&= 24\lambda^2,\\
d^2(\tilde{\mu}_{1/2,\lambda}, \tilde{\mu}_{2,\lambda}) &= (\sqrt{30\lambda^2})^2 - (\sqrt{40\lambda^2}/2)^2 = 20\lambda^2,
\end{aligned}$$

and the $\mu_{\bullet,\lambda}$ show that the Alexandrov curvature of $(P_2(\mathbb{R}^d), W_2)$ is not bounded above by $\varepsilon\lambda^{-2}$. Sending $\lambda \to 0$ proves the Alexandrov curvature of $(P_2(\mathbb{R}^d), W_2)$ is unbounded from above.

For Riemannian M, choose an origin point $p \in M$ and define the μ as above, replacing the points in $\mathbb{R}^2$ by their image under the exponential map at p. As $\lambda \to 0$, the distances approach the corresponding Euclidean distances. Thus for some $\varepsilon_0 > 0$ and all small enough λ, the $\mu_{\bullet,\lambda}$ show that the Alexandrov curvature of $(P_2(M), W_2)$

is not bounded above by $(\varepsilon + \varepsilon_0)\lambda^{-2}$, so the Alexandrov curvature of $(P_2(M), W_2)$ is unbounded from above.

Now suppose we are given $\nu \in P_2(M)$ and $r > 0$, and for each $\alpha \in (0, 1]$ and each μ above, define

$$\mu^\alpha = (1 - \alpha)\nu + \alpha\mu.$$

By the restriction property (Theorem 4.6 of [Vil09]),

$$W_2(\nu, \mu^\alpha) = \alpha W_2(\nu, \mu)$$

and for small enough α, each μ^α is in $B(\nu, r)$. Similarly,

$$W_2(\mu^\alpha_{0,\lambda}, \mu^\alpha_{1,\lambda}) = \alpha W_2(\mu_{0,\lambda}, \mu_{1,\lambda}), \ldots, W_2(\mu^\alpha_{1/2,\lambda}, \mu^\alpha_{2,\lambda}) = \alpha W_2(\mu_{1/2,\lambda}, \mu_{2,\lambda}),$$

so again for small enough λ, the $\mu^\alpha_{\bullet,\lambda}$ show that the Alexandrov curvature of $B(\nu, r)$ is not bounded above by $(\varepsilon+\varepsilon_0)\lambda^{-2}\alpha^{-2}$, and hence the Alexandrov curvature of $B(\nu, r)$ is unbounded from above.

Note that the above argument also works with only minor alteration in the setting of n-dimensional Alexandrov spaces of curvature bounded below. One must simply use Theorem 2.1.20 to obtain a nearby regular point and apply Lemma 2.1.19. Thus we have shown:

Proposition 4.2.3. *If $\kappa \in \mathbb{R}$, $\mathcal{X}$ is a Riemannian manifold or finite-dimensional Alexandrov space of curvature bounded below,* $\dim \mathcal{X} \geq 2$ *and U is open in* $(P_2(\mathcal{X}), W_2)$*, then U is not an Alexandrov space of curvature bounded above by κ.*

Furthermore, approximating the μ above by $\mu_k \in (P_2^{ac}(\mathcal{X}), W_2)$ yields:

Corollary 4.2.4. *If $\kappa \in \mathbb{R}$, $\mathcal{X}$ is a Riemannian manifold or finite-dimensional Alexandrov space of curvature bounded below,* $\dim \mathcal{X} \geq 2$ *and U is open in $P_2^{ac}(\mathcal{X})$, then U is not an Alexandrov space of curvature bounded above by κ.*

It is therefore necessary to make fairly strong assumptions on the probability space $\mathcal{P} \subset P(\mathcal{X})$ in order to have nonpositive curvature.

4.3 Gaussian Steiner problems in $(P_2(\mathbb{R}^d), W_2)$

We now consider Steiner problems in $(P_2(\mathbb{R}^d), W_2)$ where the boundary data $\mu_1, \ldots, \mu_k$ consists of Gaussian measures. This will allow us to work in a nonpositively curved probability space, as well as give Steiner vertex degree results for a class of non-compactly supported boundary data.

The Gaussian measures on $\mathbb{R}^d$ are given by

$$\gamma_{m,V}\, dx = \left(\frac{1}{2\pi}\right)^{d/2} \frac{1}{\sqrt{\det V}} \exp\left[-\frac{1}{2}\langle x - m, V^{-1}(x - m)\rangle\right] dx$$

for any $m \in \mathbb{R}^d$ and and $d \times d$ symmetric positive definite matrix V. Here

$$E[\gamma_{m,V}] = m, \text{ and } \operatorname{Cov}(\gamma_{m,V}) = V,$$

as defined below.

We now define the operators E and Cov. One may consult [Dud89] for the relevant details. Suppose we fix some Gaussian measure μ_0 on $\mathbb{R}^d$. Then for any vector-valued random variable $F : R^d \to R^d$ in $L^2(\mathbb{R}^d, \mu_0)$, the law of F is $\nu = F_{\#}\mu_0$. Given $\nu \in P_2(\mathbb{R}^d)$, one can construct F such that $\operatorname{law}(F) = \nu$, again with respect to μ_0. Let the components of F be labeled $F = (f_1, f_2, ..., f_d)$. We define the expectation vectors $E(\nu)$ and $E(F)$ by

$$E(\nu) := E(F) := (E(f_1), E(f_2), ..., E(f_d)),$$

where $E(f_i) = \int_{\mathbb{R}^d} f_i(x)\, d\mu_0$ is the usual expected value of the random variable f_i.

These integrals are finite as $F \in L^2(\mathbb{R}^d, \mu_0)$ implies $F \in L^1(\mathbb{R}^d, \mu_0)$ since $\mu_0(\mathbb{R}^d) = 1$.

For random variables $F, G : \mathbb{R}^d \to \mathbb{R}^d$, $\mathrm{Cov}(F, G)$ denotes the matrix whose (i, j) entry is

$$\mathrm{Cov}(f_i, g_j) = E(f_i g_j) - E(f_i)E(g_j).$$

Assuming again that $\mathrm{law}(F) = \nu$, we define the matrix of covariances for ν by

$$\mathrm{Cov}(\nu) := \mathrm{Cov}(F) := \mathrm{Cov}(F, F),$$

which reduces down to the variance $\mathrm{Var}(\nu)$ if $d = 1$. One may check that $E(\nu)$ and $\mathrm{Cov}(\nu)$ are independent of the choice background measure μ_0 and the choice of random vector F. However, we will have to work with a specific choice of random vectors below in order to work with $\mathrm{Cov}(F, G)$; one can have $\mathrm{law}(F_1) = \mathrm{law}(F_2)$ and $\mathrm{law}(G_1) = \mathrm{law}(G_2)$ but still have $\mathrm{Cov}(F_1, G_1) \neq \mathrm{Cov}(F_2, G_2)$. For example, if f and g are i.i.d. random variables with mean 0 and variance 1, then $\mathrm{law}(f) = \mathrm{law}(g)$ but $\mathrm{Cov}(f, f) = 1$ while $\mathrm{Cov}(f, g) = 0$. We therefore cannot define the matrix $\mathrm{Cov}(\mu, \nu)$ unambiguously.

We denote the set of Gaussian measures by $\Gamma(\mathbb{R}^d)$, and given a $d \times d$ orthogonal matrix P we denote

$$\Gamma(\mathbb{R}^d, P) = \{\gamma_{m,V} | V \text{ is diagonalized by } P\}.$$

Following formal arguments of Otto [Ott01], Takatsu [Tak08] has shown that $\Gamma(\mathbb{R}^d)$ and $\Gamma(\mathbb{R}^d, P)$ are geodesically convex in $(P_2(\mathbb{R}^d), W_2)$, with the W_2-metric making $\Gamma(\mathbb{R}^d)$ a $d + d(d+1)/2$-dimensional Riemannian manifold and making $\Gamma(\mathbb{R}^d, P)$ isometric to $\mathbb{R}^d \times (0, +\infty)^d$. In particular,

$$W_2^2(\gamma_{m,V}, \gamma_{n,U}) = |m - n|^2 + \operatorname{tr} V + \operatorname{tr} U - 2 \operatorname{tr} \sqrt{U^{1/2} V U^{1/2}}. \tag{4.1}$$

We would like to solve the parameterized and general Monge-Kantorovich-Steiner problems in the space $(P_2(\mathbb{R}^d), W_2)$ for boundary data in $\Gamma(\mathbb{R}^d)$ and show that the solution stays in $\Gamma(\mathbb{R}^d)$. Our plan is to project

$$\varphi : \mu \to \gamma_{E[\mu],\mathrm{Cov}(\mu)},$$

assuming $\mathrm{Cov}(\mu) > 0$. The following lemma shows this is distance nonincreasing.

Lemma 4.3.1. *For $\mu, \nu \in P_2(\mathbb{R}^d)$,*

$$W_2^2(\mu, \nu) \geq |E(\mu) - E(\nu)|^2 + \operatorname{tr} U + \operatorname{tr} V - 2 \operatorname{tr} \sqrt{U^{1/2} V U^{1/2}},$$

where $U = \mathrm{Cov}(\mu)$ and $V = \mathrm{Cov}(\nu)$.

Proof. We begin by showing that for random variables $X, Y : \Omega \to \mathbb{R}^d$,

$$\operatorname{tr} \mathrm{Cov}(X, Y) \leq \operatorname{tr} \sqrt{(\mathrm{Cov}(X))^{1/2} \mathrm{Cov}(Y) (\mathrm{Cov}(X))^{1/2}}. \tag{4.2}$$

Since, $\operatorname{tr} A^2 \leq (\operatorname{tr} A)^2$, it suffices to show

$$\operatorname{tr} \mathrm{Cov}(X, Y) \leq \sqrt{\operatorname{tr} (\mathrm{Cov}(X))^{1/2} \mathrm{Cov}(Y) (\mathrm{Cov}(X))^{1/2}}.$$

Traces are independent of choice of orthonormal basis and $\mathrm{Cov}(X)$ is real and symmetric, so choose a diagonalizing orthonormal basis for $\mathrm{Cov}(X)$. In this basis, we may write $\mathrm{Cov}(X)$ as the diagonal matrix with entries λ_i on the diagonal and $\mathrm{Cov}(Y) = (b_{ij})$. Then $(\mathrm{Cov}(X))^{1/2}$ has entries $\sqrt{\lambda_i}$,

$$(\mathrm{Cov}(X))^{1/2} \mathrm{Cov}(Y) (\mathrm{Cov}(X))^{1/2} = (\sqrt{\lambda_i \lambda_j} b_{ij}),$$

and

$$\sqrt{\operatorname{tr}(\operatorname{Cov}(X))^{1/2}\operatorname{Cov}(Y)(\operatorname{Cov}(X))^{1/2}} = \sqrt{\sum_{i=1}^{d}\lambda_i b_{ii}}.$$

Now

$$\operatorname{tr}\operatorname{Cov}(X,Y) = \sum_{i=1}^{d}\operatorname{Cov}(X_i,Y_i) \le \sum_{i=1}^{d}\sqrt{\operatorname{Var}(X_i)\operatorname{Var}(Y_i)}$$

since the correlation between X_i and Y_i is at most 1. (Note that if either variation is zero, then $\operatorname{Cov}(X_i,Y_i) = 0 = \sqrt{\operatorname{Var}(X_i)\operatorname{Var}(Y_i)}$.) Furthermore,

$$\sum_{i=1}^{d}\sqrt{\operatorname{Var}(X_i)\operatorname{Var}(Y_i)} = \sum_{i=1}^{d}\sqrt{\lambda_i b_{ii}} \le \sqrt{\sum_{i=1}^{d}\lambda_i b_{ii}},$$

so we have proven equation (4.2).

We also have

$$\begin{aligned}\operatorname{tr}\operatorname{Cov}(X,Y) &= \sum_{i=1}^{d}\operatorname{Cov}(X_i,Y_i)\\ &= \sum_{i=1}^{d}E[X_iY_i] - E[X_i]E[Y_i]\\ &= E[X\cdot Y] - E[X]\cdot E[Y],\end{aligned}$$

so

$$\operatorname{tr}\operatorname{Cov}(X,Y) = E[X\cdot Y] - E[X]\cdot E[Y]. \tag{4.3}$$

In particular,

$$\operatorname{tr}\operatorname{Cov}(X) = E[X\cdot X] - E[X]\cdot E[X]. \tag{4.4}$$

We may reformulate the Wasserstein distance as

$$W_2^2(\mu,\nu) = \inf E[|X-Y|^2],$$

where the infimum is taken over all random variables $X, Y : \mathbb{R}^{2d} \to \mathbb{R}^d$ with laws μ and ν respectively. For any such X, Y,

$$\begin{aligned}
E[|X-Y|^2] &= E[X^2] + E[Y^2] - 2E[X \cdot Y] \\
&= E[X^2] + E[Y^2] - 2E[X] \cdot E[Y] - 2(E[X \cdot Y] - E[X] \cdot E[Y]) \\
&\overset{(4.3)}{=} E[X^2] + E[Y^2] - 2E[X] \cdot E[Y] - 2 \operatorname{tr} \operatorname{Cov}(X,Y) \\
&\overset{(4.2)}{\geq} E[X^2] + E[Y^2] - 2E[X] \cdot E[Y] \\
&\qquad - 2 \operatorname{tr} \sqrt{(\operatorname{Cov}(X))^{1/2} \operatorname{Cov}(Y) (\operatorname{Cov}(X))^{1/2}} \\
&= (E[X] - E[Y])^2 + E[X^2] - E[X] \cdot E[X] + E[Y^2] \\
&\qquad - E[Y] \cdot E[Y] - 2 \operatorname{tr} \sqrt{(\operatorname{Cov}(X))^{1/2} \operatorname{Cov}(Y) (\operatorname{Cov}(X))^{1/2}} \\
&\overset{(4.4)}{=} |E[X] - E[Y]|^2 + \operatorname{tr} \operatorname{Cov}(X) + \operatorname{tr} \operatorname{Cov}(Y) \\
&\qquad - 2 \operatorname{tr} \sqrt{(\operatorname{Cov}(X))^{1/2} \operatorname{Cov}(Y) (\operatorname{Cov}(X))^{1/2}} \\
&= |E(\mu) - E(\nu)|^2 + \operatorname{tr} U + \operatorname{tr} V - 2 \operatorname{tr} \sqrt{U^{1/2} V U^{1/2}}.
\end{aligned}$$

Since the inequality holds for any X, Y, it holds for the infimum as well. $\square$

If we define

$$\mathcal{P}(\mathbb{R}^d) = \{\mu \in P_2(\mathbb{R}^d) | \operatorname{Cov}(\mu) > 0\}$$

and

$$\mathcal{P}(\mathbb{R}^d, P) = \{\mu \in P_2(\mathbb{R}^d) | \operatorname{Cov}(\mu) > 0, \operatorname{Cov}(\mu) \text{ is diagonalized by } P\},$$

then we immediately obtain solutions to Steiner problems by applying the projection map φ and the local compactness of $\Gamma(\mathbb{R}^d)$ and $\Gamma(\mathbb{R}^d, P)$.

Corollary 4.3.2. *Given boundary data* $\mu_1, \ldots, \mu_k$ *in* $\Gamma(\mathbb{R}^d)$*, the parameterized and general forms of the Monge-Kantorovich-Steiner problem, the Monge-Steiner problem and the Steiner problem in* $(\mathcal{P}(\mathbb{R}^d), W_2)$ *have solutions in* $\Gamma(\mathbb{R}^d)$*.*

Corollary 4.3.3. *Given boundary data* $\mu_1, \ldots, \mu_k$ *in* $\Gamma(\mathbb{R}^d, P)$*, the parameterized and general forms of the Monge-Kantorovich-Steiner problem, the Monge-Steiner problem and the Steiner problem in* $(\mathcal{P}(\mathbb{R}^d, P), W_2)$ *have solutions in* $\Gamma(\mathbb{R}^d, P)$*. The solutions are also Steiner graphs in* $\mathbb{R}^{2d}$*. In particular, for the general Steiner problems there is a Steiner solution which has a canonical representative* $\Phi : G \to P_2(\mathbb{R})$ *such that*

1. *G is a tree.*

2. *For any vertex v in $G \setminus \Phi^{-1}(\{\mu_1, \ldots, \mu_k\})$, the degree of v is three and all pairs of geodesics in $\Phi(G)$ meeting at $\Phi(v)$ do so with an angle of $2\pi/3$.*

The structural result above follows by the variational argument we used before, as $\Gamma(\mathbb{R}^d, P)$ is flat, hence nonnegatively curved.

For clarity of results, we will now concentrate on the case $d = 1$. Here we have that advantage that $\Gamma(\mathbb{R}) = \Gamma(\mathbb{R}, P) = \mathbb{R} \times (0, +\infty)$. Define the projection

$$\varphi : \mu \to \begin{cases} \gamma_{E[\mu],\mathrm{Var}(\mu)} & \text{if } \mathrm{Var}(\mu) > 0, \\ \delta_{E[\mu]} & \text{if } \mathrm{Var}(\mu) = 0. \end{cases},$$

and note that (4.1) still holds for this interpretation of $\gamma_{m,0}$. Thus Lemma 4.3.1 implies that $\varphi : P_2(\mathbb{R}) \to \mathbb{R} \times [0, +\infty)$ is a contraction mapping. Also, φ fixes $\Gamma(\mathbb{R})$. If we are given boundary data $\mu_1, \ldots, \mu_k$ in $\Gamma(\mathbb{R})$, then for some $\varepsilon > 0$, $\mathrm{Var}(\mu_i) > \varepsilon$ for all i, and we may retract $\mathbb{R} \times [0, \varepsilon]$ to $\mathbb{R} \times \{\varepsilon\}$ without increasing the network length in $\overline{\Gamma(\mathbb{R})} = \mathbb{R} \times [0, +\infty)$. We obtain in this manner Steiner solutions over the full space $P_2(\mathbb{R})$ which stay in $\Gamma(\mathbb{R})$.

Theorem 4.3.4. *Given boundary data $\mu_1, \ldots, \mu_k$ in $\Gamma(\mathbb{R})$, the parameterized and general forms of the Monge-Kantorovich-Steiner problem, the Monge-Steiner problem and the Steiner problem in $(P_2(\mathbb{R}), W_2)$ have solutions in $\Gamma(\mathbb{R})$. The solutions are also Steiner graphs in $\mathbb{R}^2$, and the Steiner ratio for this restricted class of boundary data is $\sqrt{3}/2$.*

Proof. The preceding discussion proves existence of a solution in $\Gamma(\mathbb{R})$, which may then be viewed as a solution to the corresponding planar Steiner problem. The Steiner ratio result then follows by the proof of the Gilbert-Pollak conjecture [DH90]. $\square$

The preceding Steiner ratio result cannot hold for arbitrary boundary data.

Proposition 4.3.5. *If M is a Riemannian manifold of nonnegative curvature, then*

$$\rho(P_2(M)) \leq \lim_{d\to\infty} \rho(\mathbb{R}^d) \leq \frac{\sqrt{3/2}}{2^{3/2}-1}.$$

Proof. It suffices to find for any $d \in \mathbb{N}$ and any $\varepsilon > 0$ a finite set $S \subset P_2(M)$ such that

$$\frac{L_s(S)}{L_a(S)} \leq \rho(\mathbb{R}^d) + \varepsilon.$$

By definition there is a finite set $S' \subset \mathbb{R}^d$ such that

$$\frac{L_s(S')}{L_a(S')} \leq \rho(\mathbb{R}^d) + \frac{\varepsilon}{3}.$$

Choose a compact $N \subset M$ and some $\mu \in P_2^{ac}(M)$ with support in N. Then by Theorem 4.1.2, the tangent cone K_μ of $P(N)$ at μ is an infinite-dimensional Hilbert space, so we may choose an orthonormal set $\nu_1, \ldots, \nu_d \in K_\mu$. In this way, we obtain a subset of K_μ isometric to $\mathbb{R}^d$. Let $S_1 \subset K_\mu$ be the finite set corresponding to S'. We may approximate S_1 by a finite set S_2 such that each point is on a ray corresponding to a direction of some geodesic in $P_2(N)$, and

$$\frac{L_s(S_2)}{L_a(S_2)} \leq \frac{L_s(S_1)}{L_a(S_1)} + \frac{\varepsilon}{3} = \frac{L_s(S')}{L_a(S')} + \frac{\varepsilon}{3}.$$

We may rescale by λ about the origin to obtain λS_2 with

$$\frac{L_s(\lambda S_2)}{L_a(\lambda S_2)} = \frac{L_s(S_2)}{L_a(S_2)},$$

and for small enough λ the exponential map will be well-defined on λS_2. Since

$$\left| \frac{L_s(\lambda S_2)}{L_a(\lambda S_2)} - \frac{L_s(\exp(\lambda S_2))}{L_a(\exp(\lambda S_2))} \right| \to 0 \text{ as } \lambda \to 0,$$

for small enough λ we have

$$\frac{L_s(\exp(\lambda S_2))}{L_a(\exp(\lambda S_2))} \leq \frac{L_s(\lambda S_2)}{L_a(\lambda S_2)} + \frac{\varepsilon}{3} \leq \rho(\mathbb{R}^d) + \varepsilon.$$

Sending $\varepsilon \to 0$ and applying the bound of Chung and Gilbert [CG76] gives the desired result. $\square$

Appendix A

Alexandrov curvature of convex hypersurfaces in Hilbert space

In this appendix, the following result is established:

Theorem A.0.6. *If C is an open set in a Hilbert space H and $\overline{C}$ is locally convex, then ∂C is a nonnegatively curved Alexandrov space under the induced length metric.*

Questions of this sort go back to [Ale48], where Alexandrov defined Alexandrov curvature and showed that it characterizes boundaries of locally convex bodies in $\mathbb{R}^3$. This was generalized by Buyalo to the case of locally convex sets of full dimension in a Riemannian manifold in [Buy76]. If the ambient manifold has a positive lower bound κ on sectional curvature, it has also been shown in [AKP08] that the convex boundary has Alexandrov curvature $\geq \kappa$.

The proof of Theorem A.0.6 relies on approximating ∂C by smooth manifolds, where the connection between curvature and convexity is well understood. Due to the possibly infinite dimension of H, we cannot smooth by integrating over H against a mollifier. As currently known smoothing operators for infinite dimensional spaces do not preserve convexity, we proceed by integrating over a suitably chosen finite dimensional subspace. Lemma A.1.4 shows this can be done in such a way that the curvature of ∂C is controlled by the curvature of smooth, finite-dimensional approximating manifolds. A similar approximation of infinite-dimensional curvature by finite dimensional curvature is outlined in [Hal00].

A.1 Approximation by smooth manifolds

In this section, we prove two technical lemmas which allow us to approximate $C^{1,1}$ convex functions f on a Hilbert space by convex functions that are smooth on a finite-dimensional linear subspace. This enables us to control the Alexandrov curvature of graph_f, the graph of f in $H \times \mathbb{R}$, via the sectional curvature of the approximating smooth graphs.

It will be helpful to fix notation for polygonal paths.

Definition A.1.1. *For two points p, q in a vector space V, $\sigma_{pq} : [0,1] \to V$ denotes the constant speed linear path:*

$$\sigma_{pq}(t) = (1-t)p + tq.$$

Definition A.1.2. *A path $\tau : [0,1] \to V$ is called a polygonal path if it can be written in the form*

$$\tau(t) = \sum_{i=1}^{k-1} \sigma_{p_i p_{i+1}}(kt - i) 1_{[i/k,(i+1)/k]}(t)$$

for some set of points $p_1, \ldots, p_k \in V$. Here 1_A denotes the characteristic function of the set A. If S is a set of linear subspaces of V and $p_i - p_j$ lies in an element of S for all i and j, then τ is called S-polygonal.

Lemma A.1.3. *Let $f : V \to (X, d)$ be a λ-bi-Lipschitz map from a Banach space V onto a metric space (X, d). For any rectifiable curve $\sigma : [0,1] \to X$ and any $\varepsilon > 0$, there exists a polygonal path $\tau : [0,1] \to V$ such that $f \circ \tau(0) = \sigma(0)$, $f \circ \tau(1) = \sigma(1)$, $\forall t \in [0,1]$, $|\sigma(t) - f \circ \tau(t)| < \varepsilon$ and $|l(\sigma) - l(f \circ \tau)| < \varepsilon$.*

Proof. For each rectifiable curve $\sigma_0 : [0,1] \to X$ and $\varepsilon > 0$, define

$$B^1_\varepsilon(\sigma_0) = \{\sigma : [0,1] \to X; \forall t \in [0,1], d(\sigma_0(t), \sigma(t)) < \varepsilon, |l(\sigma_0) - l(\sigma)| < \varepsilon\}.$$

For each rectifiable curve $\sigma_0 : [0,1] \to V$ and $\varepsilon > 0$, define

$$B^2_\varepsilon(\sigma_0) = \{\sigma : [0,1] \to V; \forall t \in [0,1], |\sigma_0(t) - \sigma(t)| < \varepsilon, |l(\sigma_0) - l(\sigma)| < \varepsilon\}.$$

Fix a rectifiable curve $\sigma_0 : [0,1] \to V$ and $\varepsilon > 0$. For any $\sigma \in B^2_\varepsilon(\sigma_0)$, for all $t \in [0,1]$,

$$|\sigma_0(t) - \sigma(t)| < \varepsilon \implies |f \circ \sigma_0(t) - f \circ \sigma(t)| < \lambda\varepsilon.$$

Furthermore,

$$\begin{aligned}
|l(f \circ \sigma_0) - l(f \circ \sigma)| &\le |l(\sigma_0) - l(\sigma)| + |l(f \circ \sigma_0) - l(\sigma_0)| + |l(f \circ \sigma) - l(\sigma)| \\
&\le \varepsilon + l(f \circ \sigma_0) + l(\sigma_0) + l(f \circ \sigma) + l(\sigma) \\
&\le \varepsilon + \lambda l(\sigma_0) + l(\sigma_0) + \lambda l(\sigma) + l(\sigma) \\
&\le \varepsilon + (\lambda + 1) l(\sigma_0) + (\lambda + 1)(l(\sigma_0) + \varepsilon) \\
&\le 2(\lambda + 1)(\varepsilon + l(\sigma_0)).
\end{aligned}$$

So for $\varepsilon' = 2(\lambda + 1)(\varepsilon + l(\sigma_0))$,

$$B^2_\varepsilon(\sigma_0) \subset f^{-1}(B^1_{\varepsilon'}(f \circ \sigma_0))$$

By a similar argument, for any rectifiable curve $\sigma_0 : [0,1] \to X$ and $\varepsilon > 0$,

$$f^{-1}(B^1_\varepsilon(\sigma_0)) \subset B^2_{\varepsilon'}(f^{-1} \circ \sigma_0),$$

for $\varepsilon' = 2(\lambda + 1)(\varepsilon + l(\sigma_0))$. Thus the B^2's and $f^{-1}(B^1)$'s determine equivalent topologies on the space of rectifiable curves $\sigma : [0,1] \to V$. Polygonal paths are dense under the B^2-topology, so they are dense under the $f^{-1}(B^1)$-topology. $\square$

Lemma A.1.4. *Let $f : \Omega \to \mathbb{R}$ be a $C^{1,1}$ convex function, where Ω is a domain in a Hilbert space H. For any $x_0 \in \Omega$, there exists $R > 0$ such that Y, the graph of f over $B_R(x_0)$, satisfies the quadruple condition*

$$\tilde{\angle} bac + \tilde{\angle} cap + \tilde{\angle} pab \leq 2\pi$$

for any quadruple $(a; b, c, p)$ of distinct points, under the induced length metric d from $H \times \mathbb{R}$.

Proof. f is convex, hence Lipschitz continuous on a neighborhood Ω' of x_0 for some Lipschitz constant $L \geq 1$ (Proposition 2.3 of [BT05]). Let $\hat{f} : \Omega' \to \hat{f}(\Omega') \subset \mathrm{graph}_f$ be defined by $\hat{f}(x) = (x, f(x))$, and note that $\hat{f}$ is $\sqrt{1+L^2}$-bi-Lipschitz. Choose $R > 0$ such that $B_{3R}(x_0) \subset \Omega'$. Suppose that $(a; b, c, p)$ is a quadruple of distinct points such that

$$\tilde{\angle} bac + \tilde{\angle} cap + \tilde{\angle} pab = 2\pi + \varepsilon_0 > 2\pi,$$

where $(a; b, c, p) = (\hat{f}(a'); \hat{f}(b'), \hat{f}(c'), \hat{f}(p'))$ and $a', b', c', p' \in B_R(x_0)$. The comparison angles vary continuously in the intrinsic distances, so there exists $\varepsilon > 0$ such that if $(A; B, C, D)$ is a quadruple of points in some other metric space (X_1, d_1) with

$$|d(a,b) - d_1(A,B)| < \varepsilon, \quad |d(a,c) - d_1(A,C)| < \varepsilon, \quad |d(a,p) - d_1(A,P)| < \varepsilon,$$
$$|d(b,c) - d_1(B,C)| < \varepsilon, \quad |d(b.p) - d_1(B,P)| < \varepsilon, \quad |d(c,p) - d_1(C,P)| < \varepsilon,$$

then

$$\tilde{\angle} BAC + \tilde{\angle} CAP + \tilde{\angle} PAB = 2\pi + (\varepsilon_0/2) > 2\pi.$$

By Lemma A.1.3, we may approximate $d(a,b)$ by the length of the image under $\hat{f}$

of a polygonal path τ_1 determined by points $a' = q_1, q_2, \ldots, q_{k_1-1}, b' = q_{k_1} \in B_{2R}(x_0)$ such that

$$d(a,b) + (\varepsilon/3) \geq \sum_{i=1}^{k_1-1} l(\hat{f} \circ \sigma_{q_i q_{i+1}}) = l(\hat{f} \circ \tau_1) \geq d(a,b).$$

Similarly, we may approximate $d(a,c)$ by the image under $\hat{f}$ of a polygonal path determined by points $a' = q_{k_1+1}, q_{k_1+2}, \ldots, c' = q_{k_2} \in B_{2R}(x_0)$ such that

$$d(a,c) + (\varepsilon/3) \geq \sum_{i=k_1+1}^{k_2-1} l(\hat{f} \circ \sigma_{q_i q_{i+1}}) \geq d(a,c).$$

Continue in this manner choosing $q_{k_2+1}, q_{k_2+2} \ldots, q_{k_3}, \ldots, q_{k_6}$ to approximate the remaining four intrinsic distances.

The $k_6 + 1$ points $q_1, \ldots, q_{k_6}, x_0$ lie in a k_6-dimensional subspace of H, which we will identify as $\mathbb{R}^n$, $n = k_6$, with $H = \mathbb{R}^n \times \tilde{H}$. Let $\varphi_\delta : \mathbb{R}^n \to \mathbb{R}$ be the standard C^∞ mollifier supported on the δ-ball, and define $f_\delta : B_{5R/2}(x_0) \to \mathbb{R}$ by $f_\delta = f * \varphi_\delta$, where the convolution occurs in the $\mathbb{R}^n$-variables and $\delta < R/2$. Let $\hat{f}_\delta(x) = (x, f_\delta(x))$. As f is assumed to be convex and $C^{1,1}$, it is easy to check the following properties:

1. $f_\delta|_{B_{2LR}(x_0)\cap(\mathbb{R}^n\times\{y\})}$ is C^∞ for every $y \in \tilde{H}$.

2. $f_\delta|_{B_{2LR}(x_0)\cap(\mathbb{R}^n\times\{y\})}$ is L-Lipschitz for every $y \in \tilde{H}$.

3. $f_\delta \to f$ pointwise as $\delta \to 0$.

4. For every $y \in \tilde{H}$, on $B_{2LR}(x_0) \cap (\mathbb{R}^n \times \{y\})$, $\nabla_{\mathbb{R}^n} f_\delta \to \nabla_{\mathbb{R}^n} f$ uniformly as $\delta \to 0$.

5. For every line $\Lambda \subset H$, on $B_{2LR}(x_0) \cap \Lambda$, $\nabla_\Lambda f_\delta \to \nabla_\Lambda f$ uniformly as $\delta \to 0$.

6. For every polygonal path $\sigma : [0,1] \to B_{2R}(x_0)$, $l(\hat{f}_\delta \circ \sigma) \to l(\hat{f} \circ \sigma)$. This convergence is uniform on sets of the form

$$\{\sigma : [0,1] \to B_{2R}(x_0); l(\sigma) < C, \sigma \text{ is } S\text{-polygonal}\},$$

where $C \in \mathbb{R}$ and S is the union of the set of linear subspaces of $\mathbb{R}^n$ and a finite set of linear subspaces of H.

7. f_δ is convex.

The first four properties are statements about convolution of a function of n variables against φ_δ which are well known. For the fifth and seventh properties, one fixes a line $\Lambda \subset H$ and again applies standard results for convolution against φ_δ in $\mathbb{R}^n$, or in $\mathbb{R}^{n+1}$ if necessary. Finally, the sixth property is a direct corollary of the previous two properties.

Let Y_δ denote the graph of f_δ over $B_{2R}(x_0)$ with metric d_δ induced by $H \times \mathbb{R}$, and let $Y_{\delta,n}$ denote the graph of f_δ over $B_{2R}(x_0) \cap (\mathbb{R}^n \times \{0\})$ with metric $d_{\delta,n}$ induced by $(\mathbb{R}^n \times \{0\}) \times \mathbb{R}$. Note that $f_\delta|_{B_{2R}(x_0)\cap(\mathbb{R}^n\times\{0\})}$ is a C^∞ convex function over a domain in $(\mathbb{R}^n \times \{0\})$, so $Y_{\delta,n}$ is a Riemannian manifold of nonnegative sectional curvature. In particular, it satisfies the quadruple condition. We will obtain a contradiction by showing

$$\begin{aligned}
&|d(a,b) - d_{\delta,n}(\hat f_\delta(a'), \hat f_\delta(b'))| < \varepsilon, \quad |d(a,c) - d_{\delta,n}(\hat f_\delta(a'), \hat f_\delta(c'))| < \varepsilon,\\
&|d(a,p) - d_{\delta,n}(\hat f_\delta(a'), \hat f_\delta(p'))| < \varepsilon, \quad |d(b,c) - d_{\delta,n}(\hat f_\delta(b'), \hat f_\delta(c'))| < \varepsilon,\\
&|d(b,p) - d_{\delta,n}(\hat f_\delta(b'), \hat f_\delta(p'))| < \varepsilon, \quad |d(c,p) - d_{\delta,n}(\hat f_\delta(c'), \hat f_\delta(p'))| < \varepsilon.
\end{aligned}$$

Let $C = d(a,b) + d(a,c) + \cdots + d(c,p) + \varepsilon$. Recall that τ_1 is the polygonal path determined by $q_1, \ldots, q_{k_1}$.

$$l(\tau_1) \le l(\hat f \circ \tau_1) = \sum_{i=1}^{k_1-1} l(\hat f \circ \sigma_{q_i q_{i+1}}) \le d(a,b) + (\varepsilon/3) < C,$$

so choosing δ_0 small with respect to C and $S = \mathbb{R}^n$, we have $l(\hat f_\delta \circ \tau_1) \le l(\hat f \circ \tau_1) + (\varepsilon/3)$

for $\delta < \delta_0$. $\hat{f}_\delta \circ \tau_1 : [0,1] \to Y_{\delta,n}$ is a path from $\hat{f}_\delta(a')$ to $\hat{f}_\delta(b')$, so

$$d_{\delta,n}(\hat{f}_\delta(a'), \hat{f}_\delta(b')) \leq l(\hat{f}_\delta \circ \tau_1) \leq l(\hat{f} \circ \tau_1) + (\varepsilon/3) \leq d(a,b) + (2\varepsilon/3).$$

Applying Lemma A.1.3 again, choose a polygonal path $\tau_2 : [0,1] \to B_{2R}(x_0) \cap (\mathbb{R}^n \times \{0\})$ such that

$$d_{\delta,n}(\hat{f}_\delta(a'), \hat{f}_\delta(b')) \geq l(\hat{f}_\delta \circ \tau_2) - (\varepsilon/6).$$

Note that

$$l(\tau_2) \leq l(\hat{f}_\delta \circ \tau_2) \leq d_{\delta,n}(\hat{f}_\delta(a'), \hat{f}_\delta(b')) + (\varepsilon/6) \leq d(a,b) + (5\varepsilon/6) < C.$$

Expand S by including each linear subspace of H parallel to a segment of τ_2, and decrease δ_0 if necessary for this new S. For $\delta < \delta_0$,

$$l(\hat{f}_\delta \circ \tau_2) \geq l(\hat{f} \circ \tau_2) - (\varepsilon/3),$$

so

$$d_{\delta,n}(\hat{f}_\delta(a'), \hat{f}_\delta(b')) > l(\hat{f} \circ \tau_2) - \varepsilon \geq d(a,b) - \varepsilon.$$

The remaining inequalities follow in a similar manner, for the same choice of C and δ_0 chosen for an augmentation of $\mathbb{R}^n$ by a finite number of linear subspaces of H. So for $\delta < \delta_0$, the quadruple $(\hat{f}_\delta(a'); \hat{f}_\delta(b'), \hat{f}_\delta(c'), \hat{f}_\delta(p'))$ violates the quadruple condition in the Riemannian manifold of nonnegative sectional curvature $Y_{\delta,n}$. Therefore our original assumption is false and Y satisfies the quadruple condition. $\square$

A.2 Proof of Theorem A.0.6

Proof of Theorem A.0.6. We must prove the quadruple condition holds in a neighborhood of every $x_0 \in \partial C$. Let $C' = B_{2\rho}(x_0) \cap C$, where ρ is chosen small enough to make C' convex. Note that the intrinsic balls of radius ρ about x_0 are the same for C and C'. Choose a point $y \in C'$, and $r \in (0, \rho/2)$ such that $B_{2r}(y) \subset C'$. Let H' be the hyperplane through x_0 with normal vector $y - x_0$. For any $x \in H' \cap B_{2r}(x_0)$, let L_x be the line through x spanned by $y - x_0$. $L_x \cap C'$ is convex and C' is open and bounded, so $L_x \cap C'$ is a bounded interval. $x + (y - x_0) \in L_x \cap C'$, so $L_x \cap C' \neq \emptyset$. Considering $y - x_0$ as the upward direction, let $f(x)$ denote the $\mathbb{R}$-coordinate of the bottom endpoint of $L_x \cap C'$ in $H' \times \mathbb{R}$. $f : H' \cap B_{2r}(x_0) \to \mathbb{R}$ is then a convex function, as the epigraph is convex. Furthermore, the graph of f is a neighborhood of x_0 in $\partial C'$, and thus also in ∂C since $2r < \rho$.

f is convex, so taking r smaller if necessary, f is Lipschitz continuous for some Lipschitz constant $L \geq 1$. As shown in [LL86], for all small enough $\varepsilon > 0$, the inf-sup-convolution

$$g_\varepsilon(x) = \inf_{z \in H' \cap B_{2r}(x_0)} \sup_{y \in H' \cap B_{2r}(x_0)} \left[f(y) - \frac{\|y - z\|_H^2}{2\varepsilon} + \frac{\|x - z\|_H^2}{\varepsilon} \right]$$

is a $C^{1,1}$ convex function on $H' \cap B_r(x_0)$, g_ε is L-Lipschitz, and $g_\varepsilon \to f$ uniformly on $H' \cap B_r(x_0)$.

By Lemma A.1.4, the graph of g_ε over $H' \cap B_R(x_0)$ satisfies the quadruple condition for $R = r/3$. The graph of f over $H' \cap B_R(x_0)$ then satisfies the quadruple condition by continuity. □

Bibliography

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[AKP08] Stephanie Alexander, Vitali Kapovitch, and Anton Petrunin. An optimal lower curvature bound for convex hypersurfaces in Riemannian manifolds. *Illinois J. Math.*, 52(3):1031–1033, 2008.

[Ale48] A. D. Alexandrov. *Vnutrennyaya Geometriya Vypuklyh Poverhnosteĭ*. OGIZ, Moscow-Leningrad, 1948.

[AT04] Luigi Ambrosio and Paolo Tilli. *Topics on analysis in metric spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BGP92] Yu. Burago, M. Gromov, and G. Perel′man. A. D. Aleksandrov spaces with curvatures bounded below. *Uspekhi Mat. Nauk*, 47(2(284)):3–51, 222, 1992.

[BH99] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften*

[Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999.

[BT05] Frédéric Bernard and Lionel Thibault. Prox-regular functions in Hilbert spaces. *J. Math. Anal. Appl.*, 303(1):1–14, 2005.

[Buy76] S. V. Buyalo. Shortest paths on convex hypersurfaces of a Riemannian space. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 66:114–132, 207, 1976. Studies in topology, II.

[CG72] Jeff Cheeger and Detlef Gromoll. On the structure of complete manifolds of nonnegative curvature. *Ann. of Math. (2)*, 96:413–443, 1972.

[CG76] F. R. K. Chung and E. N. Gilbert. Steiner trees for the regular simplex. *Bull. Inst. Math. Acad. Sinica*, 4(2):313–325, 1976.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[DH90] D.-Z. Du and F. K. Hwang. An approach for proving lower bounds: solution of Gilbert-Pollak's conjecture on Steiner ratio. In *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)*, pages 76–85. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.

[DS96] Ding-Zhu Du and Warren D. Smith. Disproofs of generalized Gilbert-Pollak conjecture on the Steiner ratio in three or more dimensions. *J. Combin. Theory Ser. A*, 74(1):115–130, 1996.

[Dud89] Richard M. Dudley. *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[GGJ77] M. R. Garey, R. L. Graham, and D. S. Johnson. The complexity of computing Steiner minimal trees. *SIAM J. Appl. Math.*, 32(4):835–859, 1977.

[GP68] E. N. Gilbert and H. O. Pollak. Steiner minimal trees. *SIAM J. Appl. Math.*, 16:1–29, 1968.

[Hal00] Stephanie Halbeisen. On tangent cones of Alexandrov spaces with curvature bounded below. *Manuscripta Math.*, 103(2):169–182, 2000.

[IT94] Alexandr O. Ivanov and Alexei A. Tuzhilin. *Minimal networks*. CRC Press, Boca Raton, FL, 1994. The Steiner problem and its generalizations.

[Kan04a] L. V. Kantorovich. On a problem of Monge. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):15–16, 2004. Reprinted from *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 3 (1948), no. 2; translation in *J. Math. Sci. (N. Y.)* 133 (2006), no. 4, 1383.

[Kan04b] L. V. Kantorovich. On mass transportation. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 312(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11):11–14, 2004. Reprinted from *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 37 (1942), no. 7-8; translation in *J. Math. Sci. (N. Y.)* 133 (2006), no. 4, 1381–1382.

[LL86] J.-M. Lasry and P.-L. Lions. A remark on regularization in Hilbert spaces. *Israel J. Math.*, 55(3):257–266, 1986.

[LV09] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[Mil67] A. D. Milka. Metric structure of a certain class of spaces that contain straight lines. *Ukrain. Geometr. Sb. Vyp.*, 4:43–48, 1967.

[MY06] Chikako Mese and Sumio Yamada. The parameterized Steiner problem and the singular plateau Problem via energy. *Trans. Amer. Math. Soc.*, 358(7):2875–2895 (electronic), 2006.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[Per91] Grigori Perelman. Alexandrov's space with curvatures bounded from below II. Preprint, 1991.

[Pro56] Yu. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.*, 1:177–238, 1956.

[Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[Tak08] Asuka Takatsu. On Wasserstein geometry of the space of Gaussian measures, 2008.

[Top64] V. A. Toponogov. The metric structure of Riemannian spaces of non-negative curvature containing straight lines. *Sibirsk. Mat. Ž.*, 5:1358–1369, 1964.

[Vil09] Cédric Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

Vita

Jonathan Dahl was born on December 28, 1981, in Bridgeport, Connecticut. He received a Bachelor of Science in Mathematics in 2004 from the University of Maryland. He enrolled in Johns Hopkins University in the fall of 2004 and received a Master of Arts in Mathematics from Hopkins in 2006. He defended this thesis on March 3, 2010.