DISSERTATION

# VIETORIS–RIPS METRIC THICKENINGS AND WASSERSTEIN SPACES

Submitted by

Joshua Mirth

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: Henry Adams

Christopher Peterson
Amit Patel
Richard Eykholt

ABSTRACT

VIETORIS–RIPS METRIC THICKENINGS AND WASSERSTEIN SPACES

If the vertex set, $X$, of a simplicial complex, $K$, is a metric space, then $K$ can be interpreted as a subset of the Wasserstein space of probability measures on $X$. Such spaces are called simplicial metric thickenings, and a prominent example is the Vietoris–Rips metric thickening. In this work we study these spaces from three perspectives: metric geometry, optimal transport, and category theory. Using the geodesic structure of Wasserstein space we give a novel proof of Hausmann's theorem for Vietoris–Rips metric thickenings. We also prove the first Morse lemma in Wasserstein space and relate it to the geodesic perspective. Finally we study the category of simplicial metric thickenings and determine effects of certain limits and colimits on homotopy type.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

# LIST OF FIGURES

DISSERTATION

VIETORIS–RIPS METRIC THICKENINGS AND WASSERSTEIN SPACES


Submitted by

Joshua Mirth

Department of Mathematics


In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: Henry Adams

Christopher Peterson
Amit Patel
Richard Eykholt

# Chapter 1

# Introduction

Simplicial metric thickenings are, roughly speaking, simplicial complexes which can be viewed as subspaces of the Wasserstein space of probability measures. The purpose of this thesis is to expound upon the Wasserstein space to the applied topology community, and to answer questions about the topology of simplicial metric thickenings, in particular the Vietoris–Rips metric thickening. Three different perspectives come together in this work: metric geometry, optimal transport, and category theory; each of which contribute to our understanding of simplicial metric thickenings in different ways.

Chapter 2 begins with an overview of a longstanding problem: what is the homotopy type of the Vietoris–Rips complex of a given metric space? This question, in some form, dates to the work of Gromov in the late 1980s [1] and of Hausmann in the early 1990s [2]. It has captured increasing attention in recent years due to its relevance in applied topology. Here the focus is on the related question of the homotopy type of Vietoris–Rips metric thickenings, which were introduced in [3] and are defined in Section 2.3. These also provide the primary example of a simplicial metric thickening.

Key to the definition of a simplicial metric thickening is the Wasserstein metric on probability measures. The history of the Wasserstein metric and its origins in optimal transport theory are the subject of Section 3.1. Analytical details of the Wasserstein space $\mathscr{P}(X)$ are provided in Section 3.2 and Section 3.3.

Chapter 4 begins our investigation of the homotopy type of Vietoris–Rips metric thickenings in earnest. Of great importance to the homotopy type of the Vietoris–Rips metric thickening, $\mathscr{V}\mathscr{R}(X;r)$, is the curvature of the space $X$ and the existence of a center of mass of probability distributions. Section 4.2 considers this in detail. Section 4.3 then proves versions of Hausmann's theorem in several different curvature settings. These results are similar to the main theorem in [3], but the proofs are original and emphasize the geometry of Wasserstein space. Original corollaries dependent upon our proof technique are given in Section 4.4.

Wasserstein space, $\mathscr{P}(X)$, inherits to some extent whatever infinitesimal structure the base space, $X$, possesses. This motivates Chapter 5, which describes the differential structure of Wasserstein space, and develops an original version of the first Morse lemma which holds for sublevel sets of Wasserstein space. The relation between this structure and the results in Chapter 4 is considered in Section 5.3.

The last chapter introduces a categorical perspective. Simplicial metric thickenings are shown to be a particular example of a restricted comma category, a construction introduced in Section 6.1. This framework is used to prove results about the homotopy type of products and coproducts of simplicial metric thickenings. Finally, Section 6.5 gives a version of Dowker's theorem for simplicial metric thickenings.

Essential definitions are found in Section 2.1 and Section 2.3. The reader interested in background on Wasserstein space should peruse Chapter 3 and Section 5.1. Original results are concentrated in Chapters 4 to 6, particularly Sections 4.3, 4.4, 5.2, 6.1, 6.2 and 6.4. Lastly, Chapter 6 can be read almost independently, aside from the major definitions.

# Chapter 2

# Background and Related Work

The motivating problem of this thesis is to understand the homotopy types of certain geometric simplicial complexes—Vietoris–Rips and Čech complexes—and a related class of metric spaces called simplicial metric thickenings. Section 2.1 provides the basic definitions of these simplicial complexes. Section 2.2 then gives an extensive history of the problem and discusses the motivation from applied topology. Lastly, Section 2.3 introduces the simplicial metric thickenings, which are a much more recent construction, and discusses the previous literature on them.

## 2.1 Overview

Metric spaces and simplicial complexes are both mathematical objects that inherently possess geometric and topological features. Simplicial complexes admit a very simple homology theory, so for topological computations it is often necessary to turn a space into a simplicial complex. The Vietoris–Rips and Čech complexes are classical methods of doing this when the original space has a metric, or distance function associated to it.

The more classical construction is the Čech complex. Throughout $(X,d)$ is a metric space (often just denoted $X$), and $r$ is a non-negative real number.

**Definition 2.1.1.** *The **Čech complex** of $(X,d)$ at scale parameter $r$ is the simplicial complex $\check{\mathrm{C}}(X;r)$ with vertex set $X$ and with a simplex $\sigma$ corresponding to every finite subset of $X$ such that there exists a ball $\mathrm{B}_r(y)$ with $\sigma \subseteq \mathrm{B}_r(y)$.*

Equivalently, the Čech complex can be defined as the **nerve** of $\mathscr{U}_r = \{\mathrm{B}_r(x) \mid x \in X\}$, the covering of $X$ by balls of radius $r$. (This equivalence will be discussed further in Chapter 6.) The balls in Definition 2.1.1 may be either open or closed, and when the distinction is important the notation will be $\check{\mathrm{C}}_<(X;r)$ and $\check{\mathrm{C}}_\leq(X;r)$. Definition 2.1.1 is the **intrinsic Čech complex**. If $Y \supseteq X$

is a larger ambient space, the **ambient Čech complex**, $\check{C}(X,Y;r)$ can be defined by requiring that $\sigma \subseteq B_r(y)$ for some $y$ in $Y$.

The Čech complex has a long history in algebraic topology due to the nerve theorem:

**Theorem 2.1.2.** *Let $T$ be a paracompact topological space and $\mathscr{U}$ an open cover of $T$ such that for any open sets $U_1,\dots,U_n$ in $\mathscr{U}$, the intersection $U_1 \cap \cdots \cap U_n$ is either empty or contractible. Then the nerve simplicial complex of $\mathscr{U}$ is homotopy equivalent to $T$.*

A proof can be found in [4, Corollary 4G.3] and originates in the work of Pavel Alexandrov [5]. If the Čech complex at scale $r$ satisfies the contractible intersection requirement, then $\check{C}(X;r) \simeq X$ because $\check{C}(X;r)$ is a nerve complex. The Čech complex does not always satisfy this requirement: see Figure 2.1 for a counterexample.

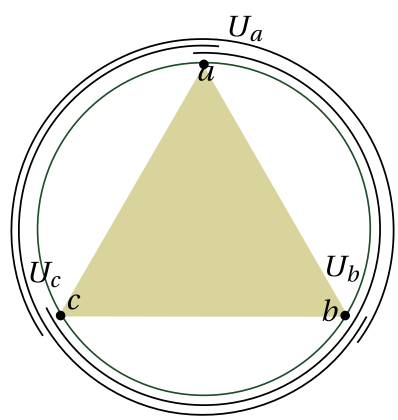


**Figure 2.1:** The open sets $U_a$, $U_b$, and $U_c$ intersect at $a$, $b$, and $c$, so the nerve complex is a 2-simplex.

The Vietoris–Rips complex is similar to the Čech complex, except that containment in a ball is replaced by containment in an arbitrary set of given diameter. The **diameter** of $\sigma \subseteq X$ is defined to be

$$\operatorname{diam}(\sigma) := \sup_{x,y\in\sigma} d(x,y).$$

Of course, for finite subsets the supremum is actually a maximum.

**Definition 2.1.3.** *The **Vietoris–Rips** complex of a metric space $X$ at scale parameter $r$ is the simplicial complex $\mathrm{VR}(X;r)$ with vertex set $X$ and with a simplex $\sigma$ corresponding to every finite* *subset of X such that*

$$\mathrm{diam}(\sigma) < r.$$

Specifically, this defines the **open** Vietoris–Rips complex. Requiring $\mathrm{diam}(\sigma) \leq r$ gives the **closed** Vietoris–Rips complex. The notation $\mathrm{VR}_{<}(X;r)$ and $\mathrm{VR}_{\leq}(X;r)$ will distinguish the two when necessary.

The Vietoris–Rips complex is clearly closely related to the Čech complex. Several precise connections can be made:

1. There is an interleaving, $\mathrm{VR}(X;r) \subseteq \check{\mathrm{C}}(X;r) \subseteq \mathrm{VR}(X;2r)$, since a set $\sigma$ of diameter $r$ is contained in a ball of radius $r$ centered at any point in $\sigma$, and because a ball of radius $r$ has diameter at most $2r$.

2. If $X$ is a geodesic space, then the 1-skeleta of $\mathrm{VR}(X;2r)$ and $\check{\mathrm{C}}(X;r)$ are the same. A ball of radius $r$ centered at the midpoint of the geodesic between $x$ and $y$ has diameter at most $2r$ and contains $x$ and $y$ whenever $d(x,y) < 2r$.

3. If $X$ is a hyperconvex metric space, then $\mathrm{VR}(X;2r) = \check{\mathrm{C}}(X;r)$. Hyperconvexity is addressed at the end of Section 2.2. An example is $\mathbb{R}^2$ with the $L^\infty$ norm.

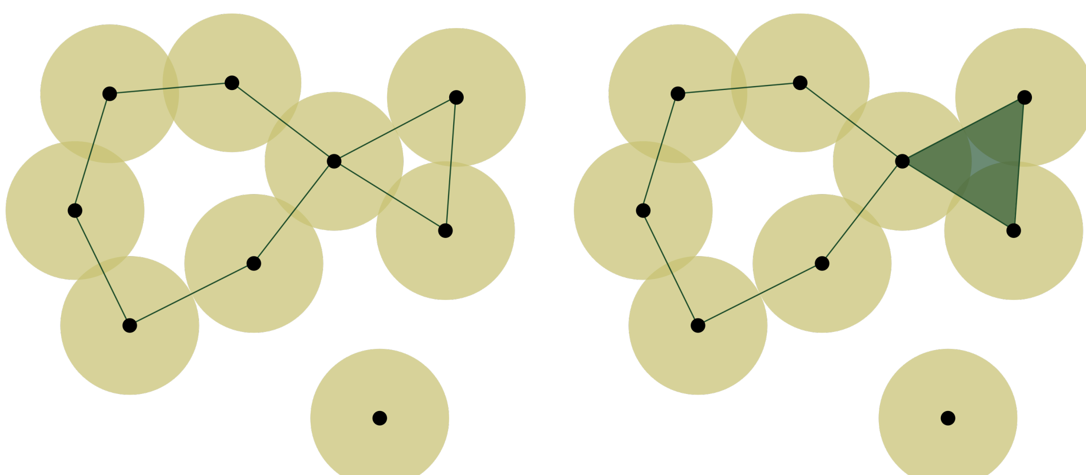

**Figure 2.2:** The Čech (left) and Vietoris–Rips (right) complexes on the same vertex set.

The idea that $\mathrm{VR}(X;r)$ and $\check{\mathrm{C}}(X;r)$ turn metric spaces into simplicial complex can be made precise in a categorical sense by describing $\mathrm{VR}(\square;r)$ and $\check{\mathrm{C}}(\square;r)$ as functors from the category of metric spaces and 1-Lipschitz maps to the category of simplicial complexes. This viewpoint plays a major role in Chapter 6.

## 2.2 Related Work

This section traces the history of the Vietoris–Rips complex, discusses its use in applied topology, and gives an overview of some of the recent work on the question of homotopy types of Vietoris–Rips complexes.

### 2.2.1 Prehistory

The Vietoris–Rips complex originates in the work of Leopold Vietoris [6], who was a pioneer in the study of homology. In 1927 homology had been developed for simplicial complexes, but it was not clear how to associate homology groups to general topological spaces (themselves a new concept at the time). Thus, Vietoris devised the Vietoris–Rips complex as a tool for associating a canonical simplicial complex with any metric space.

To eliminate the arbitrary choice of scale parameter, $r$, he takes the limit as $r \to 0$. More precisely, let $z_k$ be an $n$-cycle in the chain group of $\mathrm{VR}(X;\varepsilon_k)$. Consider a sequence of $n$-cycles $(z_k)_{k=1}^{\infty}$ with $\varepsilon_k \to 0$ as $k \to \infty$. Such a sequence is **fundamental** if for every $\delta > 0$ there exists an $N_\delta$ such that $z_l - z_m = \partial\sigma$ for some $n+1$-chain $\sigma$ in $\mathrm{VR}(X;\delta)$ for all $l, m > N_\delta$. A null fundamental sequence is one for which $z_k = \partial\sigma$ for all $k > N_\delta$. The group of fundamental sequences modulo null sequences is then taken as the homology of the metric space $X$ in dimension $n$ [7]. Geometrically, fundamental sequences contain cycles which "exist" at $r = 0$ and so describe the homology of $X$ itself. Vietoris does not seem to have considered the geometry of $\mathrm{VR}(X;r)$ at any fixed positive scale parameters.

Much later, the same complex appeared in the context of geometric group theory. In [1] Mikhail Gromov defines a polyhedron, $P_d(X)$, which is identical to the Vietoris–Rips complex,

$\mathrm{VR}(X;d)$. Since Gromov attributed several lemmas about $P_d(X)$ to Elihu Rips, the space $P_d(X)$ became known as the Rips complex in the geometric group theory literature. Rips and Gromov were concerned with a problem almost exactly opposite that of Vietoris. For example, [1, Lemma 1.7.A] says, "Let $X$ be a $\delta$-hyperbolic space such that every $x \in X$ can be joined by a segment with a fixed reference point $x_0 \in X$. Then the polyhedron $P_d(X)$ is contractible for all $d \geq 4\delta$." If $X$ is bounded, then when $r$ is greater than or equal to the diameter of $X$ the complex $\mathrm{VR}(X;r)$ is contractible. At this point $\mathrm{VR}(X;r)$ is the complete simplicial complex on the set $X$. (The same is true if $X$ is unbounded if $\mathrm{VR}(X;\infty)$ is defined correctly.) So Gromov and Rips are interested in finding large scale parameters, possibly less than the diameter of $X$, where the Vietoris–Rips complex becomes contractible. Of course, $X$ itself is quite likely not contractible, so at these scales the Vietoris–Rips complex is not typically representative of the topology of $X$, though this parameter contains some geometric information.

The first to recognize that these were the same construction was Jean-Claude Hausmann [2]. More significantly, he was the first to consider the topology of $\mathrm{VR}(X;r)$ at intermediate scales and to raise the question of the structure of the Vietoris–Rips complex of spheres. Since Hausmann's work is fundamental to the study of Vietoris–Rips complexes it is worth considering in some detail.

**Definition 2.2.1.** *A **geodesic space** is a metric space $(X,d)$ in which every pair of points $(x,y)$ admits at least one path $\gamma_y^x\colon [0,T] \to X$ with $\gamma_y^x(0) = x$, $\gamma_y^x(T) = y$, and $\mathrm{length}(\gamma_y^x) = d(x,y)$. Such paths are called **geodesics**.*[1]

The first example of a geodesic space is a Riemannian manifold such the $n$-sphere, $\mathbb{S}^n$.

[1]Geodesics as defined here are sometimes called *minimal* geodesics and should not be confused with geodesics in the sense of Riemannian geometry. In that setting a geodesic is, heuristically, the straightest possible path connecting $x$ and $y$. It can be shown that there is always a Riemannian geodesic which is a geodesic in the sense of Definition 2.2.1. However, there are may be non-minimal Riemannian geodesics. For example, following any great circle on the sphere gives a geodesic path, but there are both long and short directions of traversal and only the shorter satisfies Definition 2.2.1. Throughout this document "geodesic" always means minimal geodesic, even when referring to Riemannian manifolds.

**Definition 2.2.2.** *For any geodesic space, define $r(X)$ to be the least upper bound of the set of real numbers $r$ which satisfy*

1. *If $d(x,y) < 2r$ then there is a unique geodesic from $x$ to $y$,*
2. *if $x$, $y$, and $z$, are three points with $d(x,y)$, $d(x,z)$, and $d(y,z)$ all less than $r$, then any point $u$ on the shortest geodesic from $x$ to $y$ satisfies $d(z,u) \leq \max\{d(x,z), d(y,z)\}$, and*
3. *if $\alpha$ and $\beta$ are arc-length parametrized geodesics[2] with $\alpha(0) = \beta(0)$ and if $0 \leq s, s' < r$ and $0 \leq t \leq 1$, then $d(\alpha(ts), \beta(ts')) \leq d(\alpha(s), \beta(s'))$.*

All three conditions hold trivially for $r = 0$. A compact Riemannian manifold $M$ always has $r(M) > 0$, but in general $r(X)$ need not be positive. The second condition in particular is closely related to the convexity and curvature conditions which will be discussed in Section 4.2. We can now state Hausmann's main theorem, [2, Theorem 3.5]:

**Theorem 2.2.3.** *Let $X$ be a geodesic space and suppose $r(X) > 0$. Then for any $0 < \varepsilon \leq r(X)$, the Vietoris–Rips complex $\mathrm{VR}(X;\varepsilon)$ is homotopy equivalent to $X$.*

*Proof.* The proof involves constructing a map $T\colon \mathrm{VR}(X;\varepsilon) \to X$, showing that $T$ induces an isomorphism on fundamental groups and all homology groups, and implicitly invoking Whitehead's theorem to show that $T$ is a homotopy equivalence. To define $T$ use the axiom of choice to choose a total order of all points in $X$. The construction of $T$ depends on this choice and is therefore non-canonical. Every simplex $\sigma$ in $\mathrm{VR}(X;\varepsilon)$ can then be described uniquely by a finite list of ordered vertices, $\sigma = [x_0, \ldots, x_n]$. Then a map $T_\sigma$ from the standard $n$-simplex, $\Delta_n$, to $X$ is determined by mapping vertices of $\Delta_n$ to $x_0, \ldots, x_n$ and convex combinations of vertices to the corresponding geodesic convex combination in $X$. This requires that $\varepsilon < r(X)$. Finally, the map $T$ is given by $\sigma \mapsto T_\sigma$. □

[2] A geodesic $\gamma$ is arc-length parametrized if $d(\gamma(s), \gamma(s')) = |s - s'|$ for all $s$, $s'$.

Intuitively $X$ ought to include into $\mathrm{VR}(X;\varepsilon)$ by mapping a point $x$ to the vertex $[x]$. However, this is *not* the homotopy inverse to $T$ as it is not a continuous map; in fact, no such inverse is explicitly constructed in Hausmann's proof. This will be important in Section 2.3.

Hausmann concludes his discussion of $\mathrm{VR}(X;\varepsilon)$ by posing two motivating problems:

1. What can be said about the homotopy type of $\mathrm{VR}(Y;r)$ when $Y \subseteq X$? If $Y$ is sufficiently dense in $X$, is there a homotopy equivalence $\mathrm{VR}(Y;r) \simeq X$?

2. What is the homotopy type of $\mathrm{VR}(X;\varepsilon)$ when $\varepsilon > r(X)$? In particular, if $\varepsilon' > \varepsilon$, does

$$\pi_k(\mathrm{VR}(X;\varepsilon)) \cong 0 \text{ for all } 0 < k \leq n$$

   imply that $\pi_k(\mathrm{VR}\left(X;\varepsilon'\right) \cong 0$ for all $0 < k \leq n$?

The first of Hausmann's questions was addressed relatively quickly, by Janko Latschev [8].

**Theorem 2.2.4.** *Let $X$ be a closed Riemannian manifold. Then there exists an $\varepsilon_0 > 0$ and a $\delta_\varepsilon > 0$ for every $0 < \varepsilon \leq \varepsilon_0$ so that if $Y$ is a metric space and the Gromov–Hausdroff distance $d_{\mathrm{GH}}(X,Y) < \delta_\varepsilon$, then $\mathrm{VR}(Y;\varepsilon) \simeq X$.*

An important concept in Latschev's theorem is the Gromov–Hausdorff distance. This is a measure of distance between metric spaces.[3] The **Hausdorff distance** between two subsets $X$ and $Y$ of a metric space $(Z, d_Z)$ is

$$d_{\mathrm{H}}(X,Y) := \max\left\{\sup_{x\in X}\inf_{y\in Y} d_Z(x,y)\,,\, \sup_{y\in Y}\inf_{x\in X} d_Z(x,y)\right\}. \tag{2.1}$$

This mildly unintuitive equation can be thought of as follows: you wish to travel from $X$ to $Y$ in the ambient space $Z$, and you would like to move the shortest distance possible, corresponding to the infimum in equation (2.1). You have an adversary, however, who is to choose where

[3]Between arbitrary metric spaces, $d_{\mathrm{GH}}$ is not a true metric, though it is a metric on the set of compact metric spaces modulo isometry.

you start and who wants to make you travel as far as possible (the supremum). The Hausdorff distance is the maximum distance you can be made to travel from $X$ to $Y$ or from $Y$ to $X$.

A map $f$ between metric spaces $(X, d_X)$ and $(Y, d_Y)$ is an **isometry** if

$$d_X(x_1, x_2) = d_Y(f(x_1), f(x_2)) \text{ for all } x_1, x_2 \in X.$$

Clearly, any isometry must be an injection. The **Gromov-Hausdorff distance** from $X$ to $Y$ is $d_{\mathrm{GH}} := \inf d_{\mathrm{H}}(f(X), g(Y))$ where the infimum is taken over all isometries $f\colon X \to Z$ and $g\colon Y \to Z$, and over all metric spaces $Z$. This should be thought of as optimally aligning $X$ and $Y$ within some larger space $Z$.

If $F \subseteq X$, then $d_{\mathrm{GH}}(F, X) \leq d_{\mathrm{H}}(F, X)$, so dense samplings of $X$ are Gromov-Hausdorff close to $X$. Thus Latschev's theorem is strictly stronger than Hausmann's conjecture. In fact, if $F$ and $X$ both live in an ambient space $Z$, then $d_{\mathrm{H}}(F, X) \geq d_{\mathrm{GH}}(F, X)$. Suppose that $F$ is a finite set of points in $Z$ such that every point in $X$ is within $\varepsilon$ of some point in $F$. Then $d_{\mathrm{H}}(F, X) < \varepsilon$ so Latschev's theorem tells us that $\mathrm{VR}(F; \varepsilon) \simeq X$ for sufficiently small $\varepsilon$. Therefore, one can determine the topology of a space $X$ by computing a finite dense sample of $X$, even with some amount of noise.

The same can be accomplished with Čech complexes by choosing a finite collection of points $F$ in $X$ with sufficient density and $r$ is small. The set of balls $\mathrm{B}_r(x)$ centered at all $x \in F$ is then a good cover of $X$ and Theorem 2.1.2 applies. For subsets of Euclidean space [9] provides a rigorous statement of the probability that a random sample meets the density conditions and gives a Čech complex with the correct topology.

### 2.2.2 Applied Topology

Renewed interest in Vietoris–Rips complexes was sparked by the creation of persistent homology in the mid-2000s. Just as homology is a quantitative description of the topology of a space, persistent homology is a quantitative description of how the topology of a sequence of spaces evolves. This is no mere mathematical curiosity—a collection of data $X$, thought of as points in $\mathbb{R}^d$ or some other metric space, can be turned into a sequence of spaces by constructing $\mathrm{VR}(X;r)$ for an increasing sequence of scale parameters $r$ (and likewise with $\check{\mathrm{C}}(X;r)$). The idea that *data has shape* motivates the field of applied topology. Some of the significant early works on persistent homology include [10–13] A thorough treatment can be found in the books [14, 15]. The results most directly related to geometric simplicial complexes follow.

A **filtered topological space** $\mathbb{U}$ is a functor from the poset $(\mathbb{R}, \leq)$ to the category of topological spaces. More concretely, $\mathbb{U}$ is a collection of spaces $U_r$ indexed by real numbers $r$, along with functions $f_s^r\colon U_r \to U_s$ whenever $r \leq s$, and the requirement that if $r < s < t$, then $f_t^r = f_t^s \circ f_s^r$. In practice the functions are almost always inclusions, and the spaces simplicial complexes. The Vietoris–Rips and Čech complexes are then natural examples, since there is a canonical inclusion $\iota_s^r\colon \mathrm{VR}(X;r) \to \mathrm{VR}(X;s)$ when $r \leq s$, and likewise for Čech complexes.

In the context of persistence the metric space $X$ is typically a finite sampling from a larger unknown space $Z$. Thanks to Latschev's theorem (Theorem 2.2.4), there is reason to expect that the topology of $X$ is the same as that of $Z$ for at least some scale parameters. However, the right scale parameter is also unknown. Persistence thus chooses to be agnostic to the choice of a particular scale and considers the entire sequence of spaces $\mathrm{VR}(X;r)$ for $r \in [0, +\infty]$. (In practice the Vietoris–Rips complex is preferred to the Čech complex for computational reasons—it is much faster to compute pairwise distances than to determine containment in a ball, particularly in high-dimensional space.)

The data of persistent homology of $\mathbb{U}$ in dimension $n$, $PH_n(\mathbb{U})$ consists of the homology (with vector space coefficients) at all scales $r$ and the induced maps $H_n(f_s^r)$ for all $r$ and $s$. This appears to be too much data to be computable, but typically only a finite number of scale parameters see changes in homology, and the homology is finite-dimensional. The Vietoris–Rips complex of a finite set $X$ always has both of these properties. Thus persistence is concisely described by a **barcode**: a multi-set of intervals $[b, d]$ encoding the scale parameters at which a dimension of the homology groups is born and dies. Remarkably, there are efficient algorithms for computing persistent homology [13].

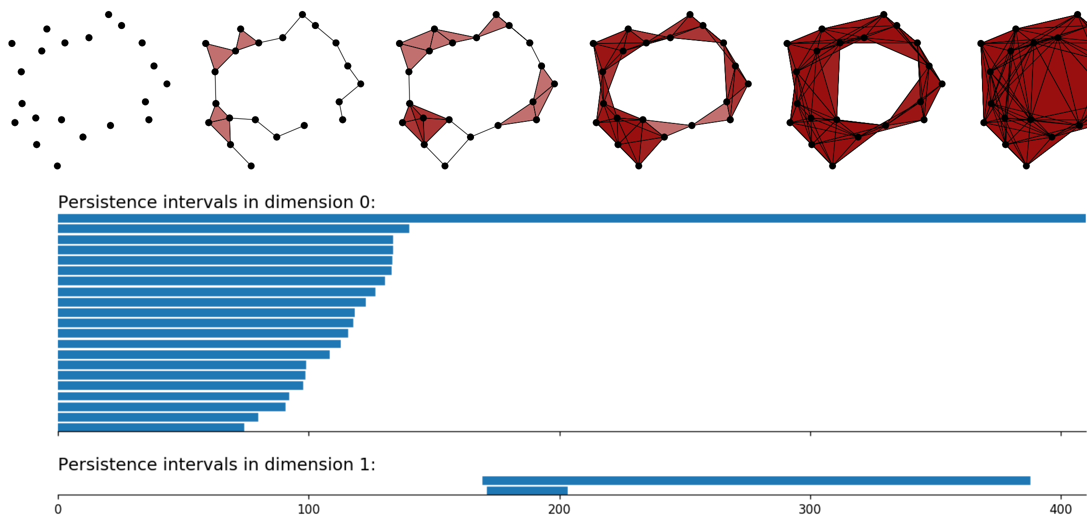


**Figure 2.3:** The Vietoris–Rips filtration (see Project 1) turns a data set into a sequence of simplicial complexes. The barcodes in dimension $i$ record the homology in dimension $i$ as the sequence progresses. Here the 0-dimensional persistent homology intervals show 21 connected components merging into a single connected component, and the 1-dimensional intervals show two 1-dimensional holes, one short-lived and the other long-lived.

An obvious question is the following: given a space $Z$ and a subset $X$, how similar is the persistent homology of $\mathrm{VR}(Z;\square)$ and $\mathrm{VR}(X;\square)$? Answering this first requires a measure of similarity for barcodes. There are a number of choices, including a version of the Wasserstein distance which is pursued more generally in Chapter 3, and as a special case, the Bottleneck distance, $d_B$. Details can be found in [14], among other sources.

This machinery leads to an important generalization of Latschev's theorem:

**Theorem 2.2.5** (Stability [16])**.** *If $X$ and $Y$ are metric spaces and $d_{\mathrm{GH}}(X,Y) < \varepsilon$, then the corresponding persistence diagrams of* $\mathrm{VR}(X;\square)$ *and* $\mathrm{VR}(Y;\square)$ *have bottleneck distance less than* $2\varepsilon$*.*

This theorem is weaker than Latschev's theorem in that it only describes homology, rather than homotopy type. Moreover, it does not state that the homology is the same at each scale parameter, or in fact at any scale parameter. What it effectively says is that the homology groups are never too far from being the same when the initial data are similar. In particular, if $X$ is a dense subset of $Z$, then the persistence of $X$ closely approximates the persistence of $Z$.

### 2.2.3 Further Studies of Vietoris–Rips Complexes

The structure of Vietoris–Rips complexes for particular spaces is now quite an important question. Indeed, if we were to know the homotopy type of $\mathrm{VR}(X;r)$ at all scale parameters for a given reference space, then we would also know its persistent homology exactly. Then if the persistent homology of $Y$ was close to that of $X$, we would have a reference space with which $Y$ could be compared.

A good first candidate space to study is the circle, $\mathbb{S}^1$. By Hausmann's theorem, $\mathrm{VR}\left(\mathbb{S}^1;\varepsilon\right) \simeq \mathbb{S}^1$ for small $\varepsilon$, and so has non-trivial topology, but since the circle is bounded, $\mathrm{VR}\left(\mathbb{S}^1;r\right)$ is contractible when $r$ is at least equal to the diameter of the circle. Therefore there is at least one change in homotopy type as the scale parameter varies.

Recall that on a circle, a metric ball $\mathrm{B}_r(x)$ is an arc centered at $x \in \mathbb{S}^1$. Therefore the Čech complex at scale $r$ is the same as the nerve complex of a collection of circular arcs. In [17] the authors determine the homotopy types of the Čech and Vietoris–Rips complexes of any finite collection of evenly-spaced points in $\mathbb{S}^1$.

**Theorem 2.2.6.** *Let $X_n$ be a set of $n$ evenly-spaced points in $\mathbb{S}^1$ with the convention that the circumference of $\mathbb{S}^1$ is $1$ and that distances are taken with respect to the arc-length distance on $\mathbb{S}^1$. Then*

$$\check{\mathrm{C}}_{\leq}\left(X_n, \mathbb{S}^1; \frac{k}{2n}\right) \simeq \begin{cases} \bigvee^{n-k-1} \mathbb{S}^{2\ell} & \text{if } \frac{k}{n} = \frac{\ell}{\ell+1} \\ \mathbb{S}^{2\ell+1} & \text{if } \frac{\ell}{\ell+1} < \frac{k}{n} < \frac{\ell+1}{\ell+2} \end{cases} \quad \text{for some } \ell \geq 0$$

*and*

$$\mathrm{VR}_{\leq}\left(X_n; \frac{k}{n}\right) \simeq \begin{cases} \bigvee^{n-2k-1} \mathbb{S}^{2\ell} & \text{if } \frac{k}{n} = \frac{\ell}{\ell+1} \\ \mathbb{S}^{2\ell+1} & \text{if } \frac{\ell}{2\ell+1} < \frac{k}{n} < \frac{\ell+1}{2\ell+3} \end{cases} \quad \text{for some } \ell \geq 0$$

Moreover, they show that a complex built on non-evenly spaced points is always homotopy equivalent to one comprised of some number of evenly spaced points [17, Theorem 5.4].

Expanding on this result, in [18] Michał Adamaszek and Henry Adams determine $\mathrm{VR}\left(\mathbb{S}^1;r\right)$ for all $r$.

**Theorem 2.2.7.** *With the same conventions as in Theorem 2.2.6,*

$$\mathrm{VR}_{<}\left(\mathbb{S}^1; r\right) \simeq \mathbb{S}^{2\ell+1} \textit{ for } \frac{\ell}{2\ell+1} < r \le \frac{\ell+1}{2\ell+3}, \quad \ell = 0, 1, 2, \ldots$$

*and*

$$\mathrm{VR}_{\le}\left(\mathbb{S}^1; r\right) \simeq \begin{cases} \mathbb{S}^{2\ell+1} & \textit{if } \frac{\ell}{2\ell+1} < r < \frac{\ell+1}{2\ell+3} \\ \bigvee^{\mathfrak{c}} \mathbb{S}^{2\ell} & \textit{if } r = \frac{\ell}{2\ell+1} \end{cases}, \quad \ell = 0, 1, 2, \ldots$$

Note that here $\mathfrak{c} = 2^{\aleph_0}$ is the cardinality of the reals, so that the last line indicates an (uncountably) infinite wedge sum of spheres. The proof of this theorem is highly combinatorial. We will sketch some of the major steps in order to contrast it with the theory of metric thickenings developed later.

*Proof.* To begin, consider a finite subset $X_n \subseteq \mathbb{S}^1$. The points in $X_n$ can be cyclically ordered by reading them clockwise, so the 1-skeleton of $\mathrm{VR}(X_n; r)$ is a cyclic graph. Let $\vec{C}_n^k$ be the cyclic graph with $n$ equally spaced vertices in which each vertex is connected by an edge to the next $k$ vertices in clockwise order. Cyclic graphs have a fundamental invariant called the **winding fraction**, defined by

$$\mathrm{wf}(\vec{G}) = \sup\left\{ \frac{k}{n} \;\middle|\; \text{there exists a cyclic homomorphism } \vec{C}_n^k \to \vec{G} \right\}.$$

A **cyclic homomorphism** is a directed graph homomorphism which respects the cyclic ordering and which is not constant whenever the domain has a directed cycle. A vertex $w$ in a cyclic graph is **dominated** by a vertex $v$ if there is a directed edge $v \to w$ and the set of vertices reached by directed edges out of $v$ is the union of $w$ and the set of vertices reached by directed edges out of $w$.

By removing dominated vertices, any cyclic graph can be dismantled to a graph $\vec{C}_n^k$ for some $n$ and $k$, and if a cyclic graph $\vec{G}$ dismantles to $\vec{C}_n^k$ then $\mathrm{wf}(\vec{G}) = \frac{k}{n}$ [18, Propositions 3.12 and 3.14]. A dismantling of cyclic graphs induces a homotopy equivalence on the associated clique complexes. The homotopy type of any (finite) Vietoris–Rips complex on the circle is then obtained by dismantling to some $\overrightarrow{C}_n^k$ and applying Theorem 2.2.6. The winding fraction is directly related to the scale parameter. In particular, $\mathrm{wf}(\mathrm{VR}(X;r)) > r - 2\varepsilon$ if $X$ is $\varepsilon$-dense in $\mathbb{S}^1$ [18, Proposition 5.2].

The last difficulty is in passing to infinite subsets, including the entire circle. It can be shown that $\mathrm{VR}(\mathbb{S}^1;r)$ is the colimit of $\mathrm{VR}(X;r)$ over all finite subsets $X$ of $\mathbb{S}^1$ ordered by inclusion. Theorem 2.2.7 follows from showing that the homotopy types of finite subsets stabilize in the poset order.[4] □

This proof depends strongly on the use of cyclic graphs, which prevents it from generalizing to spaces besides $\mathbb{S}^1$. Even changing the metric on $\mathbb{S}^1$ can alter the results, as [19]—which studies the Vietoris–Rips complexes of ellipses (as subsets of Euclidean space)—shows.

These results can be extended to give the persistence of a class of spaces called **metric gluings**. The authors of [20] completely characterize the 1-dimensional homology of the Čech complex filtration of any **metric graph**—a finite graph equipped with the path-length metric. Combining this with [18] allows the authors of [21] to characterize the Vietoris–Rips persistence of certain metric graphs in all homological dimensions.

Determining the precise homotopy type of the Vietoris–Rips complexes of general spaces and at all scales is still a formidable challenge, but there are results that classify the homology of homotopy groups of classes of spaces at different scales. Zero-dimensional persistence in fact can be easily characterized. Suppose that a metric space $X$ is the disjoint union of connected components $X_1, \ldots, X_n$. Let

$$r_{IJ} = d(X_I, X_J) = \inf_{x_i \in X_I} \inf_{x_j \in X_J} d(x_i, x_j).$$

Then the 0-dimensional Vietoris–Rips persistence has $n$ bars when $0 < r \leq \min_{I,J} r_{IJ}$. Assuming all $r_{IJ}$ are distinct, one bar dies at each successive $r_{IJ}$ as the two components are merged by the

[4] This is excluding the "singular" case with the $\leq$ convention at scale parameters $\frac{\ell}{2\ell+1}$, which require a separate analysis [18, Section 8].

appearance of an edge $[x_i, x_j]$. One can think of the scales $r_{IJ}$ as "critical" scale parameters. If $X$ is a subset of an appropriate ambient space, then the Čech persistence in dimension zero is similarly characterized.

Higher dimensions are not so easily characterized, but when $X$ is a geodesic space, Žiga Virk gives a description of the 1-dimensional persistent homology [22]. He does so by identifying conditions for a scale parameter to be critical for 1-dimensional persistence.

Consider a 1-cycle $z \in Z_1(\mathrm{VR}(X;r))$, the simplicial 1-cycles of $\mathrm{VR}(X;r)$. The goal is to give a geometric description of when $z$ is null-homologous. The data describing $z$ are a list of adjacent edges in $\mathrm{VR}(X;r)$, so $z$ can be concisely written as a sequence of vertices $x_0, x_1, \ldots, x_n$, which themselves are points in $X$. Define a **realization** of $z$ to be a path $\zeta$ in $X$ which passes through $x_0, \ldots, x_n$ and which is a geodesic between $x_i$ and $x_{i+1}$. Geodesics need not be unique, so there may be multiple paths corresponding to a given cycle; however, the lengths of all such paths are the same. Virk shows [22, Proposition 4.7, Theorem 4.8] that $z$ is null-homologous in $Z_1(\mathrm{VR}(X;r))$ if and only if $\zeta$ has length less than $3r$, or equivalently if the smallest ball enclosing $\zeta$ has radius $\frac{3r}{2}$.

A different set of techniques are used by Matthew Zaremsky in [23] to determine the Vietoris–Rips complexes of spheres at small scale parameters. He uses Bestvina–Brady discrete Morse theory with the multi-valued Morse function $(\mathrm{diam}, -\dim)\colon \mathrm{VR}(\mathbb{S}^n; r) \to \mathbb{R}^2$ to show that for $0 < r < \frac{1}{4}$ (where spheres are given circumference one), $\mathrm{VR}(\mathbb{S}^n; r) \simeq \mathbb{S}^n$ [23, Proposition 5.2].

Recently [24] uses a promising technique to determine the homotopy type of the Rips complexes of a number of spaces. The authors show that the first change in homotopy type for spheres happens after $r = \arccos\left(-\frac{1}{n+1}\right)$ where $n$ is the dimension of the sphere, here given circumference $2\pi$. They also obtain the next homotopy type of $\mathrm{VR}\left(\mathbb{S}^1; r\right)$, $\mathrm{VR}\left(\mathbb{S}^2; r\right)$, and $\mathbb{CP}^n$. Their technique is to embed the space $X$ into its **Kuratowski space** $K(X)$, the set of all functions $X \to \mathbb{R}$ equipped with the supremum norm (significantly, not the space of continuous functions). This is a hyperconvex space, so the Vietoris–Rips and Čech complexes in $K(X)$ coincide.

The Kuratowski space was studied extensively Mikhail Katz [25] which provides the foundation for the results in [24].

## 2.3 Metric Thickenings

One piece absent from the preceding discussion is [3], the major contribution of which is to introduce the Vietoris–Rips *metric thickening*, which is a major object of study in this thesis. As motivation for this new space, consider the simplicial complex topology in detail. The standard $n$-simplex, $\Delta_n$, has a topology as the convex hull of the standard basis vectors $e_0, \ldots, e_n$ in $\mathbb{R}^{n+1}$. Any simplex, $\sigma$, in a simplicial complex, $K$, has a **geometric realization**, $|\sigma|$, obtained by identifying $\sigma$ with the standard $n$-simplex in Euclidean space. The geometric realization is extended to all of $K$ by

$$|K| := \coprod_{\sigma \in K} |\sigma| / \sim$$

where $\sim$ is the identification along faces. The associated topology is also called the **simplicial complex topology** of $K$. The Vietoris–Rips complex is typically given a topology in exactly this way and this is the topology to which all of the preceding results about homotopy type refer.

When restricted to the vertex set the simplicial complex topology is discrete. Therefore $\mathrm{VR}_{\leq}(X;0)$, which is simply the set of points in $X$, is not homeomorphic to $X$ in general. The metric on $X$ has been forgotten. (This is important to the construction in Chapter 6.) Even stronger, the inclusion $x \mapsto [x]$ is not a continuous map unless $X$ is discrete.

The metric on $X$ is generally not compatible with the simplicial complex topology at all. A topological space $T$ is **metrizable** if it is homeomorphic to some metric space $X$. A simplicial complex is **locally finite** if each vertex is contained in at most a finite number of simplices. In order for a topological space $T$ to be metrizable, it must be **first-countable**, meaning that to every point $p \in T$ it is possible to associate a collection of open sets $\mathcal{U} = U_n$, $n \in \mathbb{N}$, such that any open set $V$ containing $p$ contains some $U_n$.

**Proposition 2.3.1.** *The geometric realization of a simplicial complex $K$ is metrizable only if $K$ is locally finite.*

*Proof.* If $K$ is not locally finite, then there exists a vertex $v$ in $K$ contained in an infinite number of edges, $E_n$, with $n \in \mathbb{N}$ (if $v$ is contained in an uncountable number of simplices, then any countable subset of those suffices here). Let $\mathscr{U}$ be a countable neighborhood basis. The simplicial complex topology has the property that the intersection of any open set $U_n$ with any simplex is open in that simplex, thus $U_n \cap E_n$ is open in $E_n$. For each $n \in \mathbb{N}$, choose some $a_n \in U_n \cap E_n$ such that $a_n \neq v$. Let $V = |K| \setminus \{a_n \mid n \in \mathbb{N}\}$. This is an open set in $|K|$, since its complement is a countable union of closed sets (singletons), and it contains $v$. However, it does not contain any $U_n$ since each $U_n$ contains an $a_n$, and $V$ contains no $a_n$. Therefore $|K|$ is not first-countable, and hence not metrizable. $\square$

In fact, the converse of Proposition 2.3.1 holds as well. For the stronger result, see [26, Proposition 4.2.16].

Now observe that if $X$ contains a ball, $\mathrm{B}_r(x)$, consisting of an infinite number of points, then $\mathrm{VR}(X;r)$ is not locally finite because $x$ will be the vertex of an infinite number of simplices. In particular, the Vietoris–Rips complex of an open subset of $\mathbb{R}^n$ (and therefore of any manifold) is not locally finite. Whatever properties of $X$ the Vietoris–Rips complex exhibits, it fundamentally cannot retain the metric.

At the same time, as long as $X$ is a finite set, and therefore a discrete metric space, the Vietoris–Rips complex does replicate the features of $X$, and it is this setting where the Vietoris–Rips complex is used in applied topology. This raises the question: is the simplicial complex topology the correct approach if we want to understand the limit of finite samples from a metric space?

Answering this question, of course, requires the existence of some alternative topology with better properties. The alternative studied here and in [3] is the Vietoris–Rips metric thickening.

If $\sigma = [x_0, \dots, x_n]$ is an $n$-simplex, then every point in $|\sigma|$ can be formally written as $\sum_{i=0}^{n} \lambda_i x_i$ where $0 \leq \lambda_i \leq 1$. Any point, excluding vertices, in the geometric realization of a simplicial complex $|K|$ can be written uniquely as $\sum_{i=0}^{n} \lambda_i x_i$ for some $n$, and with $0 < \lambda_i < 1$ (in other words, every point is in the interior of a unique minimal simplex), which defines **barycentric coordinates** on $K$.

Now suppose that $K$ is a simplicial complex and the set of vertices is a metric space $X$. Define a **matching** between points $\mu = \sum_{i=0}^{m} \alpha_i x_i$ and $\nu = \sum_{j=0}^{n} \beta_j x_j$ in $|K|$ to be any bistochastic matrix $M$ with marginals $\mu$ and $\nu$, that is, $M$ is an $m \times n$ matrix such that

$$M \cdot \mathbb{1} = \vec{\alpha} \text{ and } \mathbb{1} \cdot M = \vec{\beta}^T$$

where $\vec{\alpha}$ and $\vec{\beta}$ are the vectors of coefficients $\alpha_i$ and $\beta_j$ and $\mathbb{1}$ denotes the vector $[1, 1, \ldots, 1]^T$. Call the set of all matchings between $\mu$ and $\nu$, $\Gamma(\mu, \nu)$. Let $D$ be the $m \times n$ matrix with entries $D_{ij} = d(x_i, x_j)$. The **cost** of a matching $M$ is

$$\text{cost}(M) := \mathbb{1}^T \cdot M \odot D \cdot \mathbb{1}$$

where $\odot$ denotes the pointwise matrix product.

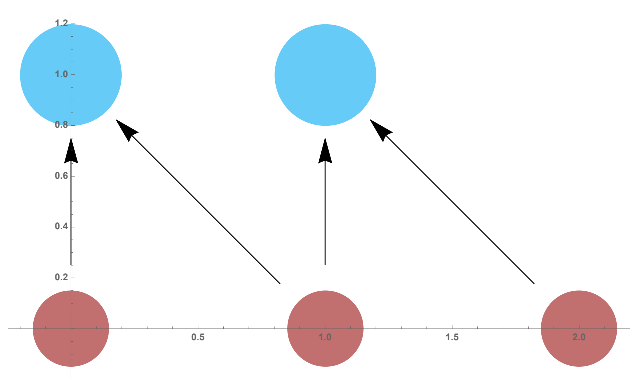


**Figure 2.4:** The optimal transport plan between $\mu = \frac{1}{3}\delta_{(0,0)} + \frac{1}{3}\delta_{(1,0)} + \frac{1}{3}\delta_{(2,0)}$ (in red) and $\nu = \frac{1}{2}\delta_{(1,1)} + \frac{1}{2}\delta_{0,1}$ (in blue) involves splitting the mass at $(1,0)$ between $(0,1)$ and $(0,2)$. (Here $\delta_p$ is the Dirac delta distribution centered at $p$.)

The 1-**Wasserstein distance** on $|K|$ is

$$W_1(\mu,\nu) := \inf_{M\in\Gamma(\mu,\nu)} \text{cost}(M).$$

This is a special case of the Wasserstein distance described in Chapter 3. The Wasserstein distance is a metric on $|K|$ and induces a topology. The geometric realization of $K$ with the Wasserstein metric will be denoted by $\mathscr{K}$.

By Proposition 2.3.1, $\mathscr{K}$ and $|K|$ with the usual topology are not generally homeomorphic. However, when $K$ has a locally finite vertex set (in particular, when $K$ is finite) then $\mathscr{K}$ and $|K|$ are homeomorphic. The map taking the formal sum $\mu = \sum_{i=0}^{n} \lambda_i x_i$ in $\mathscr{K}$ to the corresponding point in Barycentric coordinates in $|K|$ is in that case a homeomorphism.

The Vietoris–Rips complex necessarily has a metric space as its vertex set, and so admits a Wasserstein metric, giving the **Vietoris–Rips metric thickening**, $\mathscr{VR}(X;r)$. For two points $x$ and $y$ in $X$, $W_1(x,y) = d(x,y)$ since there is only one matching, the trivial one. Thus the Wasserstein metric restricted to the vertex set of $\mathscr{VR}(X;r)$ is the original metric on $X$, and $X$ isometrically embeds into $\mathscr{VR}(X;r)$.

An $r$-thickening of a metric space is a larger embedding space $Z \supseteq X$ such that $X$ is isometrically embedded in $Z$ and $d(z,X) \le r$ for all $z \in Z$. The Vietoris–Rips metric thickening is an $r$-thickening of $X$, since

$$W\left(\sum_{i=0}^{n} \lambda_i x_i, X\right) \le W\left(\sum_{i=0}^{n} \lambda_i x_i, x_0\right) = \sum_{i=0}^{n} \lambda_i d(x_i, x_0) \le r \sum_{i=0}^{n} \lambda_i = r.$$

Here the central equality holds because there is only one transport plan to any singleton.

The Vietoris–Rips thickening also satisfies a version of Hausmann's theorem. If $M$ is a Riemannian manifold, then any (weighted) finite collection of points $\mu \subseteq M$ with sufficiently small diameter has a unique Riemannian center of mass, or **Karcher mean**[5].

[5]Chapter 4 explicates the rigorous definition of the Riemannian center of mass and its role in the proof of Theorem 2.3.2.

**Theorem 2.3.2.** *Let $M$ be a complete Riemannian manifold and $\rho$ sufficiently small so that every subset of diameter at most $\rho$ has a unique Karcher mean. Then for any $0 \le r < \rho$, the Vietoris–Rips metric thickening $\mathcal{VR}(M;r)$ is homotopy equivalent to $M$.*

*Proof.* There are three key elements. First, the natural inclusion $\iota\colon M \to \mathcal{VR}(M;r)$ is continuous, since it is an isometric embedding as previously stated. Second, the map $g\colon \mathcal{VR}(M;r) \to M$ sending $\mu$ to its Karcher mean is continuous, and finally, these two maps are homotopy inverses, which follows by showing that whenever $\mu = [x_0,\dots,x_k]$ is a $k$-simplex in $\mathrm{VR}(M;r)$, then $\hat{\mu} = [x_0,\dots,x_k,g(\mu)]$ is a $(k+1)$-simplex in $\mathrm{VR}(M;r)$. This allows $\iota\circ g$ to be homotopic to the identity in $\mathcal{VR}(M;r)$ by mapping linearly from $\mu$ to $g(\mu)$ within $\hat{\mu}$. □

For the special case of spheres the bound can be extended to $r < r_n$, where $r_n$ is the diameter of the inscribed regular $(n+1)$ simplex in $\mathbb{S}^n$. The exact technique in the proof of Theorem 2.3.2 can be used to show a metric version of the nerve lemma [3, Theorem 4.4], and of Latschev's theorem [3, Corollary 6.8]. The Vietoris–Rips metric thickening then appears to enjoy all the same reconstruction properties as the Vietoris–Rips simplicial complex, with the additional benefit of preserving the metric.

The last results which need to be translated to metric thickenings are the known homotopy types of manifolds at scales beyond Hausmann's limit. Let $A_n$ be the alternating group on $n$ elements. Note that $A_{n+2}$ acts on $\Delta_{n+1}$ by rotations. This gives an group homomorphism $A_{n+2} \hookrightarrow \mathrm{SO}(n+1)$. Theorem 5.4 in [3] shows that the first change in homotopy type of $\mathcal{VR}_{\le}(\mathbb{S}^n;r)$ is at $r = r_n$, and that

$$\mathcal{VR}_{\le}\left(\mathbb{S}^n;r_n\right) \simeq \Sigma^{n+1}\frac{\mathrm{SO}(n+1)}{A_{n+2}},$$

the $(n+1)$-fold suspension of the quotient space. In particular,

$$\mathcal{VR}_{\le}\left(\mathbb{S}^1;\frac{1}{3}\right) \simeq \Sigma^2\frac{\mathrm{SO}(2)}{A_3} = \Sigma^2\frac{\mathbb{S}^1}{\mathbb{Z}/3\mathbb{Z}} = \Sigma^2\mathbb{S}^1 = \mathbb{S}^3.$$

This is noteworthy in that it is not the same as the homotopy type of $\mathrm{VR}_{\leq}\left(\mathbb{S}^1; \frac{1}{3}\right)$. Therefore, $|K|$ and $\mathscr{K}$ are not always homotopy equivalent. At the same time, it is the homotopy type of $\mathrm{VR}\left(\mathbb{S}^1; r\right)$ when $\frac{1}{3} < r < \frac{2}{5}$, which suggests that perhaps $\mathrm{VR}_{<}(X; r) \simeq \mathscr{V}\mathscr{R}_{<}(X; r)$ in general, and that $\mathrm{VR}_{\leq}(X; r)$ and $\mathscr{V}\mathscr{R}_{\leq}(X; r)$ only differ at these critical scales where the homotopy type of $\mathrm{VR}_{\leq}(X; r)$ is wild.

# Chapter 3

# Optimal Transport and Wasserstein Spaces

This chapter provides an introduction to the Wasserstein distance on spaces of probability measures. Section 3.1 discusses the optimal transport problem, which is the historical and theoretical motivation for the Wasserstein metric. Next, Section 3.2 goes into the analytical details of the Wasserstein metric and Wasserstein spaces of probability measures, establishing basic properties of the metric such as equivalence with the weak topology. Thought the Wasserstein distance is familiar to applied topologists as a metric on persistence diagrams, many of these details in more general spaces may be unfamiliar and so are covered in detail. Particular attention is payed throughout to how the theory relates to measures with finite support since these are the foundation of the Vietoris–Rips metric thickening.

Throughout $X$ is a metric space and measures on $X$ are assumed to be with respect to the Borel $\sigma$-algebra of $X$. Background on measure and probability theory can be found in [27].

## 3.1 Optimal Transport

The problem of optimal transport has a long history. In 1781 Gaspard Monge formulated it like this: Suppose that you are given a pile of dirt and an equivoluminal hole situated some distance away; in order to minimize effort (mass times distance), which shovelful of dirt should be placed in which part of the hole? [28].

The modern mathematical formulation interprets the pile and the hole as probability distributions, $\mu$ and $\nu$, on a metric space $X$. A transport plan is a function $T\colon X \to X$ which transforms $\mu$ into $\nu$ via pushforward, $\nu = T\#\mu$.[6] (Recall that the pushforward of $\mu$ along a measurable map $T$ is the measure $T\#\mu$ defined by

[6]It is possible without any significant change in theory to consider $T\colon X \to Y$, but the Wasserstein distance, and therefore the focus of this chapter, is on the case of $T\colon X \to X$.

$$T\#\mu(A) := \mu(T^{-1}(A)). \tag{3.1}$$

for any measurable set $A$.) A fundamental property of the pushforward is the change of variables formula: if $T\colon X \to X$ and $f\colon X \to \mathbb{R}$, then

$$\int_X f \circ T(x)\,\mathrm{d}\mu(x) = \int_X f\,\mathrm{d}T\#\mu(x). \tag{3.2}$$

If $\mu$ is a finitely-supported measure, then $T\#\mu$ is as well, and so if $\nu = T\#\mu$, $T$ must be a function moving the support of $\mu$ surjectively onto the support of $\nu$. Moreover, there are clearly no transport plans between measures of unequal mass; hence the problem must be restricted to probability measures (those with total mass one).

In this formalization the optimal transportation problem is to minimize the cost functional

$$I(T) := \int_X d^p(T(x), x)\,\mathrm{d}\mu(x)$$

over the set of all transport plans $T$ with $\nu = T\#\mu$, where $p \in [0, +\infty)$. (Monge considered the case $p = 1$, while modern treatments often focus on $p = 2$ or the general case.) Unfortunately, this problem often fails to have a solution.

**Example 3.1.1.** Suppose that $\mu = \frac{1}{2}\delta[x_1] + \frac{1}{2}\delta[x_2]$, and $\nu = \frac{3}{4}\delta[y_1] + \frac{1}{4}\delta[y_2]$ where $x_1$, $x_2$, $y_1$, and $y_2$ are any four distinct points in $\mathbb{R}^d$. There are no transport plans from $\mu$ to $\nu$ because for any function $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $\mu(f^{-1}(\{y_1\}))$ is either 0, $\frac{1}{2}$, or 1. The issue is that the masses in $\mu$ must be "split" in order to produce $\nu$, but no function can do so. Nor are there any transport plans from $\nu$ to $\mu$ for the same reason. This example is fundamental, and it is a major point of distinction between discrete measures and absolutely continuous ones. ♦

Solving the optimal transport problem involves a common technique: enlarge the space of possible solutions, and then prove that under some restrictions the solution actually lives in the desired subspace. In particular, any measurable function $T\colon X \to X$ and measure $\mu$ gives rise to a measure $\gamma$ on the product space $X \times X$ via $\gamma := (\mathrm{id}, T)\#\mu$. If $\pi_1$ and $\pi_2$ are the projections

$X \times X \to X$ onto the first and second factors, then $\pi_1 \# \gamma = \mu$ and $\pi_2 \# \gamma = T \# \mu$, in other words, the **marginals** of $\gamma$ are $\mu$ and $T\#\mu$. Thus a relaxation of the optimal transport problem, due to Kantorovich [29] is to search over all measures $\gamma$ on $X \times X$ with marginals $\mu$ and $\nu$. The objective function is now

$$I(\gamma) := \int_{\mathbb{R}^d \times \mathbb{R}^d} d^p(x, y) \, \mathrm{d}\gamma(x, y).$$

The measures $\gamma$ are called **transference plans** (following Villani [30]) and the collection of all such will be denoted $\Gamma(\mu, \nu)$.

**Example 3.1.2.** The measures in Example 3.1 admit transference plans. In particular, $\gamma = \frac{1}{4}\delta[(x_1, y_2)] + \frac{1}{2}\delta[(x_2, y_1)] + \frac{1}{4}\delta[(x_1, y_1)]$ is one such plan. Optimality, of course, depends on the locations of $x_1$, $x_2$, $y_1$, and $y_2$. ♦

Now three questions must be answered to solve the optimal transport problem: the existence and uniqueness of an optimal $\gamma$, and whether $\gamma$ has the form $(\mathrm{id}, T)\#\mu$ for some $T$. The Kantorovich duality theorem answers the existence and uniqueness questions, and Brenier's theorem gives a precise characterization of optimal transport plans when they exist.

**Remark 3.1.3.** Before stating these theorems in Section 3.1.1 and Section 3.1.2, it is convenient for the sake of examples to describe a formalization of the problem when restricted to measures supported on finite sets. Any such measure has the form $\mu = \sum_{i=0}^{n} \lambda_i \delta[x_i]$. A transference plan $\gamma$ between $\mu$ and $\nu = \sum_{i=0}^{m} \xi_j \delta\left[y_j\right]$ is a measure supported on (a subset of) the points $(x_i, y_j)$. Letting $\lambda$ and $\xi$ be the vectors of length $n+1$ and $m+1$ corresponding to the weights $\lambda_i$ and $\xi_j$, $\gamma$ can be written as a matrix with $\gamma_{ij}$ the weight on $(x_i, y_j)$. The marginal condition now says that $\gamma \cdot \mathbb{1} = \lambda$ and $\mathbb{1} \cdot \gamma = \xi^T$. Let $D$ be the matrix with $D_{ij} = d^p(x_i, y_j)$. Then the objective function is

$$I(\gamma) = \mathbb{1}^T \cdot D \odot \gamma \cdot \mathbb{1}$$

with $\odot$ denoting the pointwise or Hadamard matrix product.

### 3.1.1 Kantorovich Duality

The exposition here follows [30]. As a technical detail, $X$ is always a locally-compact, complete, and separable metric space.

**Theorem 3.1.4.** *For any pair* $(\mu, \nu)$, *there is a* $\gamma \in \Gamma(\mu, \nu)$ *which minimizes* $I$.

*Proof.* The proof follows that found in [31]. First, observe that the product measure $\mu \otimes \nu$ has marginals $\mu$ and $\nu$, so $\Gamma(\mu, \nu)$ is nonempty. Put the weak topology on $\Gamma(\mu, \nu)$, meaning that a sequence $\{\gamma_k\}$ converges to $\gamma$ if for any lower semi-continuous function $f\colon X \times X \to \mathbb{R} \cup +\infty$,

$$\lim_{k\to\infty} \int_{X\times X} |f| \, \mathrm{d}\gamma_k = \int_{X\times X} |f| \, \mathrm{d}\gamma.$$

Prokhorov's theorem says that any tight sequence of measures on a Polish space has a convergent subsequence and that any convergent sequence of measures is tight. (Recall that a sequence $\{\alpha_k\}$ of measures is **tight** if for every $\varepsilon > 0$ there is a compact set $K$ such that $\alpha_k(X \setminus K) < \varepsilon$ for every $k$.) Let $\{\alpha_k\}$ be any sequence in $\Gamma(\mu, \nu)$. By Prokhorov's theorem there are two compact subsets of $X$, $K_\mu$ and $K_\nu$, such that $\mu(X \setminus K_\mu) < \frac{\varepsilon}{2}$ and $\nu(X \setminus K_\nu) < \frac{\varepsilon}{2}$. The product $K_\mu \times K_\nu$ is a compact subset of $X \times X$. Then

$$\alpha_k((X \times X) \setminus (K_\mu \times K_\nu)) \le \alpha_k((X \setminus K_\mu) \times X) + \alpha_k(X \times (X \setminus K_\nu))$$

by additivity of measures, and by definition of $\Gamma(\mu, \nu)$

$$\alpha_k((X \setminus K_\mu) \times X) + \alpha_k(X \times (X \setminus K_\nu)) = \mu(X \setminus K_\mu) + \nu(X \setminus K_\nu) < \varepsilon.$$

Hence any sequence of measures has a convergent subsequence, so $\Gamma(\mu, \nu)$ is sequentially compact.

Let $\{\gamma_n\}_{n=0}^\infty$ be any sequence of measures converging weakly to $\gamma$ in $\Gamma(\mu, \nu)$. Then

$$\liminf_{n\to\infty} \int_{X\times X} d(x, y)^p \, \mathrm{d}\gamma_n(x, y) \ge \int_{X\times X} d(x, y)^p \, \mathrm{d}\gamma(x, y)$$

by the definition of weak convergence and the continuity of $d(x,y)^p$. This implies immediately that $I$ is lower semi-continuous.

The extreme value theorem says that any lower semi-continuous function from a sequentially compact space to $\mathbb{R} \cup +\infty$ achieves its minimum. $\square$

**Remark 3.1.5.** Theorem 3.1.4 is not the most general existence theorem which can be stated. Most obviously, the properties of the distance function were never used, aside from continuity, so the theorem actually holds for a much wider class of cost functions $c(x,y)$. Many different versions of this result can be found in [32, Section 4]. Worth mentioning is the result of [33] which removes the assumption of a Polish space in exchange for the mild restriction of working only with Radon measures.

If one only wants the value of the minimum, rather than a description of the transference plan, then Kantorovich's duality theorem provides a dual formulation of the problem.

**Theorem 3.1.6** (Kantorovich Duality)**.** *Let $\Phi$ be the set of all pairs of bounded, continuous functions $(\phi,\psi)$ such that $\phi(x)+\psi(y) \le d^2(x,y)$ for $\mu$-almost every $x$ and $\nu$-almost every $y$. Then*

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int d^2(x,y)\,\mathrm{d}\gamma = \max_{(\phi,\psi)\in\Phi} \int_X \phi\,\mathrm{d}\mu + \int_X \psi\,\mathrm{d}\mu.$$

A delightful intuition is provided by Cedric Villani in [30].

> Suppose for instance that you are both a mathematician and an industrialist, and want to transfer a huge amount of coal from your mines to your factories. You can hire trucks to do this transportation problem, but you have to pay them $c(x,y)$ for each ton of coal which is transported from place $x$ to place $y$. Both the amount of coal which you can extract from each mine, and the amount which each factory should receive, are fixed. As you are trying to solve the associated Monge-Kantorovich [optimal transport] problem in order to minimize the price you have to pay, another mathematician comes to you and tells you "My friend, let me handle this for you: I will ship all your coal with my own trucks and you won't have to

worry about what goes where. I will just set a price $\phi(x)$ for loading one ton of coal at place $x$, and a price $\psi(y)$ for unloading it at destination $y$. I will set the prices in such a way that your financial interest will be to let me handle *all* your transportation! Indeed, you can check very easily that for any $x$ and $y$, the sum $\phi(x)+\psi(y)$ will always be less that [*sic*] the cost the cost $c(x,y)$ (in order to achieve this goal, I am even ready to give financial compensations for some places, in the form of negative prices!)".

Using the conventions of Remark 3.1.3, Theorem 3.1.6 in the finite setting reduces precisely to the usual duality of linear programming, so it can be thought of as a generalization of a certain class of linear programs to the continuous setting.

We now turn to characterizing the optimal transport plans. While the minimum is achieved by Theorem 3.1.4, it need not be unique. The set of measures which achieve the minimum will be written $\Gamma_0(\mu,\nu)$.

**Example 3.1.7.** As a simple instance of non-uniqueness, let $X = \{x_1, x_2, x_3, x_4\}$ be a four point metric space in which $d(x_i, x_j) = d$ for all $i, j$. Any transference plan between $\mu = \lambda_1\delta[x_1] + \lambda_2\delta[x_2]$ and $\nu = \lambda_3\delta[x_3] + \lambda_4\delta[x_4]$ has equal, optimal, cost, since

$$\mathbb{1}^T \cdot \gamma \odot D \cdot \mathbb{1} = d\left(\mathbb{1}^T \cdot \gamma \cdot \mathbb{1}\right) = d$$

as long as $\gamma$ is a transference plan. ♦

### 3.1.2 Brenier's Theorem

When the cost function is $\int_X d^2(x,y)\,\mathrm{d}\gamma$ the solution to the generalized Kantorovich problem can be shown to be unique, and to in fact be given by pushforward along a function so long as the measures in question are sufficiently regular.

**Theorem 3.1.8** (Brenier)**.** *If $\mu$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$ and $\mu$ and $\nu$ have finite second-moments, then there is a unique optimal transport plan $\gamma$, and*

$$\gamma = (\mathrm{id} \times \nabla\phi)\#\mu$$

*for some convex function* $\phi$*.*

A thorough, and very readable treatment, including multiple proofs, is given in [30, Chapter 2]. The original proof appeared in [34, 35]. The statement can in fact be extended to a slightly more general class of measures that do not give mass to small sets. If the measures are not absolutely continuous, it can still be shown that the support of any optimal $\gamma$ lies on the subdifferential of a convex, lower semi-continuous function $\phi$. (The subdifferential is defined in Chapter 5.)

McCann [36] showed that, with appropriately modified definitions, the same result holds on manifolds:

**Theorem 3.1.9** (McCann)**.** *More generally, if* $M$ *is a compact Riemannian manifold and* $\mu \ll \mathrm{dVol}$*, then there is a unique optimal transference plan*

$$\gamma = (\mathrm{id} \times T)\#\mu$$

*where* $T\colon M \to M$ *satisfies* $T(x) = \exp_x(-\nabla\phi(x))$ *for some geodesically convex function* $\phi$*.*

## 3.2 The Wasserstein Metric

The optimal transport problem assigns a real number to any pair of measures, namely, the cost of the optimal transport plan between them. This section shows how that assignment is actually a metric on the space of probability measures, which gives rise to the Wasserstein space.

Let $P(X)$ denote the set of all Borel probability measures on a Polish (complete, separable) metric space $X$. Define the set of measures with **finite** $p$**-th moment**,

$$\mathscr{P}_p(X) := \left\{\mu \in P(X) \mid \int_X d(x, x_0)^p \,\mathrm{d}\mu < +\infty\right\}$$

for some $x_0 \in X$.

**Proposition 3.2.1.** *The following basic properties of $\mathscr{P}_p(X)$ hold:*

1. *The set $\mathscr{P}_p(X)$ does not depend on the choice of $x_0$.*

2. *If $X$ is a bounded metric space, then $\mathscr{P}_p(X) = P(X)$.*

3. *$\mathscr{P}_p(X)$ contains all finitely-supported probability measures.*

*Proof.* 1. Take any $x_1 \in X$. Then

$$\int_X d(x,x_1)^p \,\mathrm{d}\mu \le \int_X d(x,x_0)^p \,\mathrm{d}\mu + \int_X d(x_0,x_1)^p \,\mathrm{d}\mu,$$

and since $d(x_0,x_1)^p$ is a constant and $\mu$ is a probability measure, $\int_X d(x_0,x_1)^p \,\mathrm{d}\mu = d(x_0,x_1)^p$. Thus the difference between the two integrals is at most a finite amount.

2. Let $c \sup_{x,y\in X} d^p(x,y)$. Then $\int_X d^p(x,x_0)\,\mathrm{d}\mu \ge \int_X c\,\mathrm{d}\mu = c < \infty$.

3. Let $\mu = \sum_{i=1}^n \lambda_i \delta[x_i]$. Then $\int_X d^p(x,x_0) = \sum_{i=1}^n \lambda_i d^p(x_i,x_0) < \infty$ because the sum is finite.

□

**Definition 3.2.2.** *Let $\mu,\nu \in \mathscr{P}_p(X)$. The $p$-**Wasserstein** distance between $\mu$ and $\nu$ is*

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

The Wasserstein distance is a metric on $\mathscr{P}_p(X)$. Since $d$ is a distance, $W_p(\mu,\nu) \ge 0$. Non-degeneracy follows from observing that $W_p(\mu,\nu) = 0$ implies that $\gamma$ is supported precisely on the diagonal $x = y$ in $X \times X$. Thus $\mu = \nu$ since for any measurable function $f\colon X \to \mathbb{R}$,

$$\int_X f(x)\,\mathrm{d}\mu = \int_{X\times X} f(x)\,\mathrm{d}\gamma = \int_{X\times X} f(y)\,\mathrm{d}\gamma = \int_X f(y)\,\mathrm{d}\nu.$$

Symmetry is immediate. The triangle inequality is less clear, depending on two significant lemmas.

**Lemma 3.2.3** (Disintegration of Measure)**.** *Let $X$ and $Y$ be Polish spaces, $\pi_2\colon X\times Y\to Y$ be the projection onto $Y$, $\mu\in\mathscr{P}_p(X\times Y)$, and set $\nu=\pi_2\#\mu$. Then there exists a family of measures $\mu_y\in\mathscr{P}_p(X)$ indexed by $y\in Y$ such that*

$$\mu=\int_Y \mu_y\,\mathrm{d}\nu(y).$$

*Moreover, the map $y\mapsto\mu_y(B)$ is measurable for any measurable $B\subseteq X\times Y$, and the family $\{\mu_y\}$ is uniquely determined $\nu$-almost everywhere.*

A proof can be found in [37, page 78].

**Lemma 3.2.4** (Composition of Transport)**.** *Let $\alpha\in\mathscr{P}_p(X\times Y)$ and $\beta\in\mathscr{P}_p(Y\times Z)$ with*

$$\mu=\pi_2\#\alpha=\pi_1\#\beta\in\mathscr{P}_p(Y).$$

*Then there exists a (not necessarily unique) measure $\sigma\in\mathscr{P}_p(X\times Y\times Z)$ such that $\pi_{1,2}\#\sigma=\alpha$ and $\pi_{2,3}\#\sigma=\beta$.*

Here $\pi_{i,j}$ denotes the projection onto factors $i$ and $j$ of the product.

*Proof.* By disintegration of measure there exist families $\alpha_y$ and $\beta_y$ such that $\alpha=\int_Y\alpha_y\,\mathrm{d}\mu(y)$ and $\beta=\int_Y\beta_y\,\mathrm{d}\mu(y)$. Then $\sigma$ defined as $\sigma:=\int_y(\alpha_y\otimes\beta_y)\,\mathrm{d}\mu(y)$ satisfies the requirements of the lemma. □

**Remark 3.2.5.** It is trivial to construct a measure $\sigma\in\mathscr{P}_p(X\times Y\times Y\times Z)$ with $\pi_{1,2}\#\sigma=\alpha$ and $\pi_{3,4}\#\sigma=\beta$ by simply taking the product $\alpha\otimes\beta$. The significance of Lemma 3.2.4 is that the central index can be “contracted.” A similar composition does not hold for three measures: given $\alpha$ and $\beta$ as in the lemma and $\gamma\in\mathscr{P}_p(X\times Z)$, with $\pi_1\#\gamma=\pi_1\#\alpha$ and $\pi_2\#\gamma=\pi_2\#\beta$, there is in general no $\sigma$ with $\pi_{1,2}\#\sigma=\alpha$, $\pi_{2,3}\#\sigma=\beta$, and $\pi_{1,3}\#\sigma=\gamma$.

**Example 3.2.6.** The construction is clear when working with finitely-supported measures, interpreted as matrices. Let $A$ and $B$ be the matrices with values corresponding to the weights in

$\alpha$ and $\beta$, respectively, and $m$ the vector of weights corresponding to their common marginal. The assumption that $\mu = \pi_2\#\alpha = \pi_1\#\beta$ corresponds to requiring that

$$m = \mathbb{1}^T \cdot A = B \cdot \mathbb{1}.$$

A disintegration $A_y$ is the row of $A$ corresponding to $y$ and dividing it by $m_y$, the weight on $y$ in $m$, and likewise for $B$, except transposed. The joint measure on $X \times Y \times Z$ is then represented by a three-dimensional array whose sheets are the outer product of $A_y$ and $B_y$.

♦

**Proposition 3.2.7.** *$(\mathscr{P}_p(X), W_p)$ is a metric space.*

*Proof.* Non-degeneracy and symmetry have already been shown. To show the triangle inequality, let $\mu, \nu, \rho \in \mathscr{P}_p(X)$, with $\gamma_1$ and $\gamma_2$ the optimal plans for $(\mu, \nu)$ and $(\nu, \rho)$, respectively. By Lemma 3.2.4 there is some $\sigma$ which marginalizes to $\gamma_1$ and $\gamma_2$, and by construction $\pi_{1,3}\#\sigma$ is a transference plan between $\mu$ and $\rho$. Applying the triangle inequality in $X$ and Minkowski's inequality gives:

$$\begin{aligned}
W_p(\mu,\rho) &\le \left(\int_X d^p(x,z)\,\mathrm{d}\pi_{1,3}\#\sigma(x,z)\right)^{1/p} \\
&= \left(\int_X d^p(x,z)\,\mathrm{d}\sigma(x,y,z)\right)^{1/p} \\
&\le \left(\int_X (d(x,y)+d(y,z))^p\,\mathrm{d}\sigma(x,y,z)\right)^{1/p} \\
&\le \left(\int_X d^p(x,y)\,\mathrm{d}\sigma(x,y,z)\right)^{1/p} + \left(\int_X d^p(y,z)\,\mathrm{d}\sigma(x,y,z)\right)^{1/p} \\
&\le \left(\int_X d^p(x,y)\,\mathrm{d}\pi_{1,2}\#\sigma(x,y)\right)^{1/p} + \left(\int_X d^p(y,z)\,\mathrm{d}\pi_{2,3}\#\sigma(y,z)\right)^{1/p} \\
&= W_p(\mu,\nu) + W_p(\nu,\rho).
\end{aligned}$$

Also note that $W_p$ is finite since by the assumption of finite $p$-th moment,

$$+\infty > \int_X d^p(x,x_0)\,\mathrm{d}\mu = \int_{X\times X} d^p(x,y)\,\mathrm{d}(\mu\otimes\delta[x_0])(x,y) = W_p(\mu,\delta[x_0])$$

and then by the triangle inequality, $W_p(\mu, \nu) \leq W_p(\mu, \delta[x_0]) + W_p(\delta[x_0], \nu)$. $\square$

There are many notions of convergence in probability theory, including convergence in probability, almost sure convergence, convergence in total variation, strong convergence, and weak convergence, among others. Convergence in Wasserstein distance could be added to this list, but it turns out to be equivalent to weak convergence under minor assumptions.

Recall that a sequence $\mu_n$ converges to $\mu$ weakly if

$$\lim_{n\to\infty} \int_X f \, \mathrm{d}\mu_n = \int_X f \, \mathrm{d}\mu$$

for all continuous, bounded functions $f$. The nomenclature "weak convergence" is used because most other forms of convergence for probability measures imply weak convergence, and becuase—perhaps more rigorously—weak convergence is convergence in the dual space $C_b(X)^*$ of the space of bounded, continuous functions on $X$ with the weak-* topology, that is, the weakest topology on $C_b(X)^*$ such that the evaluation maps are continuous. By the Riesz representation theorem, probability measures can be identified with a convex subset of $C_b(X)^*$ (though not a vector subspace).

**Proposition 3.2.8.** *Let $X$ be a compact metric space. Then $\lim_{n\to\infty} W_p(\mu_n, \mu) = 0$ if and only if $\mu_n \to \mu$ weakly.*

*Proof.* Suppose first that $W_2(\mu_n, \mu) \to 0$, and let $\gamma_n$ be an optimal transport plan from $\mu_n$ to $\mu$. Then for any $f \in C_b(X)$,

$$\left| \int_X f(x) \, \mathrm{d}\mu_n(x) - \int_X f(y) \, \mathrm{d}\mu(y) \right| = \left| \int_X f(x) - f(y) \, \mathrm{d}\gamma_n(x, y) \right| \leq \int_X |f(x) - f(y)| \, \mathrm{d}\gamma_n(x, y).$$

Assume for the moment that $f$ is $k$-Lipschitz. Then

$$\int_X |f(x) - f(y)| \, \mathrm{d}\gamma_n(x, y) \leq k \int_X d(x, y) \, \mathrm{d}\gamma_n(x, y) = k W_1(\mu_n, \mu).$$

Hölder's inequality gives that

$$\int_X d(x,y)\,\mathrm{d}\gamma_n(x,y) \le \left(\int_X d^p(x,y)\,\mathrm{d}\gamma_n(x,y)\right)^{1/p}\left(\int_X 1\,\mathrm{d}\gamma_n(x,y)\right)^{1/q}$$

for any $p \in (1,+\infty)$ and $q = \frac{p}{p-1}$. Since $\gamma_n$ is a probability measure,

$$\lim_{n\to\infty}\int_X d(x,y)\,\mathrm{d}\gamma_n(x,y) \le \lim_{n\to\infty}\left(\int_X d^p(x,y)\,\mathrm{d}\gamma_n(x,y)\right)^{1/p} = \lim_{n\to\infty} W_p(\mu_n,\mu) = 0.$$

Finally, conclude by recalling that the set of Lipschitz functions is dense in $C_b(X)$ with the sup-norm (i.e. uniform convergence), so convergence must hold for any $f$.

Conversely, if $\mu_n \to \mu$ weakly, assume there is a sequence $\gamma_n \in \Gamma_0(\mu_n,\mu)$ which conveges to some $\gamma$ and this $\gamma$ is an optimal plan.[7] This means $\int d^p(x,y)\,\mathrm{d}\gamma(x,y) = 0$, so that

$$\lim_{n\to\infty} W_p^p = \lim_{n\to\infty}\int d^p(x,y)\,\mathrm{d}\gamma_n(x,y) = \int d^p(x,y)\,\mathrm{d}\gamma(x,y) = 0.$$

□

**Corollary 3.2.9.** *If $X$ is compact, then the spaces $\mathscr{P}_p(X)$ are homeomorphic, i.e. $W_p$ induces the same topology for all $p$.*

*Proof.* Compactness implies boundedness, and by Proposition 3.2.1 $\mathscr{P}_p(X)$ is the same set for all $p$ as long as $X$ is bounded. Then by Proposition 3.2.8, convergence in $W_p$ and $W_q$ are both equivalent to uniform convergence, and therefore equivalent. □

This corollary justifies using $\mathscr{P}(X)$ rather than specifying any particular $p$, as will be done throughout the remainder. Moreover, when a particular $p$ is intended it will always be $p = 2$ for reasons given in Chapter 4, unless specified otherwise.

[7] The elided details are to be found in [38].

## 3.3 Topology and Geometry of Wasserstein Space

Wasserstein space $\mathscr{P}(X)$ inherits a number of topological and geometric properties from $X$. Recall that $X$ is assumed to be locally compact, complete, and separable.

**Corollary 3.3.1.** *If $X$ is compact then $\mathscr{P}(X)$ is compact.*

*Proof.* It suffices to show that $\mathscr{P}(X)$ is closed and bounded since it is a metric space. If $X$ is bounded, then $d^2(-,-)$ is bounded and so $W_2$ is bounded. By Proposition 3.2.8, $\mathscr{P}(X)$ is closed.

□

**Proposition 3.3.2.** *If $X$ is a geodesic space, then $\mathscr{P}(X)$ is a geodesic space.*

*Proof.* The construction of the geodesics is given here. A complete proof is found at [38, Theorem 3.10]. Define the space $\mathrm{Geo}(X)$ to be the set of all constant-speed geodesics $\gamma\colon [0,1] \to X$ with the sup-norm. Let $G\colon X \times X \to \mathrm{Geo}(X)$ be the map $(x, y) \mapsto \gamma_y^x(t)$ selecting a geodesic connecting $x$ and $y$.[8] There is also an evaluation map $e\colon [0,1] \times \mathrm{Geo}(X) \to X$ defined by $e_t(\gamma) = \gamma(t)$. Let $\mu$ and $\nu$ be any measures in $\mathscr{P}(X)$ with $\gamma$ and optimal plan connecting them. Define $\Theta :=$ $G\#\gamma \in \mathscr{P}(\mathrm{Geo}(X))$ and $\theta(t) := e_t\#\Theta \in \mathscr{P}(X)$. The constant speed geodesic between $\mu$ and $\nu$ is $\theta(t)\colon [0,1] \to \mathscr{P}(X)$. □

**Example 3.3.3.** Consider the case where $\mu$ and $\nu$ are finitely-supported, $\mu = \sum_{i=0}^n \lambda_i \delta[x_i]$ and $\nu = \sum_{i=0}^m \xi_j \delta[y_j]$. An optimal transference plan $\gamma$ has the form $\gamma = \sum_{i,j} \omega_{i,j} \delta\left[(x_i, y_j)\right]$ where some $\omega_{i,j}$ may be zero. Then

$$G\#\gamma = \sum_{i,j} \omega_{i,j} \delta\left[\gamma_{y_j}^{x_i}\right] \text{ and } \theta(t) = \sum_{i,j} \omega_{i,j} \delta\left[\gamma_{y_j}^{x_i}(t)\right].$$

The difference between the last two formulas is that in the first $\delta\left[\gamma_{y_j}^{x_i}\right]$ is the point mass at $\gamma_{y_j}^{x_i}$ in the space of geodesics, while in the second $\delta\left[\gamma_{y_j}^{x_i}(t)\right]$ is the point mass at $\gamma_{y_j}^{x_i}(t)$ in $X$.

This formula for the geodesics between discrete measures will play a significant role in Chapter 4. ♦

[8] This map is not necessarily uniquely defined, in which case the geodesics in $\mathscr{P}(X)$ are also not unique.

### 3.3.1 Vietoris–Rips Metric Thickenings as Wasserstein Spaces

Section 2.3 gave the definition of Vietoris–Rips metric thickenings and defined a Wasserstein metric on them. The relation between this definition and the Wasserstein metric in this chapter is made explicit here.

Let $\mathrm{VR}(X;r)$ be a Vietoris–Rips complex and suppose that some point $m \in \mathrm{VR}(X;r)$ is given in barycentric coordinates by the formal sum $m = \sum_{i=0}^{n} \lambda_i x_i$. Define a map $\Phi\colon \mathrm{VR}(X;r) \to \mathscr{P}(X)$ by $\sum_{i=0}^{n} \lambda_i x_i \mapsto \sum_{i=0}^{n} \lambda_i \delta[x_i]$, that is, turning the formal sum into a convex sum of probability distributions. The image of $\Phi$ in $\mathscr{P}(X)$ is the Vietoris–Rips metric thickening $\mathscr{VR}(X;r)$. Specifically, $\mathscr{VR}(X;r)$ as defined in Section 2.3 is the image in $\mathscr{P}_1(X)$; however, this is homeomorphic to the image in any other $\mathscr{P}_p(X)$. Henceforth $\mathscr{VR}(X;r)$ will denote the image in $\mathscr{P}_2(X)$ unless otherwise stated.

The Vietoris–Rips thickening is not typically a geodesically convex subset.

**Example 3.3.4.** Consider $\mathscr{VR}\left(\mathbb{S}^1; \frac{2\pi}{3}+\varepsilon\right)$ where $\varepsilon$ is small ($0<\varepsilon<\frac{2\pi}{15}$). Let $\mu = \frac{1}{3}\delta[0] + \frac{1}{3}\delta\left[\frac{2\pi}{3}\right] + \frac{1}{3}\delta\left[\frac{4\pi}{3}\right]$ and $\nu = \delta[0]$. The constant speed geodesic in $\mathscr{P}(X)$ from $\mu$ to $\nu$ is

$$\gamma_\nu^\mu(t) = \frac{1}{3}\delta\left[(1-t)\frac{2\pi}{3}\right] + \frac{1}{3}\delta\left[2\pi t + (1-t)\frac{4\pi}{3}\right] + \frac{1}{3}\delta[0].$$

Note that $\mathrm{diam}(\gamma_\nu^\mu(\frac{1}{4})) = \pi > \frac{2\pi}{3}+\varepsilon$, and so $\gamma_\nu^\mu(t)$ is not contained in $\mathscr{VR}\left(\mathbb{S}^1; \frac{2\pi}{3}+\varepsilon\right)$. ♦

The definition of geodesics in Wasserstein space, and of the Vietoris–Rips metric thickening, emphasizes that there are two distinct concepts of "straight line" at play. On the one hand, there are geodesics $\gamma_\nu^\mu(t)$ in $\mathscr{P}(X)$, at least when $X$ is geodesic, and geodesics broadly speaking generalize the idea of a straight line. On the other hand, there are convex combinations $(1-t)\mu + t\nu$ for any $t \in [0,1]$ which also connect $\mu$ to $\nu$ in an intuitively "linear" way. Viewing $\mathscr{P}(X)$ as a subset of $C_b(X)^*$, the dual function space, the latter is precisely the linear vector space structure there. At first glance the linear structure is more compatible with $\mathscr{VR}(X;r)$, since if $\sigma$ and $\tau$ are faces of the same simplex in $\mathrm{VR}(X;r)$, then $(1-t)\Phi(\sigma) + t\Phi(t)$ is in $\mathscr{VR}(X;r)$. Chapter 4 however determines the homotopy type of $\mathscr{VR}(X;r)$ using only the geodesic structure. The interplay between these two elements of Wasserstein space is a significant, if not always explicit, theme in this thesis.

# Chapter 4

# Results

This chapter uses the Wasserstein space structure on $\mathcal{VR}(X;r)$ to construct a homotopy equivalence between $\mathcal{VR}(X;r)$ and $X$ under various curvature constraints on $X$. A main tool is the center of mass, discussed in Section 4.1. Curvature for metric spaces is defined in Section 4.2 and necessary curvature bounds imposed. The main theorem is in Section 4.3. This extends the results of [3] by using a homotopy along geodesics within Wasserstein space, rather than along the linear structure of the simplices. Section 4.4 concludes the chapter with new corollaries arising from the geodesic approach.

## 4.1 The Center of Mass

The center of mass of a measure $\mu$ on $\mathbb{R}^d$ is defined by

$$\overline{\mu} := \int_{\mathbb{R}^n} \vec{x}\, \mathrm{d}\mu(\vec{x}).$$

This is finite for any probability measure with finite first moment, in particular for any $\mu \in \mathscr{P}(X)$. In the case that $\mu = \sum_{i=0}^{n} \lambda_i \delta\left[x_i\right]$ is finitely supported, its center of mass is

$$\overline{\mu} = \sum_{i=0}^{n} \lambda_i \vec{x}_i.$$

The center of mass satisfies the property that

$$\overline{\mu} = \underset{\vec{y}\in\mathbb{R}^n}{\operatorname{argmin}} \int_{\mathbb{R}^n} \|\vec{x}-\vec{y}\|^2\, \mathrm{d}\mu(\vec{x}),$$

and it can also be understood statistically as the expected value of the distribution $\mu$. The center of mass gives a convenient map from $\mathscr{P}(\mathbb{R}^d)$ to $\mathbb{R}^d$.

Hermann Karcher showed how to generalize this construction to a Riemannian manifold or more general metric space with his introduction of the Riemannian center of mass in [39, 40]. The exposition that follows is drawn from [40] and [41]. A measurable function $f\colon A \to M$ from some probability space $(A, \mathscr{F}, \mathbb{P})$ to a Riemannian manifold $M$ induces a measure $\mu = f\#\mathbb{P}$ on $M$. Define a function $P_f\colon M \to \mathbb{R}$ by

$$P_f(m) \coloneqq \frac{1}{2}\int_A d(m, f(a))^2 \, \mathrm{d}\mathbb{P}(a) = \frac{1}{2}\int_M d(m,x)^2 \, \mathrm{d}\mu(x).$$

(The subscript will denote either the function or the induced measure depending on context, so here $P_f = P_\mu$.) The function $P_f$ is the weighted sum of squared distances from $m$ to the support of $\mu$. Let $\mathrm{B}_\rho(m)$ be the ball of radius $\rho$ centered at $m \in M$. If $\rho$ is sufficiently small $\mathrm{B}_\rho(m)$ is geodesically convex: there is a unique shortest geodesic between any points $x$ and $y$ in $\mathrm{B}_\rho(m)$, and the entirety of this geodesic is contained in $\mathrm{B}_\rho(m)$.

**Theorem 4.1.1** (Karcher)**.** *Suppose that* $\mathrm{supp}(\mu)$ *is contained inside a geodesically convex ball* $\mathrm{B}_\rho(m)$ *and that the sectional curvature of $M$ in* $\mathrm{B}_\rho(m)$ *is either*

- *at most* $0$*, or*
- *bounded above by some* $\Delta > 0$ *and* $\rho < \frac{1}{4}\pi\Delta^{-\frac{1}{2}}$.

*Then* $P_f\colon \overline{\mathrm{B}}_\rho(m) \to \mathbb{R}$ *is (geodesically) convex and therefore has a unique minimum* $C_f$ *in* $\mathrm{B}_\rho(m)$.

The proof is found in [40, Theorem 1.2]. The point $C_f$ may variously be called the Riemannian center of mass, the "Karcher mean," the "Fréchet mean," or the barycenter of $\mu$. Here $C_f$ will be called the center of mass, since "Karcher mean" is disputed [42] and barycenter has another meaning in the setting of simplicial complexes.

While Theorem 4.1.1 depends on the Riemannian structure of $M$, neither the definition of $P_f$ nor $C_f$ does so.

**Definition 4.1.2.** *Let $X$ be a metric space and $\mu \in \mathscr{P}(X)$. Define*

$$P_\mu(x) = \int_X d^2(x,y)\,\mathrm{d}\mu(y)$$

*as in the Riemannian setting, and define*

$$V(\mu) = \inf_x P_\mu(x), \quad K(\mu) = \operatorname*{argmin}_x P_\mu(x).$$

A priori, *$K(\mu)$ is set-valued and $K(\mu) = \emptyset$ is possible.*

Recall that a metric space is **proper** if every closed, bounded subspace is compact.

**Lemma 4.1.3.** *If $X$ is a proper metric space, then $K(\mu)$ is nonempty for every $\mu \in \mathscr{P}(X)$.*

*Proof.* The proof follows that in [43, Lemma 3.2]. Choose some $x_0 \in X$ and $r > 1$ such that $\mu(\mathrm{B}_r(x_0)) \geq \frac{1}{2}$ and

$$\int_{X \setminus \mathrm{B}_r(x_0)} d^2(x_0,x)\,\mathrm{d}\mu(x) \leq 1.$$

Such a choice is possible by the assumption that the second moment of $\mu$ is finite. This implies that

$$\int_{\mathrm{B}_r(x_0)} d^2(x_0,x)\,\mathrm{d}\mu(x) \leq r^2\mu(\mathrm{B}_r(x_0))$$

since $d(x,x_0) \leq r$ in $\mathrm{B}_r(x_0)$. This gives

$$\int_X d^2(x_0,x)\,\mathrm{d}\mu(x) \leq r^2\mu(\mathrm{B}_r(x_0)) + 1 \leq r^2 + 1.$$

If $y \in X \setminus \mathrm{B}_{3r}(x_0)$, then

$$\int_X d^2(y,x)\,\mathrm{d}\mu(x) \geq \int_{\mathrm{B}_r(x_0)} d^2(y,x)\,\mathrm{d}\mu(x) > (2r)^2\mu(\mathrm{B}_r(x_0)) \geq 2r^2$$

where the first inequality holds because $X \supseteq \mathrm{B}_r(x_0)$ and the second because $y \in \mathrm{B}_{3r}(x_0)$ and therefore $\inf_{x\in\mathrm{B}_r(x_0)} d^2(x,y) = 2r$. Hence the optimal point necessarily is contained in $\overline{\mathrm{B}}_{3r}(x_0)$.

With the assumption that $X$ is proper, $\overline{\mathrm{B}}_{3r}(x_0)$ is compact, and therefore $P_\mu$ achieves its minimum. □

If $X$ happens to be globally non-positively curved (see Section 4.2), then this can be improved:

**Proposition 4.1.4.** *If $X$ is globally non-positively curved and $\mu$ is supported in a convex closed set $V \subseteq X$, then $K(\mu)$ is in $V$.*

For proof, see [44, Proposition 6.1]. The idea is to project onto the boundary of the region $V$ and show that doing so reduces the integral. Non-positive curvature gives a canonical projection onto $V$, and guarantees that the distance from the projection to any point in the interior of $V$ is less than the original distance.

Proposition 4.1.4 does not hold in positive curvature. Let $H$ be a closed hemisphere on the 2-sphere. Certainly $H$ is geodesically convex, however, if $\mu$ is supported on two antipodal points on the equator, the pole opposite $H$ is a mean.

In order to use $K$ as a map $\mathscr{P}(X) \to X$ we would like to further establish some type of uniqueness, as well as continuity. Uniqueness will generally only hold for measures supported on sufficiently small subsets, and continuity of course depends upon uniqueness. Geometric considerations on the space $X$ will determine the necessary conditions.

## 4.2 Curvature Considerations

Existence of a minimizer requires only the minor assumption that $X$ is proper. Uniqueness is generally much harder, but one way to guarantee the uniqueness of a minimizer is to show that $P_\mu$ is a convex function. If $X$ is a geodesic space, then convexity of a function $f$ means that for any pair of points $x$ and $y$ and any geodesic $\gamma_y^x \colon [0,T] \to X$ connecting them, the inequality

$$f(\gamma_y^x(t)) \leq (1-t)f(x) + tf(y)$$

is satisfied. Convex functions have a unique minimum whenever their minimum is achieved, and a unique argmin if the inequality is strict whenever $t \in (0,T)$. Since $P_\mu$ corresponds to a sum of squared distances, convexity of $d^2(-,-)$ is closely related to the convexity of $P_\mu$, and the convexity of the distance function is closely related to the curvature of the space $X$.

Curvature for geodesic spaces is defined by comparisons to reference manifolds of constant curvature, in the simplest case, the Euclidean plane. Given a point $p$ and a geodesic curve $\gamma(t)$ in $X$ which is parametrized by arc length, define the comparison function $g(t) = d(p,\gamma(t))$. Any such function has a Euclidean comparison function $\vec{g}(t) = \|\vec{p} - \vec{\gamma}(t)\|$ defined by choosing a point $\vec{p}$ in the Euclidean plane and a parametrized straight line segment $\vec{\gamma}(t)$ such that $\|\vec{p} - \vec{\gamma}(0)\| = d(p,\gamma(0))$ and $\|\vec{p} - \vec{\gamma}(1)\| = d(p,\gamma(1))$.

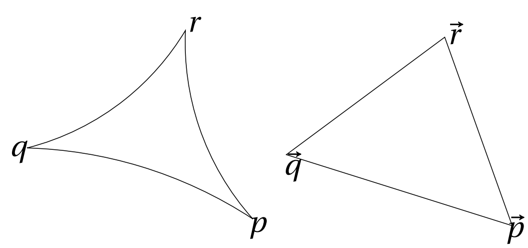


**Figure 4.1:** A triangle in a geodesic space and the comparison Euclidean triangle.

**Definition 4.2.1.** *A neighborhood $N_x$ of a point $x$ in a geodesic space $X$ is* ***non-positively curved*** *(respectively,* ***non-negatively curved****) if given any point $p \in N_x$ and a geodesic $\gamma \subseteq N_x$, the comparison function satisfies $g(t) \leq \vec{g}(t)$ (respectively, $g(t) \geq \vec{g}(t)$).*

A neighborhood $N_x$ satisfying this condition is called a **normal region**. If it is possible to take $N_x = X$ then $X$ is **globally** non-positively or non-negatively curved. If $X$ is complete and every $x \in X$ has a normal region satisfying the same curvature inequality, then $X$ is **locally** non-positively curved (respectively, non-negatively curved). Equivalent to Definition 4.2.1 is the triangle condition:

**Definition 4.2.2.** *Let $\triangle pqr$ be a triangle in $X$, that is, a set of three points $p$, $q$, and $r$, and a designated geodesic between each. Let $\triangle \vec{p}\vec{q}\vec{r}$ be a triangle in Euclidean space with the same side lengths. The geodesic space $X$ satisfies the* ***narrow triangle condition*** *if for every point $x$ on the geodesic connecting $p$ and $r$, $d(x,q) \le \|\vec{x}-\vec{q}\|$, and the* ***wide triangle condition*** *if $d(x,q) \ge \|\vec{x}-\vec{q}\|$.*

A direct consequence of the definition is that in any space of non-positive curvature the distance function is convex, meaning that for any geodesics $\gamma$ and $\tilde{\gamma}$,

$$d(\gamma(t),\tilde{\gamma}(t)) \le (1-t)d(\gamma(0),\tilde{\gamma}(0)) + td(\gamma(1),\tilde{\gamma}(1)).$$

Examples of spaces of global non-positive curvature, hereafter global NPC, include manifolds which are complete, simply-connected, and have non-positive sectional curvature [44]. (The major example is hyperbolic space.) Locally finite metric trees, that is, simply connected metric graphs, are non-positively curved. Hilbert spaces are also global NPC spaces, and in fact, this condition distinguishes Banach spaces from Hilbert spaces in the sense that a Banach space is global NPC space if and only if it is a Hilbert space. These and many other examples are found in [44, Section 3].

A key example of a space which is locally NPC but not globally NPC is the circle. Any open metric ball of radius less than $\pi$ in $\mathbb{S}^1$ is isometric to a segment of the real line and therefore satisfies the condition of Definition 4.2.1; however, the circle is the canonical example of a space of global positive curvature. More generally, any locally finite metric graph is locally NPC, since a sufficiently small neighborhood of any point resembles a tree. If $\phi\colon \mathbb{R} \to \mathbb{R}$ is a smooth convex function, then its surface of revolution is a non-compact Riemannian manifold with negative sectional curvature and non-trivial fundamental group. It is therefore not globally NPC, but it is locally NPC. A concrete example is the pseudo-sphere, the surface of constant negative curvature obtained as the surface of revolution of $f(x) = \log\left(\frac{1-\sqrt{1-x^2}}{x} - \sqrt{1-x^2}\right)$. Additional examples can be found in [45, Chapter 4].

Any complete Riemannian manifold with positive sectional curvature at each point is globally non-negatively curved in the sense of Definition 4.2.1. In particular, spheres have constant positive curvature and the sphere $\mathbb{S}^n$ is the only manifold of dimension $n$ with this property. Spaces of non-negative curvature are discussed extensively in Chapter 5. A particularly notable example is that the Wasserstein space $\mathscr{P}(X)$ has non-negative curvature if and only if $X$ does. (See Theorem 5.1.2.)

A more refined notion of curvature bounds is that of a CAT$(\kappa)$ space. In a CAT$(\kappa)$ space, the Euclidean comparison triangles of Definition 4.2.2 are replaced with standard comparison triangle in other spaces of constant curvature.

**Definition 4.2.3.** *The reference spaces of dimension $n$ and constant curvature $\kappa$, denoted $\mathbb{M}^n(\kappa)$ are the metric spaces:*

*if $\kappa = 0$ then $\mathbb{M}^n(0)$ is Euclidean space $\mathbb{R}^n$,*

*if $\kappa > 0$ then $\mathbb{M}^n(\kappa)$ is obtained from the unit $n$-sphere $\mathbb{S}^n$ by multiplying the distance function by $\frac{1}{\sqrt{\kappa}}$, and*

*if $\kappa < 0$ then $\mathbb{M}^n(\kappa)$ is obtained from hyperbolic space $\mathbb{H}^n$ by multiplying the distance function by $\frac{1}{\sqrt{-\kappa}}$.*

When $\kappa > 0$ the space $\mathbb{M}^n(\kappa)$ can also be described as the sphere of radius $\frac{1}{\sqrt{\kappa}}$.

**Definition 4.2.4.** *Let $\triangle pqr$ be a triangle in a geodesic space $X$ and $\triangle \bar{p}\bar{q}\bar{r}$ be a triangle identical side lengths in $\mathbb{M}^n(\kappa)$. The triangle $\triangle pqr$ satisfies the* CAT$(\kappa)$***-inequality*** *if for any point $x$ on the geodesic connecting $p$ and $r$ and $\bar{x}$ on the geodesic connecting $\bar{p}$ and $\bar{r}$ with $d(p,x) = d(\bar{p},\bar{x})$, the inequality $d(x,q) \le d(\bar{x},\bar{q})$ holds.*

A geodesic space has **Alexandrov curvature** $\le \kappa$ if for every $x \in X$ there is a ball $\mathrm{B}_r(x)$ in which all triangles satisfy the CAT$(\kappa)$-inequality. If all triangles in $X$ satisfy the CAT$(\kappa)$-inequality, then $X$ is a CAT$(\kappa)$**-space**.

**Lemma 4.2.5.** *Let $U$ be a* $\mathrm{CAT}(\kappa)$*-space, or more generally a geodesically convex ball satisfying the* $\mathrm{CAT}(\kappa)$*-inequality, and $\triangle pqr$ a triangle in $U$. Let $w(t) = d(\gamma_q^p(t), \gamma_q^r(t))$ be the distance between the geodesics from $p$ and $r$ to $q$, scaled so that $\gamma_q^p(1) = \gamma_q^r(1) = q$, the "width" of the triangle. Then for all $t \in [0,1]$, $w(t) \le d(p,r)$.*

*Proof.* In non-positive curvature this is immediate from the convexity of the distance function. In positive curvature it suffices to prove the lemma on the unit $n$-sphere. Recall that any ball of radius at most $\frac{\pi}{2}$ is convex. Let $a = d(p,q)$ and $b = d(r,q)$ so that $d(\gamma_q^p(t), q) = a - at$ and $d(\gamma_q^r(t), q) = b - bt$. Let $\theta$ be angle at $q$ between the geodesics $\gamma_q^p$ and $\gamma_q^r$. The result follows from the spherical law of cosines:

$$\cos(w(t)) = \cos(a-at)\cos(b-bt) + \sin(a-at)\sin(b-bt)\cos(\theta)$$

which shows that $w(t)$ is non-increasing.

In detail,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} w(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \arccos\left(\cos(a-at)\cos(b-bt) + \sin(a-at)\sin(b-bt)\cos(\theta)\right) \\
&= -\frac{(a - b\cos(\theta))\sin(a-at)\cos(b-bt) + (b - a\cos(\theta))\cos(a-at)\sin(b-bt)}{\sqrt{1 - (\cos(\theta)\sin(a-at)\sin(b-bt) + \cos(a-at)\cos(b-bt))^2}} \\
&\le 0
\end{aligned}$$

since $a, b \in [0, \frac{\pi}{2}]$ and $\theta \in [0, \pi]$. □

The notion of Alexandrov curvature is closely related to the classical definition of curvature on a Riemannian manifold:

**Theorem 4.2.6.** *A smooth Riemannian manifold has Alexandrov curvature $\le \kappa$ if and only if it has sectional curvature $\le \kappa$.*

*Proof.* See [46, Theorem 1A.6]. □

### 4.2.1 The Center of Mass in Non-positive Curvature

Because of the inherent convexity of the distance function, the center of mass in NPC spaces is relatively easy to determine. Throughout, let $N$ be a global NPC space or a normal region of a metric space $X$ on which the non-positive curvature condition of Definition 4.2.1 is satisfied. (In the latter case $N$, viewed as a metric space on its own right, is global NPC.)

**Proposition 4.2.7.** *Let $\mu \in \mathscr{P}(N)$. Then there is a unique $K(\mu) \in N$ and $K\colon \mathscr{P}(N) \to N$ is continuous.*

*Proof.* Proposition 4.3 in [44]. The proof consists in showing that $K$ is uniformly convex. □

A direct proof, in the case where $\mu = \sum_{i=0}^{n} \lambda_i \delta[x_i]$ is a finitely-supported measure, is also possible. Here $P_\mu$ is certainly continuous, since it is a weighted sum of the distance function, which is continuous. If $N = \mathbb{R}^n$, then $P_\mu$ is convex. This can be shown by computing the second derivative, or by applying the definition directly. Doing the latter gives, for any $m_1$ and $m_2$ in $\mathbb{R}^n$,

$$\begin{aligned}
P_\mu((1-t)m_1 + tm_2) &= \sum_{i=0}^{n} \lambda_i \|(1-t)m_1 + tm_2 - x_i\|^2 \\
&= \sum_{i=0}^{n} \lambda_i \|(1-t)m_1 + tm_2 - (1-t)x_i - tx_i\|^2 \\
&= \sum_{i=0}^{n} \lambda_i \|(1-t)(m_1 - x_i) + t(m_2 - x_i)\|^2 \\
&\le \sum_{i=0}^{n} \lambda_i \|(1-t)(m_1 - x_i)\|^2 + \|t(m_2 - x_i)\|^2 \\
&= \sum_{i=0}^{n} \lambda_i (1-t)^2 \|m_1 - x_i\|^2 + t^2 \|m_2 - x_i\|^2 \\
&< \sum_{i=0}^{n} \lambda_i (1-t) \|m_1 - x_i\|^2 + t\|m_2 - x_i\|^2 \\
&= (1-t)P_\mu(m_1) + tP_\mu(m_2),
\end{aligned}$$

establishing convexity.

Now let $N$ be any global NPC space. Let $m_1$ and $m_2$ be in $N$ and $\gamma_{m_2}^{m_1}$ be the geodesic connecting them. Then for $t \in (0,1)$,

$$\begin{aligned}
P_\mu(\gamma_{m_2}^{m_1}(t)) &= \sum_{i=0}^{n} \lambda_i d^2(x_i, \gamma_{m_2}^{m_1}(t)) \\
&\leq \sum_{i=0}^{n} \lambda_i \|\hat{x}_i - \hat{\gamma}_{\hat{m}_2}^{\hat{m}_1}(t)\|^2 \\
&< \sum_{i=0}^{n} \lambda_i (1-t)\|\hat{m}_1 - \hat{x}_i\| + t\|\hat{m}_2 - \hat{x}_i\|^2
\end{aligned}$$

where the first inequality follows from curvature, and the second one is the same as in the Euclidean case. This establishes that $P_\mu$ is strictly convex on $N$, so it has a unique minimum on $N$. This proof clearly shows the importance of the curvature assumption, since in positive curvature the necessary inequality would point the wrong way.

There is in fact a stronger statement than continuity available when $N$ is globally non-positively curved, namely that $K$ is 1-Lipschitz (with respect to the 1-Wasserstein metric). This follows as a consequence of Jensen's inequality for global NPC spaces [44, Theorem 6.3].

### 4.2.2 The Center of Mass in Non-negative Curvature

Determining the existence and uniqueness of centers of mass in non-negative curvature is generally more difficult. Since the distance function is not globally convex, some assumptions on the diameter of the support of the measure are necessary.

**Example 4.2.8.** Let $x$ and $x^-$ be antipodal points on the sphere, and take $\mu = \frac{1}{2}\delta[x] + \frac{1}{2}\delta[x^-]$. Then there is an equator of the sphere, every point on which is a minimizer of $P_\mu$ (see Section 4.2.2). ♦

It is for this reason that Theorem 4.1.1 imposes a diameter bound on the measure $\mu$. However, even more care is necessary. While Theorem 4.1.1 guarantees the existence of a unique mean on some $\mathrm{B}_\rho(m)$ containing the support of $\mu$, it does not guarantee that mean is the global mean, and *a priori* there may be another ball $\mathrm{B}_\rho(m')$ also containing $\mathrm{supp}(\mu)$ with a different

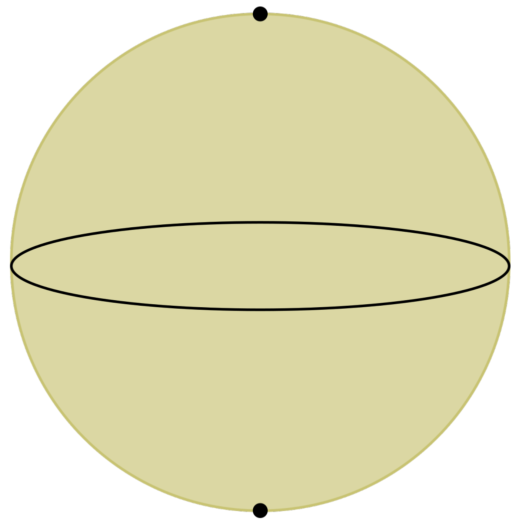

**Figure 4.2:** Every point on the equator is a mean of the poles.

local minimum. The mean guaranteed by Theorem 4.1.1 has thus been called the "solipsistic Karcher mean" since it is the mean if nothing outside of $\mathrm{B}_\rho(m)$ exists [47, 48]. The literature on Riemannian centers of mass has since improved upon Theorem 4.1.1. A stronger statement than Karcher's which guarantees a global minimum is the following theorem of Afsari [48].

**Theorem 4.2.9.** *Let $M$ be a complete Riemannian manifold with sectional curvature $\leq \Delta$ and injectivity radius $I$. Define*

$$\rho_\Delta := \frac{1}{2}\min\left\{I, \frac{\pi}{\sqrt{\Delta}}\right\}.$$

*Suppose that $\mu$ is a probability measure on $M$ with $\operatorname{supp}(\mu) \subseteq \mathrm{B}_{\rho_\Delta}(m)$ for some $m \in M$. Then there is a (globally) unique minimum of $P_\mu$ and that minimum is contained in $\mathrm{B}_{\rho_\Delta}(m)$.*

The present work does not depend on the exact bound, only that there be some bound below which centers of mass are unique, and that the location of the center of mass lies within a given region. Therefore, let $\rho(M)$ be the minimal radius guaranteeing both of these for the manifold $M$, so $\rho(M) \geq \rho_\Delta$.

Karcher's result bounds the distance between the centers of mass of two different distributions:

**Lemma 4.2.10.** *If $f, g\colon A \to \mathrm{B}_\rho(m)$ are measurable, $A$ is a probability space, and $\mathrm{B}_\rho(m)$ satisfies the conditions of Theorem 4.1.1, then*

$$d(C_f, C_g) \leq (1 + c(\delta,\Delta)\cdot(2\rho)^2)\cdot \int_A d(f(a), g(a))\,\mathrm{d}\mathbb{P}(a),$$

*where $c(\delta,\Delta)$ is some constant depending on the curvature bounds $\delta$ and $\Delta$.*

This is [40, Corollary 1.6]. An immediate and useful lemma is the following:

**Lemma 4.2.11.** *Suppose $\mathrm{B}_r(m)$ where $r \leq \rho(M)$. Then $K\colon \mathscr{P}(\mathrm{B}_\rho(m)) \to \mathrm{B}_\rho(m)$ is continuous.*

*Proof.* Let $\mu$ and $\nu$ be in $\mathscr{P}(\mathrm{B}_\rho(m))$. Let $A = \operatorname{supp}(\mu)\times\operatorname{supp}(\nu)$ and make $A$ a probability space with reference measure $\gamma$ for some $\gamma \in \Gamma_0(\mu,\nu)$. That is, $\gamma$ is an optimal transport plan between $\mu$ and $\nu$. Then $\pi_1$ and $\pi_2$ are measurable maps $A \to \mathrm{B}_\rho(m)$ with $K(\mu) = C_{\pi_1}$ and $K(\nu) = C_{\pi_2}$. Then Lemma 4.2.10 gives

$$\begin{aligned} d(K(\mu), K(\nu)) &\leq \left(1 + c(\delta,\Delta)\cdot(2\rho)^2\right)\cdot \int_A d(\pi_1(a), \pi_2(a))\,\mathrm{d}\gamma(a) \\ &= \left(1 + c(\delta,\Delta)\cdot(2\rho)^2\right)\cdot \int_{\mathrm{B}_\rho(m)} d(x,y)\,\mathrm{d}\gamma(x,y) \\ &= \left(1 + c(\delta,\Delta)\cdot(2\rho)^2\right) W_1(\mu,\nu) \end{aligned}$$

establishing the continuity of $K$ with regard to $W_1$, and therefore $W_p$ for all $p$. □

An interesting example is the unit circle, $\mathbb{S}^1$. Here the center of mass can be explicitly computed for any measure $\mu$ with $\operatorname{diam}(\operatorname{supp}(\mu)) < \pi$, so that $\operatorname{supp}(\mu) \subseteq \mathrm{B}_{\frac{\pi}{2}}(m)$ for some $m \in \mathbb{S}^1$. Note that Karcher's theorem only guarantees a unique minimum up to $\rho = \frac{\pi}{4}$. Identify $\mathbb{S}^1$ with the unit circle in $\mathbb{C}$, that is, $\{e^{i\theta} \mid \theta \in [0, 2\pi)\}$. Choose a branch of $\log\colon \mathbb{S}^1 \to \mathbb{R}$ such that $B = \mathrm{B}_{\frac{\pi}{2}}(m)$ is mapped isometrically into $\mathbb{R}$ and denote its inverse by exp.[9] Then

$$K(\mu) = \exp\left(\operatorname*{argmin}_{\theta\in\log(B)} \int_{\log(B)} |\theta-\phi|^2\,\mathrm{d}\log\#\mu(\phi)\right).$$

[9]Here isometric is meant in the sense of metric spaces, not of Riemannian manifolds.

More intuitively, the center of mass of $\mu$ can be computed by treating $\mu$ as a measure on the real line and identifying angular coordinates on $\mathbb{S}^1$ with the corresponding real numbers. So long as $\mu$ is supported on a subset of diameter less than $\pi$, all distances within the support of $\mu$ are computed without wrapping around the circle in the opposite direction, so this is an isometry.

**Example 4.2.12.** While the conditions imposed by Theorem 4.2.9 are sufficient to guarantee the existence and uniqueness of centers of mass, they are not necessary. Consider a distribution on the circle $\mu = a\delta[0] + b\delta[\pi] + c\delta[\pi+\varepsilon]$ where coordinates are given in $[0, 2\pi)$ and $\varepsilon > 0$. This is not supported in an open hemisphere and therefore fails the assumptions given above. Writing coordinates on the circle as angles,

$$P_\mu(\theta) = \begin{cases} a\theta^2 + b(\pi-\theta)^2 + c(\theta+\pi-\varepsilon)^2 & 0 \le \theta < \varepsilon \\ a\theta^2 + b(\pi-\theta)^2 + c(\pi+\varepsilon-\theta)^2 & \varepsilon < \theta \le \pi \\ a(2\pi-\theta)^2 + b(\theta-\pi)^2 + c(\pi+\varepsilon-\theta)^2 & \pi < \theta < 2\pi \end{cases}$$

Thus,

$$\frac{\mathrm{d}P_\mu}{\mathrm{d}\theta} = \begin{cases} 2a\theta + 2b(\theta-\pi) + 2c(\theta+\pi-\varepsilon) & 0 \le \theta < \varepsilon \\ 2a\theta + 2b(\theta-\pi) + 2c(\theta-\pi-\varepsilon) & \theta \le \pi \\ 2a(\theta-2\pi) + 2b(\theta-\pi) + 2c(\theta-\pi-\varepsilon) & \pi < \theta < 2\pi \end{cases}$$

The critical points are

$$\theta = 0, \varepsilon, \pi, (b-c)\pi + c\varepsilon, (b+c)\pi + c\varepsilon, \text{ and } \theta = (1+a)\pi + c\varepsilon.$$

Choosing $a = b = c = \frac{1}{3}$ and $\varepsilon = \frac{1}{10}$, the unique minimum occurs at $(1+a)\pi + c\varepsilon = \frac{38}{9}$. ♦

## 4.3 Hausmann-Type Theorems for Metric Thickenings

Recall that Hausmann's theorem (Theorem 2.2.3) says that when $r$ is sufficiently small and $M$ is a manifold, $\mathrm{VR}(M; r) \simeq M$. This section gives proofs that $\mathcal{VR}(X; r) \simeq X$ for certain bounds on $r$ in both positive and negative curvature. These proofs are all modelled on the same format, illustrated first for convex subsets of Euclidean space, that is, the zero curvature setting.

**Example 4.3.1.** Let $\Omega \subseteq \mathbb{R}^n$ be convex, and consider the Vietoris–Rips thickening $\mathcal{VR}(\Omega; r)$. Any $\mu = \sum_{i=0}^{n} \lambda_i \delta[x_i]$ has a mean $K(\mu) = \overline{\mu} = \sum_{i=0}^{n} \lambda_i x_i$. There is a geodesic $\gamma_\mu(t)$ in $\mathscr{P}(\Omega)$ connecting $\mu$ and $\delta\left[\overline{\mu}\right]$. Explicitly this is given by

$$\gamma_\mu(t) = \sum_{i=0}^{n} \lambda_i \delta\left[(1-t)x_i + t\overline{\mu}\right]$$

as seen in Section 3.3. That this curve lies in $\mathscr{P}(\Omega)$ is due to convexity; and in fact it lies in $\mathcal{VR}(\Omega; r)$ since it clearly remains a finitely-supported measure, and

$$\operatorname{diam}(\gamma(t)) = (1-t)\operatorname{diam}(\mu)$$

since for any pair $x_i$ and $x_j$ in $\operatorname{supp}(\mu)$,

$$\|((1-t)x_i + t\overline{\mu}) - ((1-t)x_j + t\overline{\mu})\| = \|(1-t)x_i - (1-t)x_j\| = (1-t)\|x_i - x_j\|.$$

In particular this holds for any pair determining the diameter of $\mu$. Now identify $\Omega$ with the space $\delta[\Omega]$ of point masses in $\mathscr{P}(\Omega)$. Since we have established that $K\colon \mathscr{P}(\Omega) \to \Omega$ is continuous, there are compositions of continuous maps $K \circ \delta\colon \Omega \to \Omega$ and $\delta \circ K\colon \mathcal{VR}(\Omega; r) \to \mathcal{VR}(\Omega; r)$. Obviously $K \circ \delta = \mathrm{id}_\Omega$ since the mean of a point mass is the only point in its support. Conversely there is a continuous map

$$H(\mu, t)\colon \mathcal{VR}(\Omega; r) \times [0,1] \to \delta[\Omega]$$

given by $H(\mu, t) = \gamma_\mu(t)$, which is a homotopy between $\delta \circ K$ and $\mathrm{id}_{\mathcal{VR}(\Omega;r)}$. In conclusion, $\mathcal{VR}(\Omega; r)$ is homotopy equivalent to $\Omega$, or more precisely, is contractible. ♦

This example is not a surprising result, but the method employed can be directly transferred to each of the settings discussed in Section 4.2. In short, we will be able to show that $\mathcal{VR}(X; r) \simeq X$ under the conditions that

1. any $\mu \in \mathcal{VR}(X;r)$ has a unique center of mass,

2. $K\colon \mathcal{VR}(X;r) \to X$ is continuous, and

3. the geodesic $\gamma_\mu$ connecting $\mu$ to $\delta\left[\overline{\mu}\right]$ does not increase in diameter.

### 4.3.1 Riemannian Manifolds

The first result here stated is the same as [3, Theorem 4.2]. However, the proof technique followed here is more in the spirit of the Wasserstein metric, and has novel corollaries discussed in Section 4.4.

**Theorem 4.3.2.** *Let $M$ be a complete Riemannian manifold. Then for any $0 \le r < \rho(M)$ there is a homotopy equivalence $\mathcal{VR}(M;r) \simeq M$.*

*Proof.* There is as usual an inclusion map $\delta\colon M \to \mathcal{VR}(M;r)$ which is an isometry. By the assumptions on $\rho$ and Theorem 4.2.9, any measure in $\mathcal{VR}(M;r)$ has a unique center of mass, thus $K\colon \mathcal{VR}(M;r) \to M$ is well-defined, and it is continuous by Lemma 4.2.11. Indeed, if $\varepsilon^2 = r' - r$ and $W_2(\mu,\nu) < \varepsilon$, then there exist points $x_i \in \operatorname{supp}(\mu)$ and $y_j \in \operatorname{supp}(\nu)$ with $d^2(x_i, y_j) < \varepsilon^2$. Then $d(x_i, y_j) < \varepsilon$, and since by assumption $d(x_i, x_{i'}) \le r$ for all $i'$ and $d(x_i, y_{j'}) \le d(x_i, y_j) + d(y_j, y_{j'}) < \varepsilon + r = r'$, we have that

$$\operatorname{diam}(\operatorname{supp}(\mu) \cup \operatorname{supp}(\nu)) < r'$$

thereby guaranteeing continuity of $K$.

Let $\gamma_\mu(t)$ be the constant speed geodesic in $\mathscr{P}(M)$ connecting $\mu$ to $K(\mu)$ and recall that

$$\gamma_\mu(t) = \sum_{i=0}^{n} \lambda_i \gamma_{\overline{\mu}}^{x_i}(t).$$

Then by Lemma 4.2.5, the diameter of $\gamma_\mu(t)$ is decreasing, since $\gamma_\mu(t)$ is contained in a ball with sectional curvature bounded above by $\Delta$ and therefore satisfies the CAT$(\kappa)$-inequality for $\kappa = \Delta$.

This means we have a well-defined and continuous homotopy $H(\mu, t)\colon \mathcal{VR}(M;r) \times I \to M$ between $\delta \circ K$ and $\mathrm{id}_{\mathcal{VR}(M;r)}$ given by $H(\mu, t) = \gamma_\mu(t)$, while $K \circ \delta = \mathrm{id}_M$. $\square$

### 4.3.2 Non-Positively Curved Spaces

**Theorem 4.3.3.** *Let $X$ be a local NPC space. Take $\rho(m)$ be the largest radius such that $\mathrm{B}_\rho(m)$ is NPC and geodesically convex. Let $r' = \inf_{m\in M} \rho(m)$. Then for any $0 \le r < r'$, there is a homotopy equivalence $\mathcal{VR}(M;r) \simeq M$.*

*Proof.* Identical to the proof of Theorem 4.3.2, except that the diameter is non-increasing directly by the assumption of non-positive curvature. $\square$

**Remark 4.3.4.** For the sphere, $\mathbb{S}^n$, this version gives $r' = \frac{\pi}{2}$, while Theorem 4.3.2 only permits $r' = \frac{\pi}{4}$. Both are less than the known first parameter at which the homotopy type changes, which is the diameter of an inscribed regular $(n+1)$-dimensional simplex.

Note that it is possible to have $r' = 0$, and that if $M$ is not global NPC, then $r' < +\infty$. In addition to the circle, this version applies to the torus, or any manifold with a flat metric. It also applies to a locally finite metric graph, $G$, where $r'$ is one-quarter the diameter of the smallest loop in $G$.

**Corollary 4.3.5.** *Let $M$ be a complete, simply-connected Riemannian manifold of non-positive sectional curvature. Then $\mathcal{VR}(M;r) \simeq M$ for all $r \in [0, +\infty)$.*

*Proof.* Immediate. $\square$

For global NPC spaces the result of Section 4.3.1 can be strengthened considerably:

**Corollary 4.3.6.** *Let $X$ be a global NPC geodesic space. Then $\mathcal{VR}(X;r) \simeq X$ for all $r \in [0, +\infty]$. In particular, $\mathcal{VR}(X;r)$ is always contractible.*

*Proof.* Apply Proposition 4.2.7 to get that $K$ is well-defined and continuous. Then by convexity of the distance function the usual homotopy is well-defined. $\square$

Vietoris–Rips simplicial complexes (instead of metric thickenings) of trees have been studied in [49, 50], but the analysis is in some part made more complicated due to the difficulties of the simplicial complex topology. Corollary 4.3.6 determines that $\mathcal{VR}(T;r)$ is contractible for any tree $T$.

While we have focused primarily on Vietoris–Rips complexes, the proofs here are easily adapted to the Čech complex as well. Recall that the Čech complex of $X$ contains a simplex whenever all the vertices are contained in a metric ball of radius $r$. We also know that if $\mathrm{supp}(\mu) \subseteq \mathrm{B}_r(m)$, then $K(\mu) \in \mathrm{B}_r(m)$ by Theorem 4.2.9 or Proposition 4.1.4. Thus the flow $\gamma_\mu(t)$ never leaves $\mathrm{B}_r(m)$. Therefore the proofs of Theorem 4.3.2 and Theorem 4.3.3 immediately give:

**Theorem 4.3.7.** *Let $X$ and $r$ satisfy the assumptions of Theorem 4.3.2 or Theorem 4.3.3. Then there is a homotopy equivalence $\check{\mathcal{C}}\left(X;\frac{r}{2}\right) \simeq X$.*

## 4.4 Novel Corollaries

A significant feature of the theorems in Section 4.3 is that they consist of following a geodesic flow in Wasserstein space. A consequence is that the support of the measure $H(\mu,t)$ never increases in cardinality as $t$ varies.

**Definition 4.4.1.** *The $k$**-skeleton** of a simplicial complex $L$ is the subcomplex consisting of simplices of dimension at most $k$. The $k$-skeleton is denoted $L^{(k)}$.*

The Vietoris–Rips complex of course has a $k$-skeleton. Analogously, the $k$-skeleton of the Vietoris–Rips metric thickening is the subspace $\mathcal{VR}^{(k)}(X;r)$ of $\mathcal{VR}(X;r)$ which consists of measures supported on at most $k+1$ points. The $k$-skeleton of the Čech metric thickening is defined analogously.

**Corollary 4.4.2.** *Let $X$ and $r$ satisfy the requirements of Theorem 4.3.2 or Theorem 4.3.3. Then there are homotopy equivalences $\mathcal{VR}^{(k)}(X;r) \simeq X$ and $\check{\mathcal{C}}^{(k)}(X;r) \simeq X$ for any $k \in \mathbb{N}$.*

*Proof.* The homotopy $H(\mu,t)$ along geodesic curves is supported on at most as many points as $\mu$, and therefore restricts to $\mathcal{VR}^{(k)}(X;r)$, and likewise for $\check{\mathcal{C}}^{(k)}(X;r)$. □

**Example 4.4.3.** The Borsuk–Ulam states that given an odd map $f\colon S^n \to \mathbb{R}^n$, there exists a point $x \in S^n$ with $f(x)$ equal to the origin in $\mathbb{R}^n$. In [51], generalizations are given for odd maps into higher-dimensional codomains, namely $f\colon S^n \to \mathbb{R}^k$ with $k \geq n$. In this context, one obtains a set $X \subseteq S^n$ of diameter bounded from below such that the convex hull of $f(X)$ contains the origin in $\mathbb{R}^k$. Carathéodory's theorem implies that the cardinality of $X$ can be taken to be at most $k+1$. Corollary 4.4.2, or related ideas, could be potentially be used to strengthen this bound to give a set $X$ of diameter at most $k$. ♦

Alternatively, it is natural to consider spaces of measures which are not restricted to containing only those of finite support. Define

$$\mathcal{VR}^{(\infty)}(X;r) = \{\mu \in \mathcal{P}(X) \mid \operatorname{diam}(\operatorname{supp}(\mu)) \leq r\}$$

(or < depending on convention). Similarly, define

$$\check{\mathcal{C}}^{(\infty)}(X;r) = \{\mu \in \mathcal{P}(X) \mid \exists\, x \in X \text{ with } \operatorname{supp}(\mu) \subseteq \overline{\mathrm{B}}_r(x)\}.$$

**Corollary 4.4.4.** *Let $X$ and $r$ satisfy the requirements of Theorem 4.3.2, Theorem 4.3.3, or Theorem 4.3.7. Then $\mathcal{VR}^{(\infty)}(X;r) \simeq X$, or, respectively, $\check{\mathcal{C}}^{(\infty)}(X;r) \simeq X$.*

*Proof.* Since Theorem 4.2.9 and Proposition 4.2.7 both apply to arbitrary measures, the proofs of Theorem 4.3.2 and Theorem 4.3.3 apply directly to $\mathcal{VR}^{(\infty)}(X;r)$. □

A motivation for the study of the infinite Vietoris–Rips thickening is the following proposition, which shows that an increasing sequence of samples from a space converges in a precise sense to $\mathcal{VR}^{(\infty)}(X;r)$.

**Proposition 4.4.5.** *Let $M$ be a compact manifold with volume form* dVol. *Let $(X_n)_{i=1}^{\infty}$ be an increasing sequence of subsets of $M$ where $X_n$ is a random sample of $n$ points from* dVol. *Then the sequence of spaces $\mathcal{VR}(X_n;r)$ converges to $\mathcal{VR}^{(\infty)}(X;r)$ almost surely in Gromov–Hausdorff convergence.*

*Proof.* Since $\mathscr{VR}(X_n;r)$ and $\mathscr{VR}^{(\infty)}(M;r)$ are all subsets of $\mathscr{P}(M)$, it suffices to show convergence in Hausdorff distance. For any $n$, $\mathscr{VR}(X_n;r)$ includes into $\mathscr{VR}^{(\infty)}(M;r)$ isometrically. By compactness, there there is an $\varepsilon_n$ such that $X_n$ is an $\varepsilon_n$-net in $M$, and with probability one $\varepsilon_n \to 0$ as $n \to \infty$. By [45, Corollary 7.3.28], $\mathscr{VR}(X_i;r) \to \mathscr{VR}^{(\infty)}(M;r)$ almost surely in Gromov–Hausdorff distance. □

Note that $\mathscr{VR}^{(\infty)}(M;r)$ is compact if and only if $M$ is compact and the $\leq$ convention is used. Thus these assumptions are necessary for the Gromov–Hausdorff distance to be well-defined. Since $\mathscr{VR}(M;r)$ is not closed, it is never compact (barring the trivial case where $M$ is finite).)

# Chapter 5

# Morse Theory on Wasserstein Space

This chapter discusses the differential or Riemannian structure of Wasserstein space, developed originally in [52] and further in [43, 53, 54] among others. Every Alexandrov $X$ space has an infinitesimal structure, and whenever $X$ is Alexandrov, so is $\mathscr{P}(X)$. Therefore it is possible to develop the idea of a gradient flow in Wasserstein space [54]. Using this framework, a version of the first Morse lemma for functions on Wasserstein space is given.

In the case where $M$ is a manifold, a stronger differential structure on $\mathscr{P}(M)$ can be developed. An overview is given in Section 5.3, and the relation between this structure and the technique used to prove Theorem 4.3.2 is discussed.

## 5.1 Infinitesimal Structure of Wasserstein Space

A Riemannian structure on a manifold $M$, intuitively, is the data required to locally describe distances and angles. This admits various equivalent formalizations. One is to define a tangent vector at a point $p \in M$ as an equivalence class of curves (smooth maps $\gamma\colon (-1,1) \to M$) such that $\gamma(0) = p$, and $\gamma$ and $\widetilde{\gamma}$ are identified if under any coordinate chart $\phi$ the derivatives $\frac{\mathrm{d}}{\mathrm{d}t}(\phi \circ \gamma)$ and $\frac{\mathrm{d}}{\mathrm{d}t}(\phi \circ \widetilde{\gamma})$ are equal. The tangent space is the set of all equivalence classes of curves, and it inherits a vector space structure from $\mathbb{R}^d$ via the coordinate charts. Intuitively, the set of curves through a point determine all of the directions of motion away from that point.

A Riemannian metric $g$ on $M$ is a function $g(-,-)\colon T_pM \times T_pM \to \mathbb{R}$ at each tangent space which varies smoothly from point to point on $M$. The angle $\theta$ between two tangent vectors $u$ and $v$ can be defined by $\cos(\theta) = \frac{g(u,v)}{g(u,u)g(v,v)}$. With this data it is possible to define Riemannian geodesics and to show that every equivalence class in the tangent space is represented by a constant-speed geodesic.

When $X$ is a geodesic space, not necessarily a manifold, these definitions can be worked out essentially in reverse to construct an infinitesimal structure. The additional assumption that $X$ is an Alexandrov space will be useful.

**Definition 5.1.1.** *A geodesic metric space $X$ is an* ***Alexandrov space*** *of curvature $\geq \kappa$ if given any three points $x, y, z$ in $X$, geodesics $\gamma_y^x$ and $\gamma_y^z$, and any three comparision points $\hat{x}, \hat{y}, \hat{z}$ and comparison geodesics $\hat{\gamma}_y^x$ and $\hat{\gamma}_y^z$ in $\mathbb{M}^2(\kappa)$, the triangle condition*

$$d(\gamma_y^x(t), \gamma_y^z(s)) \geq d_{\mathbb{M}^2(\kappa)}(\hat{\gamma}_y^x(t), \hat{\gamma}_y^z(s))$$

*holds for all $s, t, \in [0,1]$. If $\kappa > 0$ the equality is only required to hold for triangles in which $d(x,y) + d(y,z) + d(z,x) < \frac{2\pi}{\sqrt{\kappa}}$.*

The definition of an Alexandrov space should be compared to that of CAT$(\kappa)$ spaces (Definition 4.2.4). Heuristically an Alexandrov space is one in which triangles are at least as wide as they are in the comparison space $\mathbb{M}^2(\kappa)$. A Riemannian manifold $M$ is an Alexandrov space of curvature $\geq \kappa$ if the sectional curvature of $M$ is at least $\kappa$ everywhere. Manifolds of constant curvature are both Alexandrov and CAT$(\kappa)$ spaces.

**Theorem 5.1.2.** *$\mathscr{P}(X)$ is an Alexandrov space of curvature $\geq 0$ if and only if $X$ is.*

*Proof.* Since $\delta \colon X \to \mathscr{P}(X)$ is an isometric embedding, the only if is immediate. The converse is shown in [55, Proposition 2.10.(iv)]. □

The same does not hold for the CAT$(\kappa)$ property. Euclidean space is CAT(0), yet $\mathscr{P}(\mathbb{R}^d)$ does not satisfy the CAT$(\kappa)$ condition for $\kappa = 0$, as shown by [38, Example 3.21] which constructs a triangle of three point masses in $\mathscr{P}(\mathbb{R}^2)$ that does not satisfy the CAT(0) condition.

### 5.1.1 Geodesics and Tangent Cones

Any Alexandrov space has an infinitesimal structure given by geodesics and tangent cones. This section focuses on how that structure manifests for $\mathscr{P}(X)$ when $X$ is a compact Alexandrov space with curvature $\kappa \geq 0$. The exposition follow [54] and [38, Chapter 7]. The idea is to define appropriate equivalence classes of curves to serve as tangent vectors. Recall that a curve is a continuous function $\mu(t)\colon I \to \mathscr{P}(X)$, where $I$ is some (usually closed) interval containing 0.

**Definition 5.1.3.** *The **metric derivative** of a curve $\mu(t)$ in $\mathscr{P}(X)$ is defined by*

$$\left|\dot{\mu}\right|(t) := \lim_{h\to 0} \frac{W_2(\mu(t+h), \mu(t))}{h}$$

*whenever this limit exists.*

Since the metric derivative contains no directional information, it is the "speed" rather than the "velocity" of the curve. A sufficient condition to guarantee the existence of the metric derivative is that the curve in question be 2-absolutely continuous.

**Definition 5.1.4.** *A curve $\mu(t)\colon [0,1] \to \mathscr{P}(M)$ is 2**-absolutely continuous** if there exists a function $f \in L^1(0,1)$ such that*

$$d(\mu(t), \mu(s)) \leq \int_t^s f(r)\,\mathrm{d}r$$

*for all $s > t$, $s, t \in (0,1)$.*

Despite the confluence of terminology, 2-absolutely continuous does not in any way imply that the measure $\mu(t)$ is absolutely continuous with respect to some reference measure. In fact, 2-absolute continuity is typically defined for general metric spaces which are not Wasserstein spaces.

**Theorem 5.1.5.** *If $\mu(t)$ is 2-absolutely continuous, then $\left|\dot{\mu}\right|$ exists, and for almost all $t$, $\left|\dot{\mu}\right|(t) \leq f(t)$ as given in Definition 5.1.4.*

*Proof.* [52, Theorem 1.1.2]. □

Observe that a constant speed geodesic is necessarily 2-absolutely continuous since it can always be assumed to be parametrized by arc length. Doing so gives that $W_2(\mu(t), \mu(s)) = |t - s|$ by definition, and so $f(t) = 1$ satisfies Definition 5.1.4.

**Example 5.1.6.** An informative example is the curve $\gamma_t := (1-t)\delta[a] + t\delta[b]$, which is not 2-absolutely continuous. Indeed,

$$\lim_{h\to 0}\frac{W_2(\gamma_{t+h},\gamma_t)}{h} = \lim_{h\to 0}\frac{\sqrt{h}d(a,b)}{|h|} = \lim_{h\to 0}\frac{d(a,b)}{\sqrt{h}} = \infty.$$

This shows that a curve consisting of point masses cannot be 2-absolutely continuous unless the mass distribution remains constant. This curve involves mass "teleporting" from one point to another. ♦

Let $\Sigma'_\mu\mathscr{P}(X)$ denote the set of constant speed geodesics $\gamma\colon [0,\varepsilon]\to\mathscr{P}(X)$ with $\gamma(0)=\mu$. Define an equivalence relation on $\Sigma'_\mu\mathscr{P}(X)$ by $\gamma\sim\tilde{\gamma}$ if $\gamma(t)=\tilde{\gamma}(t)$ for all $t\in[0,\delta]$ for some $\delta>0$. Given two geodesics in $\Sigma'_\mu\mathscr{P}(X)$ and $s,t\geq 0$, define

$$\sigma_\mu((\alpha,s),(\beta,t)) := \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}W_2(\alpha(\varepsilon s),\beta(\varepsilon t)).$$

This limit exists by [54, Theorem 3.6]. The angle, $\angle_\mu$, between two geodesics at the point $\mu$ is determined by

$$\cos(\angle_\mu(\alpha,\beta)) := 1-\frac{\sigma_\mu((\alpha,1),(\beta,1))^2}{2}.$$

The angle is a pseudo-metric on $\Sigma'_\mu\mathscr{P}(X)$.

**Definition 5.1.7.** *The space of directions in $\mathscr{P}(X)$ at $\mu$ is*

$$\Sigma_\mu\mathscr{P}(X) := \overline{\Sigma'_\mu\mathscr{P}(X)/\sim}$$

*where $\alpha\sim\beta$ if $\angle_\mu(\alpha,\beta)=0$, and the closure is taken with respect to the pseudo-metric $\angle_\mu$.*

Form the cone $C_\mu\mathscr{P}(X) := \Sigma_\mu\mathscr{P}(X)\times[0,+\infty)/(\alpha,0)\sim(\beta,0)$. This is called the **tangent cone** at $\mu$. An element $(\alpha,s)$ of the tangent cone should be thought of as describing a direction, given by the geodesic $\alpha$ and a distance to travel along that geodesic, given by the number $s\in[0,+\infty)$. The cone over $\Sigma'_\mu \mathscr{P}(X)$ will be denoted $C'_\mu \mathscr{P}(X)$. Note that the completion of $C'_\mu \mathscr{P}(X)$ is $C_\mu \mathscr{P}(X)$, so the completion and cone can be formed in either order.

### 5.1.2 Differentiable Functions

Now that a tangent space is defined, a method of assigning tangent vectors to curves and gradient vectors to functions will complete the differential structure.

**Definition 5.1.8.** *A curve $\gamma(t)$ is **right differentiable** at $0$ if there is a $v \in C_\mu \mathscr{P}(X)$ such that, for any sequence $\varepsilon_i \to 0$ and any associated sequence of unit speed geodesics $\alpha_i$ from $\mu = \gamma(0)$ to $\gamma(\varepsilon_i)$, the sequence*

$$\lim_{i\to\infty} \left( \alpha_i, \frac{W_2(\mu, \gamma(\varepsilon_i))}{\varepsilon_i} \right) = v$$

This assigns a velocity to a curve $\gamma$ which need not be a geodesic. The vector $v$ is denoted $\gamma'(0)$.

The class of differentiable functions will be those which are lower semi-continuous and $\lambda$-convex. These definitions are stated for a general metric space $X$, but apply immediately to Wasserstein space.

**Definition 5.1.9.** *A function $f\colon X \to \mathbb{R}$ is **lower semi-continuous** at $x \in X$ if*

$$\liminf_{y\to x} f(y) \geq f(x).$$

**Definition 5.1.10.** *Fix $\lambda \in \mathbb{R}$. A function $f\colon X \to \mathbb{R}$ is $\lambda$**-convex** along a curve $\gamma(t)$ if for all $x \in X$ and all $t \in [0,1]$, the inequality*

$$f(\gamma(t)) \leq (1-t) f(\gamma(0)) + t f(\gamma(1)) - \frac{\lambda}{2}(1-t)t d^2(\gamma(0), \gamma(1))$$

*holds.*

Intuitively convexity means that the region above the graph of $f$ is a convex region. The generalization to $\lambda$-convexity for $\lambda < 0$ permits this region to be nearly convex with $\lambda$ measuring the amount of second-order "smoothing" required to make it so. Note that $\lambda$-convexity for $\lambda > 0$ is stronger than simple convexity, while $\lambda < 0$ is weaker.

**Example 5.1.11.** The function $f(x) = x^4$ is a convex function on $\mathbb{R}$. The function $g(x) = x^2(x-1)(x-2)$ is $\lambda$-convex for a negative $\lambda$. Note that the $g(x)$ has two distinct local minima, while a strictly convex function can have at most one. ♦

Let $K_\lambda(\mathscr{P}(M))$ be the space of bounded, lower semi-continuous, real-valued functions on $\mathscr{P}(M)$ which are $\lambda$-convex along all minimal geodesics. [10] Definition 5.1.3 gave a definition of the derivative of a curve analogous to its speed in $\mathscr{P}(M)$. For a function $f \in K_\lambda(\mathscr{P}(M))$, there is a similar derivative which corresponds to the slope of $f$, but without directional information.

**Definition 5.1.12.** *The* ***absolute gradient*** *of $f$ at $\mu$ is*

$$\left|\nabla f\right|(\mu) := \max\left\{0, \limsup_{\nu\to\mu} \frac{f(\mu) - f(\nu)}{W_2(\mu,\nu)}\right\}.$$

Observe immediately from the definition that if $\mu$ is a local minimum of $f$, then $\left|\nabla f\right| = 0$ because the limit superior term will always be negative. The absolute gradient may be infinite, but is always lower semi-continuous.

There is also a "directional derivative." Fix a $\mu \in \mathscr{P}(M)$ at which $\left|\nabla f\right|(\mu) < +\infty$ and choose a unit speed geodesic $\gamma(t)$ with $\gamma(0) = \mu$ and a number $s \geq 0$. Take $v = (\gamma, s)$ in $C'_\mu\mathscr{P}(M)$. Define

$$D'_\mu f(v) := \lim_{\varepsilon\to 0^+} \frac{f(\gamma(\varepsilon s) - f(\mu)}{\varepsilon} \tag{5.1}$$

to be the directional derivative along $v$. Lastly, Extend the derivative to any $v \in C_\mu\mathscr{P}(X)$ by

$$D_\mu f(v) := \liminf_{w\to v} D'_\mu f(w)$$

[10] For slightly more generality, bounded can be replaced by "$f(x) < +\infty$ on a nonempty set of points", and the following can be restricted to the subset $\mathscr{P}(X)^*$ of points for which a given function is finite. This is the approach taken in [54], but the generality will not be needed here.

where $w \in C'_\mu \mathscr{P}(M)$. The directional derivative is actually an element of the tangent space $C_\mu \mathscr{P}(X)$, and [54, Lemma 4.2] gives a way of identifying a gradient vector from it:

**Lemma 5.1.13.** *For any $\mu \in \mathscr{P}(X)$ with $\left|\nabla f\right|(\mu) < \infty$, there is a unique $\gamma \in \Sigma_\mu \mathscr{P}(X)$ such that $D_\mu f(\gamma) = -\left|\nabla f\right|(\mu)$.*

In light of this, define the **negative gradient vector** of $f$ at $\mu$ to be $\nabla f(\mu) := (\gamma, \left|\nabla f\right|(\mu)) \in C_\mu \mathscr{P}(X)$.

**Definition 5.1.14.** *A continuous, locally Lipschitz curve $\xi\colon [0,T) \to \mathscr{P}(X)$ is a **gradient curve** of $f$ if $\xi$ is right differentiable, $\left|\nabla f\right|(\xi(t)) < \infty$ for all $t \in (0,T)$ and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\xi(t) = \left|\nabla f\right|(\xi(t)).$$

*A gradient curve is called **complete** if $T = +\infty$.*

The following results from [54] establish the foundation for a Morse theory of Wasserstein spaces. Up to this point all of the statements in this section do not depend on having curvature $\kappa > 0$, only on having some lower bound, but the following do require a positive curvature assumption. Nevertheless, these results ought to hold in some form without such an assumption, but they are required for a technical reason.

**Theorem 5.1.15.** *Assume that $f \in K_\lambda(\mathscr{P}(X))$. For every $\mu$ in $\mathscr{P}(X)$ there exists a complete gradient curve $\xi\colon [0,\infty) \to \mathscr{P}(X)$ of $f$ with $\xi(0) = \mu$.*

*Proof.* Theorem 5.11 in [54]. □

**Theorem 5.1.16.** *Let $\xi$ and $\zeta$ be gradient curves of $f \in K_\lambda(\mathscr{P}(X))$. Then*

$$W_2(\xi(t), \zeta(t)) \le \exp(-\lambda t) W_2(\xi(0), \zeta(0)).$$

*In particular, there is a unique, continuous gradient flow $G\colon \mathscr{P}(X) \times [0,\infty) \to \mathscr{P}(X)$.*

*Proof.* Theorem 6.2 in [54]. □

These two theorems establish that a continuous gradient flow always exists. The next proposition gives an important control on the function value along a gradient curve.

**Proposition 5.1.17.** *Let $\xi(t)$ be a gradient curve of $f$. Then for any $t > 0$,*

$$f(\xi(t)) = f(\xi(0)) - \int_0^t \left(\left|\nabla f\right|(\xi(s))\right)^2 \,\mathrm{d}s.$$

*Proof.* Proposition 5.12 in [54]. □

## 5.2 Toward a Morse Theory

The machinery of Section 5.1 provides the foundation for a version of the first Morse Lemma for Wasserstein spaces. Recall the classical version of the first Morse lemma:

**Theorem 5.2.1** (First Morse Lemma)**.** *Let $M$ be a smooth manifold and $f\colon M \to \mathbb{R}$ a smooth function. Suppose that $[a,b]$ contains no critical values of $f$, and that $f^{-1}([a,b])$ is compact. Then $f^{-1}((-\infty,b])$ and $f^{-1}((-\infty,a])$ are homeomorphic, and $f^{-1}((-\infty,b])$ deformation retracts to $f^{-1}((-\infty,a])$.*

A proof sketch is as follows: assume without loss of generality that $M$ has a Riemannian metric $g$. Compute the gradient of $f$, and let $X$ be vector field equal to $\nabla f$ on $f^{-1}([a,b])$ and zero outside of a small neighborhood of $f^{-1}([a,b])$. The flow associated to $X$ defines a deformation retraction (in fact, a diffeomorphism) from $f^{-1}((-\infty,b])$ to $f^{-1}((-\infty,a])$. The standard reference is [56].

This proof will be transported to the setting of Section 5.1. The greatest difference is that general vector fields are not yet defined; however, the proof can be constructed "by hand" from the gradient flow without an intermediate vector field at the cost of giving up the homeomorphism.

Throughout this section $X$ is a compact, non-negatively curved Alexandrov space. Let $f$ be a lower semi-continuous, $\lambda$-convex function on $M$. Call a $\mu \in \mathscr{P}(M)$ a critical point of $f$ if either $|\nabla f|(\mu) = 0$ or $f$ is discontinuous at $\mu$. If $M$ is compact, then $\mathscr{P}(M)$ is compact and so is any closed bounded subset of $\mathscr{P}(M)$. In particular, $f^{-1}((-\infty, a])$ is compact for any $a$, and $f^{-1}([a,b])$ is compact for any $a < b$.

**Theorem 5.2.2.** *Suppose that* $[a,b]$ *contains no critical values of* $f$*, so that* $f^{-1}([a,b])$ *contains no critical points. Then* $f^{-1}((-\infty,a]) \simeq f^{-1}((-\infty,b])$.

*Proof.* By Theorem 5.1.16 there is a continuous gradient flow $G$ associated to $f$. Proposition 5.1.17 gives

$$f(G(\mu,t)) = f(\mu) - \int_0^t |\nabla f|(G(\mu,s))\,\mathrm{d}s.$$

The lack of critical points in $f^{-1}([a,b])$ guarantees that $|\nabla f| > 0$ on $f^{-1}([a,b])$, so that $f(G(\mu,t))$ is strictly decreasing on same. Compactness then ensures that there is some $t = \tau(\mu)$ such that $f(G(\mu,\tau(\mu)) = a$, and this varies continuously in $\mu$. Let $T = \max_\mu \tau(\mu)$. Of course, for $\nu \in f^{-1}((-\infty,a])$, $\tau(\nu) = 0$.

Define a deformation retract $R(\mu,t)\colon f^{-1}((-\infty,b]) \times [0,T] \to f^{-1}((-\infty,b])$ by

$$R(\mu,t) = \begin{cases} G(\mu,t) & t \le \tau(\mu) \\ G(\mu,\tau(\mu)) & t \ge \tau(\mu) \end{cases}$$

Since $R(\mu,0) = \mu$ and $f(G(\mu,T)) = a$, the desired homotopy is achieved. □

**Remark 5.2.3.** Since $\mathscr{P}(X)$ is a convex subset of $C_b(X)^*$, a Banach space, $\mathscr{P}(X)$ is contractible for any $X$. The informational content of Morse theory on $\mathscr{P}(X)$ is then necessarily more about the sublevel sets of $f$ at intermediate scales than the topology of $\mathscr{P}(X)$ itself.

Theorem 5.2.2 is analogous to the classical first Morse lemma, which naturally suggests the question:

**Question 1.** Does some version of the second Morse lemma hold for Wasserstein spaces? In particular, if $a$ is a critical value of $f$, is there a useful relation between $f^{-1}((-\infty, a-\varepsilon])$ and $f^{-1}((-\infty, a+\varepsilon])$?

Even in the classical case, the second Morse lemma is considerably more subtle. The first Morse lemma does not require that $f$ be a Morse function, while the second does. At a minimum a second Morse lemma for Wasserstein space should require some non-degeneracy assumption on the critical points of $f$.

Examples of functions in $K_\lambda(\mathscr{P}(M))$ are often closely related to geometric partial differential equations. An extensive discussion can be found in [57].

For the study of Vietoris–Rips complexes, one would ideally be able to apply Theorem 5.2.2 to the function $D\colon \mathscr{P}(M) \to \mathbb{R}$ given by $D(\mu) = \operatorname{diam}(\operatorname{supp}(\mu))$. The infinite Vietoris–Rips thickening described in Section 4.4 can be defined as

$$\mathscr{V}\mathscr{R}^{(\infty)}(M;r) := D^{-1}((-\infty, r]) \subseteq \mathscr{P}(X).$$

Similarly, the Vietoris–Rips thickening is

$$\mathscr{V}\mathscr{R}_{\leq}(M;r) = D^{-1}((-\infty, r]) \subseteq \mathscr{I}(X)$$

where the domain of $D$ is restricted to finitely-supported measures. However, $D$ is not $\lambda$-convex.

**Example 5.2.4.** Let $\mu = \frac{1}{2}\delta[\varepsilon] + \frac{1}{2}\delta[\pi - \varepsilon]$ and $\nu = \frac{1}{2}\delta[\pi + \varepsilon] + \frac{1}{2}\delta[2\pi - \varepsilon]$ (see figure). Clearly, $W_2(\mu, \nu) = 2\varepsilon$, $D(\mu) = D(\nu) = \pi - 2\varepsilon$, and the diameter of their connecting geodesic is

$$D(\gamma(t)) = \begin{cases} (1-2t)(\pi - 2\varepsilon) + 2t\pi & 0 \leq t \leq \frac{1}{2} \\ (2-2t)\pi + (2t-1)(\pi - 2\varepsilon) & \frac{1}{2} \leq t \leq 1 \end{cases}.$$

To be $\lambda$-convex the inequality

$$D(\gamma(t) \leq (1-t)D(\mu) + tD(\nu) - \frac{\lambda}{2}(1-t)tW_2^2(\mu, \nu)$$

must hold for some fixed $\lambda$ and all $t$, $\mu$, and $\nu$. The right hand side is $r(t) = \pi - 2\varepsilon - \frac{\lambda}{2}(1-t)t(2\varepsilon)^2$. Now consider $t = \frac{1}{2}$. There $D(\gamma(\frac{1}{2})) = \pi$, while $r(\frac{1}{2}) = \pi - 2\varepsilon - \frac{\lambda}{2}\varepsilon^2$. For any $\lambda$ there is a choice of $\varepsilon$ such that $r(\frac{1}{2}) < \pi$, and so $D$ is not $\lambda$-convex. ♦

## 5.3 Riemannian Structure

When $M$ is a Riemannian manifold, the differential structure on $\mathscr{P}(M)$ can be strengthened. This section describes this additional structure and then discusses how the homotopies in Section 4.3 are related to gradient flows of this sort. Throughout assume $M$ is a smooth, compact, Riemannian manifold. Let $C_c^\infty(M)$ be the set of compactly-supported smooth functions on $M$, and denote by $\nabla C_c^\infty$ the set of gradients of such functions,

$$\nabla C_c^\infty(M) := \{\nabla f \mid f \in C_c^\infty(M)\}.$$

As a consequence of McCann's theorem (Theorem 3.1.9), at any absolutely continuous $\mu \in \mathscr{P}(M)$, every direction is associated to the gradient of a function. Therefore $C'_\mu\mathscr{P}(M) \cong \nabla C_c^\infty(M)$, and this can be taken as the definition of the tangent space.

More precisely, define $L^2(M, TM; \mu)$ to be the set of Borel vector fields[11] $X$ on $M$ such that $\|X\|_\mu^2 := \int_M \langle X, X\rangle_g^2 \, \mathrm{d}\mu < +\infty$. As usual with $L^p$ spaces, vector fields which differ on sets of measure zero are identified. There is an inner product on $L^2(M, TM; \mu)$ given by

$$(X, Y)_\mu := \int_M \langle X, Y\rangle_g \, \mathrm{d}\mu. \tag{5.2}$$

With this inner product, $L^2(M, TM; \mu)$ is a Hilbert space. There is a natural map of $\nabla C_c^\infty(M)$ into $L^2(M, TM; \mu)$: by definition $\nabla f$ is compactly supported and Borel, and because $f$ is smooth and compactly supported, $\|\nabla f\|$ is bounded, and thus square-integrable with respect to any $\mu$.

[11] That is, sections of the tangent bundle, $TM$, which are Borel-measurable, not-necessarily smooth, maps.

**Definition 5.3.1.** *The **tangent space**, $T_\mu\mathscr{P}(M)$, of $\mu$ in $\mathscr{P}(M)$ is the closure of $\nabla C_c^\infty(M)$ within $L^2(M, TM;\mu)$ under the topology induced by the inner product.*

The tangent space is a Hilbert space since it is a closed linear subspace and $T_\mu\mathscr{P}(M)$ inherits the inner product from $L^2(M, TM;\mu)$ through restriction.

**Example 5.3.2.** Suppose $\mu$ is supported on a finite set of points, $x_1,\dots,x_n$. Then $T_\mu\mathscr{P}(M) = L^2(M, TM;\mu)$, and $T_\mu\mathscr{P}(M) \cong \oplus_{i=1}^n T_{x_i}M$ canonically as vector spaces. The equality holds because vector fields are identified on sets of measure zero, so only the values of a vector field $X$ at $x_1,\dots,x_n$ are relevant. The inner product then reduces to

$$(X,Y)_\mu = \sum_i \lambda_i \langle X(x_i), Y(x_i)\rangle_g$$

where $\lambda_i$ is the weight at $x_i$. Since $\|X\|$ is finite at any given point, the sum is finite for any vector field. It is easy to construct a function $f$ such that $\nabla f(x)$ is equal to a desired vector at each of $x_1,\dots,x_n$, hence the identification with $\oplus_{i=1}^n T_{x_i}M$. Note however that while the vector space isomorphism is canonical, the natural inner product is not the same, in particular, $T_\mu\mathscr{P}(M)$ generally gives a *weighted* sum of the component inner products.

If $\mu = \delta[p]$, then this implies that $T_\mu\mathscr{P}(M) \cong T_pM$. Thus $T_\mu\mathscr{P}(M)$ is clearly a much more restrictive set of directions than $C_\mu\mathscr{P}(M)$ when $\mu$ is discrete. ♦

The tangent space $T_\mu\mathscr{P}(M)$ admits an inclusion into $C_\mu\mathscr{P}(M)$, the tangent cone. Given a $\phi \in C_c^\infty(M)$, a geodesic near $\mu$ can be obtained by $t \mapsto (\mathrm{id} + t\nabla\phi)\#\mu$, which is an element of $\Sigma\mu\mathscr{P}(M)$.[12] This is actually injective and isometric [38, Theorem 7.1]. When $\mu$ is an absolutely continuous measure this is also surjective. Intuitively, this is because every geodesic beginning at an absolutely continuous measure is obtained by a function. For discrete measures this is far from true. The only measures obtained from $\delta[p]$ by a pushforward of a function are other $\delta$ measures.

[12] Technically this gives a representative of an equivalence class.

**Example 5.3.3.** Let $\alpha = \frac{1}{2}\delta[a] + \frac{1}{2}\delta[b]$, and $\beta = \frac{3}{4}\delta[a] + \frac{1}{4}\delta[b]$ for some points $a, b \in M$, and consider the curve $\mu(t) = t\alpha + (1-t)\beta$. Reasoning similar to Section 5.1.1 shows that $\mu(t)$ is not 2-absolutely continuous. There is a geodesic between $\alpha$ and $\beta$,

$$\gamma(t) = \frac{1}{2}\delta[a] + \frac{1}{4}\delta[b] + \frac{1}{4}\gamma_b^a(t).$$

where $\gamma_b^a(t)$ is the geodesic from $a$ to $b$ in $M$. Therefore there are vectors in the tangent cone $C_{\delta[a]}\mathscr{P}(M)$ which describe the direction "within simplices." This data is *not* contained in the tangent space $T_{\delta[a]}\mathscr{P}(M)$ since the mass at $a$ has to "split" instantaneously. ♦

An advantage of this stronger Riemannian structure is that gradients velocities can be stated more explicitly. Let $\mu(t)$ be a curve. The **continuity equation** for $\mu(t)$ is

$$\frac{\partial \mu}{\partial t} + \operatorname{div}_\mu(X_t) = 0 \tag{5.3}$$

and a solution is a vector field $X_t$ on $M$ such that Equation (5.3) holds in the sense of distributions, namely

$$\int_{(-1,1)} \int_M \left( \frac{\partial f}{\partial t} + df(X_t) \right) \mathrm{d}\mu_t \, \mathrm{d}t = 0 \tag{5.4}$$

for all compactly-supported test functions, $f$. A solution is guaranteed whenever $\mu(t)$ is 2-absolutely continuous [52, Theorem 8.3.1], and this solution is interpreted as the velocity of $\mu(t)$.

For a given function $F \colon \mathscr{P}(M) \to \mathbb{R}$ it is possible to construct a vector field by computing the gradient of $F$. The **subdifferential** at $\mu$ of $\partial_\mu F$ is the set of vector fields $\xi$ (on $\mathscr{P}(M)$) defined by

$$\lim_{\nu\to\mu} \frac{F(\nu) - F(\mu) - \sup_{\gamma\in\Gamma_0(\mu,\nu)} \iint_{\mathbb{R}^n\times\mathbb{R}^n} \langle \xi(x), y - x\rangle \, \mathrm{d}\gamma(x,y)}{W_2(\nu,\mu)} \geq 0 \tag{5.5}$$

and the **superdifferential**, $\partial^\mu F$ is the set of vector fields $\xi$ such that

$$\lim_{\nu\to\mu} \frac{F(\nu) - F(\mu) - \sup_{\gamma\in\Gamma_0(\mu,\nu)} \iint_{\mathbb{R}^n\times\mathbb{R}^n} \langle \xi(x), y - x\rangle \, \mathrm{d}\gamma(x,y)}{W_2(\nu,\mu)} \leq 0 \tag{5.6}$$

The functional $F$ is said to be differentiable at $\mu$ if there is a (not necessarily unique) vector field in $\partial_\mu F \cap \partial^\mu F$, and that vector field is the gradient of $F$ at $\mu$, denoted $\nabla_\mu F$. Of course, the limits here need not exist for generic functions $F$, but they do when $F$ is $\lambda$-convex, which is not surprising given the theory of Section 5.1.

There are several fundamental examples where the gradient can be computed explicitly.

**Example 5.3.4.** Let $f\colon M \to \mathbb{R}$, and define the potential energy function $F\colon \mathscr{P}(M) \to \mathbb{R}$ by

$$F(\mu) := \int_M f \, \mathrm{d}\mu.$$

Then the gradient of $F$ is $\nabla f$, so long as $\nabla f \in L^2(M, TM; \mu)$. ♦

**Example 5.3.5.** Let $w\colon M \times M \to R$ be symmetric. The interaction energy function is

$$W(\mu) := \frac{1}{2} \int_M w(x, y) \, \mathrm{d}(\mu \otimes \mu)(x, y).$$

The gradient of $W$ is $\nabla W * \mu$ whenever this is in $L^2(M, TM; \mu)$. ♦

A curve $\mu(t)$ is a gradient flow of $F$ if the velocity vector field $V_t$ associated to $\mu(t)$ is the gradient of $F$ for all $t$, or more generally, if $F$ is $\lambda$-convex and $V_t$ is in the subdifferential of $F(\mu_t)$ for almost all $t$. If $f$ or $W$ satisfies appropriate convexity conditions, then so do $F$ and $W$ and gradients flows exist [38, Proposition 4.33, Theorem 4.35].

### 5.3.1 Homotopies as Gradient Flows

Consider the function $\mathscr{P}(M)$ defined by

$$J(\mu) := \frac{1}{2} \int_{M \times M} d^2(x, y) \, \mathrm{d}(\mu \otimes \mu)(x, y).$$

This has the form of Section 5.3. Ignore for the moment the $\lambda$-convexity requirement (which may or may not be satisfied depending on $M$), and suppose that $\mu$ is supported on a sufficiently small ball in $M$ so that $\exp_x^{-1}(y)$ is well-defined for any $x, y \in \operatorname{supp}(\mu)$. Then Section 5.3 suggests that

$$\nabla J_\mu(x) = \nabla \frac{1}{2} d^2(x, y) * \mu = \int_M \exp_x^{-1}(y)\,\mathrm{d}\mu(y)$$

The gradient flow along $J$, starting at $\mu$, is given by pushing $\mu$ forward along the flow of the vector field $\nabla J_\mu$.

Compare this to gradient of $P_\mu$, which Karcher derives [40] to be

$$\operatorname{grad} P_\mu(x) = -\int_A \exp_x^{-1}(f(a))\,\mathrm{d}\mathbb{P}(a) = -\int_M \exp_x^{-1}(y)\,\mathrm{d}\mu(y), \tag{5.7}$$

assuming that $\operatorname{supp}(\mu) \subseteq \mathrm{B}_r(m)$ for some $m \in M$ and $r < \rho(M)$.

**Example 5.3.6.** When $M = \mathbb{R}^d$, the flow along $\nabla J_\mu$ is exactly the homotopy given in Section 4.3. It is easy to see that $d^2(x, y) = \|x - y\|^2$ is a convex function on $\mathbb{R}^d$, and so by [38, Proposition 4.33] $J$ is convex along interpolating curves, and therefore $\nabla J_\mu$ is the vector field

$$\nabla J_\mu(x) = \int_{\mathbb{R}^d} y - x\,\mathrm{d}\mu(y) = \overline{\mu} - x.$$

A flow along this vector field is a solution to the differential equation

$$\dot{x} = \overline{\mu} - x \quad , \quad x(0) = x_0$$

which has solution given by $x(t) = (1 - \exp(-t)\overline{\mu}) + \exp(-t)x_0$ which is the straight line from $x_0$ to $\overline{\mu}$. Rescaling to have constant speed and pushing $\mu$ forward along the flow gives precisely the map $\gamma_\mu(t)$ used in Section 4.3. ♦

In the manifold setting there is no longer perfect agreement between the geodesic flow and the gradient flow of $J$ because in positive curvature $\nabla P_\mu$ does not point directly to $\overline{\mu}$, but only approximately [40, Theorme 1.5.2]. Heuristically, however, the flows are similar, and it would be possible to replace the geodesic flow in the proof of Theorem 4.3.2 with a gradient flow. Doing so does not improve the bounds on Theorem 4.3.2. Much beyond the small radius dictated by Theorem 4.2.9 the gradient of $J$ cannot be well-defined.

**Proposition 5.3.7.** *Let $m \in M$ and $\mu \in \mathscr{P}(X)$. Then any open ball $\mathrm{B}_r(\mu) \subseteq \mathscr{P}(M)$ contains a measure supported on $m$.*

*Proof.* Take $\mu' = (1-\varepsilon)\mu + \varepsilon\delta[x]$. If $\varepsilon < \frac{r}{W_2(\delta[x],\mu)}$, then $\mu' \in \mathrm{B}_r(\mu)$. □

As a result of Proposition 5.3.7, any sublevel set of $J$ always contains points supported at the cut locus of $\overline{\mu}$, and so $\exp^{-1}$ will not be defined.

# Chapter 6

# Simplicial Metric Thickenings

The Vietoris–Rips metric thickening is the geometric realization of the Vietoris-Rips simplicial complex equipped with an alternative topology. *A priori*, not every simplicial complex admits a topology in the same way, since the vertex set need not be a metric space, nor does every metric space contain the type of simplicial complex structure which the Vietoris–Rips metric thickening does. This chapter introduces a general category of simplicial metric thickenings which generalize the Vietoris–Rips and Čech metric thickenings.

Section 6.1 develops a general categorical construction, the restricted comma category, of which the category of simplicial metric thickenings is a particular example described in Section 6.2 and Section 6.3. Some results about the homotopy type of simplicial metric thickenings under limit and colimit operations is discussed in Section 6.4.2. Finally, a version of Dowker's theorem for simplicial metric thickenings is presented in Section 6.5.

Intuitively one would like the definition of simplicial metric thickening to be something like the following:

**Definition 6.0.1.** *A* ***simplicial metric thickening*** *of a metric space $X$ is a subspace $\mathscr{K}$ of $\mathscr{I}(X)$ which satisfies:*

1. *The image of $\delta\colon X \to \mathscr{P}(X)$ is contained in $\mathscr{K}$, and*

2. *If $\mu \in \mathscr{K}$ and $\nu \ll \mu$, then $\nu \in \mathscr{K}$.*

As a point of comparison, recall the definition of an abstract simplicial complex:

**Definition 6.0.2.** *An* ***abstract simplicial complex*** *on a set $V$ is a subset $\mathscr{K}$ of $2^V$ consisting only of finite sets which satisfies*

1. *The image of the map $v \mapsto \{v\}$ is in $\mathscr{K}$, and*

2. *If $\sigma \in \mathscr{K}$ and $\tau \subseteq \sigma$, then $\tau \in \mathscr{K}$.*

Definition 6.0.1 does in fact describe the objects of the category of simplicial metric thickenings, and could be furnished with appropriate morphisms to produce that category. However, this approach leaves the structure of the category itself a bit opaque.

## 6.1 Comma Categories

For standard definitions in category theory the reader is encouraged to peruse [58]. One abuse of notation that will be used is to write $c \in \mathsf{C}$ when $c$ is an object of the category $\mathsf{C}$.

**Definition 6.1.1.** *Given functors $S\colon \mathsf{A} \to \mathsf{C}$ and $T\colon \mathsf{B} \to \mathsf{C}$, the* ***comma category*** $(S \downarrow T)$ *has as objects all triples $(a, b, \phi)$ where $a \in \mathsf{A}$, $b \in \mathsf{B}$, and $\phi\colon Sa \to Tb$, and as morphisms all pairs $(f_{\mathsf{A}}, f_{\mathsf{B}})$ with $f_{\mathsf{A}} \in \hom \mathsf{A}(a, a')$ and $f_{\mathsf{B}} \in \hom \mathsf{B}(b, b')$, such that the following diagram commutes.*

$$\begin{tikzcd} Sa \arrow[r, "\phi"] \arrow[d, "Sf_1"'] & Tb \arrow[d, "Tf_2"] \\ Sa' \arrow[r, "\phi'"'] & Tb' \end{tikzcd}$$

**Definition 6.1.2.** *The* ***restricted comma category*** $[S \downarrow T]$ *is the subcategory defined to contain all objects $(a, b, \phi)$ of $(S \downarrow T)$ such that $\phi$ is an isomorphism.*

In arbitrary comma categories the order of the source functor $S$ and target functor $T$ is important: $(S \downarrow T)$ and $(T \downarrow S)$ are not equivalent as categories in general. Restricted comma categories are less particular.

**Proposition 6.1.3.** *The categories $[S \downarrow T]$ and $[T \downarrow S]$ are isomorphic.*

*Proof.* Let $\Phi\colon [S \downarrow T] \to [T \downarrow S]$ be defined on objects by $(a, b, \phi) \mapsto (b, a, \phi^{-1})$ and on morphisms by $(f_{\mathsf{A}}, f_{\mathsf{B}}) \mapsto (f_{\mathsf{B}}, f_{\mathsf{A}})$. This is well-defined since the commutativity of the diagram on the left below implies the commutativity of the diagram on the right.

$$\begin{tikzcd} Sa \arrow[r, "\phi"] \arrow[d, "Sf_1"'] & Tb \arrow[d, "Tf_2"] \\ Sa' \arrow[r, "\phi'"'] & Tb' \end{tikzcd} \qquad \begin{tikzcd} Sa \arrow[d, "Sf_1"'] & Tb \arrow[l, "\phi^{-1}"'] \arrow[d, "Tf_2"] \\ Sa' & Tb' \arrow[l, "\phi'^{-1}"] \end{tikzcd}$$

The natural transformation $\Phi$ is an isomorphism because $(\phi^{-1})^{-1} = \phi$. □

The main theorems in Section 6.4 are about various types of limits and colimits. Restricted comma categories inherit these structures from their source and target categories.

Observe that any comma category $(S \downarrow T)$ has two functors $P_\mathsf{A} \colon (S \downarrow T) \to \mathsf{A}$ and $P_\mathsf{B} \colon (S \downarrow T) \to \mathsf{B}$, the **domain** and **codomain** functors. These are given by sending a triple $(a, b, \phi)$ to $a$ and to $b$, respectively, and by sending morphisms $(f_A, f_B)$ to $f_A$ and $f_B$, respectively.

**Lemma 6.1.4.** *Let* $\mathsf{J}$ *be an index category. Suppose that* $\mathsf{A}$ *and* $\mathsf{B}$ *possess colimits under* $\mathsf{J}$*-shaped diagrams and that* $S$ *and* $T$ *preserve colimits under* $\mathsf{J}$*-shaped diagrams. Then* $(S \downarrow T)$ *possesses colimits under* $\mathsf{J}$*-shaped diagrams.*

*Proof.* Let $D \colon \mathsf{J} \to (S \downarrow T)$ be a diagram in the comma category, and denote the objects in its image by $(a_j, b_j, \phi_j)$ for $j \in \mathsf{J}$. Then $P_\mathsf{A} D \colon \mathsf{J} \to \mathsf{A}$ and $P_\mathsf{B} D \colon \mathsf{J} \to \mathsf{B}$ are J-shaped diagrams in $\mathsf{A}$ and $\mathsf{B}$, and so have colimits $\ell_a$ and $\ell_b$. There is a natural transformation $\Phi = (\phi_j)_{j \in J} \colon SP_\mathsf{A} D \Longrightarrow TP_\mathsf{B} D$. Observe that $SP_\mathsf{A} D \colon \mathsf{J} \to \mathsf{C}$ is a diagram in $\mathsf{C}$ with colimit $S\ell_a$ because $S$ preserves limits. Let $Z \colon P_\mathsf{B} D \to \ell_\mathsf{B}$ denote the cocone natural transformation. Then $TZ\Phi \colon SP_\mathsf{A} D \to \ell_B$ is a cocone over $SP_\mathsf{A} D$, so there exists a unique morphism $\psi \colon S\ell_\mathsf{A} \to T\ell_\mathsf{B}$.

The colimit of $D$ is $(\ell_\mathsf{A}, \ell_\mathsf{B}, \psi)$. Indeed, suppose that $(a, b, \chi)$ is a cocone over $D$. Then there are unique morphisms $f_1 \colon \ell_\mathsf{A} \to a$ and $f_2 \colon \ell_\mathsf{B} \to b$ because composition with $P_\mathsf{A}$ or $P_\mathsf{B}$ gives diagrams in $\mathsf{A}$ and $\mathsf{B}$. The morphism $(f_1, f_2) \in \hom(S \downarrow T)((\ell_\mathsf{A}, \ell_\mathsf{B}, \phi), (a, b, \chi))$ is well-defined because everything in sight commutes. □

$$\begin{tikzcd}
\cdots \arrow[r] & (a_i, b_i, \phi_i) \arrow[r] \arrow[drr] & \cdots \arrow[r] & (a_j, b_j, \phi_j) \arrow[r] \arrow[dl] & \cdots \\
& & (\ell_{\mathsf{A}}, \ell_{\mathsf{B}}, \psi) & &
\end{tikzcd}$$

**Figure 6.1:** The colimit under an arbitrary diagram $J$ is $(\ell_{\mathsf{A}}, \ell_{\mathsf{B}}, \psi)$.

**Lemma 6.1.5.** *Let* $\mathsf{J}$ *be an index category. Suppose that* $\mathsf{A}$ *and* $\mathsf{B}$ *possess limits over* $\mathsf{J}$*-shaped diagrams and that* $S$ *and* $T$ *preserve limits over* $\mathsf{J}$*-shaped diagrams. Then* $(S \downarrow T)$ *possesses limits over* $\mathsf{J}$*-shaped diagrams.*

*Proof.* Dualize the proof of Lemma 6.1.4. □

**Lemma 6.1.6.** *Suppose that* $\mathsf{A}$ *and* $\mathsf{B}$ *have all (co)limits of all* $\mathsf{J}$*-shaped diagrams, and that* $S$ *and* $T$ *preserve (co)limits over* $\mathsf{J}$*-shaped diagrams. Let* $D\colon \mathsf{J} \to [S \downarrow T]$ *be a diagram in* $[S \downarrow T]$ *viewed as a subcategory of* $(S \downarrow T)$*. Then the (co)limit of* $D$ *is an object of* $[S \downarrow T]$*.*

*Proof.* Using the notation in the proof of Lemma 6.1.4, there are natural transformations $\Phi$ and $\Phi^{-1}$. This means that $S\ell_{\mathsf{A}}$ is a colimit under $TP_{\mathsf{B}}D$ and $T\ell_{\mathsf{B}}$ is a colimit under $SP_{\mathsf{A}}D$. Hence there are unique morphisms $\psi\colon S\ell_{\mathsf{A}} \to T\ell_{\mathsf{B}}$ and $\xi\colon T\ell_{\mathsf{B}} \to S\ell_{\mathsf{A}}$ and these are necessarily isomorphisms. □

**Lemma 6.1.7.** *Let* $P_{\mathsf{A}}$ *and* $P_{\mathsf{B}}$ *be the domain and codomain functors from* $(S \downarrow T)$ *to* $\mathsf{A}$ *and* $\mathsf{B}$*, repsectively. If* $T$ *has a left adjoint, then so does* $P_{\mathsf{A}}$*, and if* $S$ *has a right adjoint, then so does* $P_{\mathsf{B}}$*.*

*Proof.* Assume that $T$ has a left adjoint $L$, with counit $\varepsilon\colon LT \to \mathrm{id}_{\mathsf{B}}$ and unit $\eta\colon \mathrm{id}_{\mathsf{C}} \to TL$. Define $\widetilde{L}\colon \mathsf{A} \to (S \downarrow T)$ by

$$\mathsf{A} \ni a \mapsto (a, LSa, \eta_{Sa}) \in (S \downarrow T)$$

$$\hom \mathsf{A}(a_1, a_2) \ni f \mapsto (f, LSf) \in \hom (S \downarrow T)(\widetilde{L}a_1, \widetilde{L}a_2)$$

$$\begin{tikzcd}
 & {[S \downarrow T]} \arrow[dl, bend right, "P_{\mathsf{A}}"'] \arrow[dr, bend left, "P_{\mathsf{B}}"] & \\
\mathsf{A} \arrow[ur, dashed, bend left, "\widetilde{L}"] \arrow[dr, bend right, "S"'] & & \mathsf{B} \arrow[dl, bend left, "T"] \\
 & \mathsf{C} \arrow[ur, bend left, "L"] &
\end{tikzcd}$$

**Figure 6.2:** The setup of Lemma 6.1.7.

Observe that $P_{\mathsf{A}}\widetilde{L} = \mathrm{id}_{\mathsf{A}}$ so there is trivially a unit $\widetilde{\eta}\colon \mathrm{id}_{\mathsf{A}} \to P_{\mathsf{A}}\widetilde{L}$. Construct a counit $\widetilde{\varepsilon}\colon \widetilde{L}P_{\mathsf{A}} \to \mathrm{id}_{(S\downarrow T)}$ by defining $\widetilde{\varepsilon}_{(a,b,\phi)} = (\mathrm{id}_a, \varepsilon_b \circ L\phi)$.

Dualizing this shows that if $S$ has a right adjoint $R$ then $P_{\mathsf{B}}$ thas a right adjoint $\widetilde{R}$. $\square$

**Corollary 6.1.8.** *Let $P_{\mathsf{A}}$ and $P_{\mathsf{B}}$ be the domain and codomain functors from $[S\downarrow T]$ to $\mathsf{A}$ and $\mathsf{B}$, repsectively. If $S$ has a left or right adjoint, then so does $P_{\mathsf{B}}$, and likewise if $T$ has a left or right adjoint, so does $P_{\mathsf{A}}$.*

*Proof.* Apply Lemma 6.1.7 and Proposition 6.1.3. $\square$

## 6.2 Category of Simplicial Metric Thickenings

To formalize simplicial thickenings as comma categories, recall the definitions of the categories of simplicial complexes and of metric spaces.

Simplicial complexes have been defined in Definition 6.0.2. The appropriate morphisms are simplicial maps:

**Definition 6.2.1.** *Let $K$ and $L$ be a simplicial complexes with vertex sets $K^0$ and $L^0$. A **simplicial map** is a function $f\colon K^0 \to L^0$ such that if $\sigma$ is a simplex of $K$, then $f(\sigma)$ is a simplex of $L$.*

The category of simplicial complexes, $\mathsf{sCpx}$, has abstract simplicial complexes as objects and simplicial maps as morphisms. This category possesses finite products and coproducts.

The categorical product of simplicial complexes $K$ and $L$ is the simplicial complex $K\prod L$, where $(\sigma,\tau)\in K\prod L$ is a simplex whenever $\sigma\in K$ and $\tau\in L$ [59, Definition 4.25]. The coproduct, $K\coprod L$, is the disjoint union simplicial complex.

**Definition 6.2.2.** *Let $X$ and $Y$ be metric spaces and $c\in[0,+\infty)$. A function $f\colon X\to Y$ is $c$-Lipschitz if $d(f(x),f(x'))\leq kd(x,x')$ for all $x,x'\in X$. Functions which are $1$-Lipschitz may be called **short**.*

Lipschitz functions are, of course, continuous. The category of metric spaces, Met, has metric spaces as objects and short maps as morphisms. While this is a standard definition (it is the same used in [60], for example), there are alternative definitions in the literature, where either the morphisms are less-restricted, or the axioms of a metric space are relaxed. In particular, the morphisms may be allowed to be all maps which are $k$-Lipschitz for some $k\in[0,\infty)$, or simply continuous maps. The latter is the structure of the category of metric spaces as a full subcategory of Top. Many of the constructions here do not depend on the choice of morphisms for Met and so Met$^*$ is used to be agnostic about this choice.

The metric space axioms may also be relaxed when defining a category of metric spaces. Recall that the classical definition of a metric space is a set $X$ equipped with a function

$$d(\cdot,\cdot)\colon X\times X\to[0,+\infty)$$

such that

- $d(x,x)=0$, and $d(x,y)=0$ if and only if $x=y$,
- $d(x,y)=d(y,x)$ for all $x,y\in X$, and
- $d(x,z)\leq d(x,y)+d(y,z)$ for all $x,y,z\in X$.

Allowing $d(x,y)=\infty$ gives an **extended metric space**, while allowing $d(x,y)=0$ when $x\neq y$ gives a **pseudo-metric space**, and allowing $d(x,y)\neq d(y,x)$ is a **quasi-metric space**. Combining all of the above relaxations gives Lawvere metric spaces, or categories enriched in the monoidal poset $([0,+\infty],\leq,+)$ [61]. Here classical metric spaces and extended pseudo-metric spaces are used, with the latter denoted by pMet. Of course, Met is a full subcategory of pMet.

The category Met has finite products. If $X$ and $Y$ are metric spaces, the product $X\times Y$ is the cartesian product of the underlying sets with the supremum norm:

$$d((x,y),(x',y'))=\max\{d(x,x'),d(y,y')\}.$$

Arbitrary products do not exist. For example $\mathbb{R}^{\mathbb{R}}$ is a product in the category of topological spaces, but it is not metrizable (it is not first-countable).

Coproducts do not exist in Met; however, in Section 6.4.2 we define a pointed category of metric spaces where the coproduct structure, the wedge sum, is of more interest. One advantage of pMet is the existence of (unpointed) coproducts as well as products. The coproduct $X\coprod Y$ is the set $X\sqcup Y$ with $d(x,y)=+\infty$ for $x\in X$ and $y\in Y$ (all other distances are unchanged). The product is the same as in Met.

Both the categories of metric spaces and simplicial complexes possess canonical functors to Set. For metric spaces, the functor $U$ is given by forgetting the metric $d$,

$$U\colon \mathsf{Met}\ni(X,d)\mapsto X\in\mathsf{Set}$$
$$f\colon (X,d_X)\to(Y,d_Y)\mapsto f\colon X\to Y$$

For abstract simplicial complexes, the functor $\square^0$ is given by forgetting the subset structure,

$$\square^0\colon \mathsf{sCpx}\ni K\mapsto K^0\in\mathsf{Set}$$
$$f\colon K\to L\mapsto f|_{K^0}\colon K^0\to L^0$$

When no ambiguity is possible the symbols $U$ and $\square^0$ will not appear explicitly, with only $X$ or $K^0$ referring to the underlying sets.

**Definition 6.2.3.** *The category* $\mathsf{MetTh}$ *of simplicial metric thickenings is the restricted comma category* $[U \downarrow \square^0]$*. Explicitly, objects are triples* $(X, K, \phi)$*, in which* $X$ *is a metric space,* $K$ *is an abstract simplicial complex, and* $\phi\colon K^0 \to X$ *is an isomorphism of sets, and a morphism between* $(X, K, \phi)$ *and* $(Y, L, \psi)$ *is a pair* $(f\colon X \to Y, g\colon K \to L)$ *such that the following commutes in* $\mathsf{Set}$*:*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\phi"'] & Y \arrow[d, "\psi"] \\ K^0 \arrow[r, "g"'] & L^0 \end{tikzcd}$$

The source category of $U$ can be either $\mathsf{Met}$ or $\mathsf{pMet}$ and this choice is distinguished by $\mathsf{MetTh}$ and $\mathsf{pMetTh}$.

**Proposition 6.2.4.** *Both the domain functor* $P_{\mathsf{pMet}}\colon \mathsf{pMetTh} \to \mathsf{pMet}$ *and the codomain functor* $P_{\mathsf{sCpx}}\colon \mathsf{pMetTh} \to \mathsf{sCpx}$ *have left and right adjoints. In addition, the functor* $P_{\mathsf{Met}}$ *also defines a functor* $\mathsf{MetTh} \to \mathsf{Met}$ *with left and right adjoints.*

*Proof.* As per Corollary 6.1.8 it suffices to show that $U$ and $\square^0$ have adjoints. Starting with $\square^0$, the right adjoint is the complete simplicial complex functor, $C$, and the left adjoint is the trivial complex functor, $T$.

Let $D_r\colon \mathsf{Set} \to \mathsf{pMet}$ be the functor giving every set the discrete metric where all distances are equal to $r$. The right adjoint of $U$ is $D_0$ and the left adjoint is $D_\infty$. These are not defined for $\mathsf{Met}$, and so $P_{\mathsf{sCpx}}$ has adjoints only in $\mathsf{pMetTh}$, and not in $\mathsf{MetTh}$. □

**Proposition 6.2.5.** *If* $\mathsf{pMet}$ *and* $\mathsf{sCpx}$ *each possess limits over diagrams of shape* $\mathsf{J}$*, then so does* $\mathsf{pMetTh}$*. If* $\mathsf{Met}$ *and* $\mathsf{sCpx}$ *each posses limits over diagrams of shape* $\mathsf{J}$*, then so does* $\mathsf{MetTh}$*. In particular,* $\mathsf{pMetTh}$ *possesses finite products and coproducts, and* $\mathsf{MetTh}$ *possesses finite products.*

*Proof.* As described in Proposition 6.2.4, $\square^0$ and $U$ both have left and right adjoints. Therefore, both are continuous and cocontinuous functors, i.e., they preserve small limits and colimits. By Lemma 6.1.6, $[U \downarrow \square^0]$ has limits of any small diagram for which limits exist in both $\mathsf{Met}$ and $\mathsf{sCpx}$. □

## 6.3 Metric Realization Functor

This section shows that every object of $\mathsf{pMetTh}$ can be realized as a space satisfying Definition 6.0.1. This metric space is called the **metric realization** of the simplicial metric thickening. It was first introduced in [3] and is related to [60].

**Definition 6.3.1.** *The* ***metric realization functor*** $\square^{\mathrm{m}}\colon \mathsf{pMetTh} \to \mathsf{pMet}$ *is specified by the following data:*

- *For each simplicial thickening $\mathcal{K} = (X, K, \phi)$ in $\mathsf{MetTh}$, let $\mathcal{K}^{\mathrm{m}}$ be the sub-metric space of $\mathcal{P}(X)$ of all probability measures $\mu$ such that $\operatorname{supp}(\mu) = \sigma$ for some $\sigma \in K$.*
- *For each morphism $(f, g)\colon (X, K, \phi) \to (Y, L, \psi)$, let $(f, g)^{\mathrm{m}}$ be the morphism taking $\mu = \sum_{i=1}^{n} \lambda_i \delta[x_i]$ to $f^{\mathrm{m}}(\mu) = \sum_{i=1}^{n} \lambda_i \delta\left[f(x_i)\right]$.*

This also restricts to a functor $\square^{\mathrm{m}}\colon \mathsf{MetTh} \to \mathsf{Met}$.

There is no difficulty in allowing pseudo-metric spaces here, even though Chapter 3 only treats classical metric spaces. If $X$ contains some point $x'$ with $d(x', x) = \infty$ for some $x$ (and hence all $y$ within finite distance of that $x$), then no measure with $x, x' \in \operatorname{supp}(\mu)$ is in $\mathcal{P}(X)$ due to the finite moments condition. Pseudo-metric spaces also have a natural topology and a well-defined Borel $\sigma$-algebra, so $\mathcal{P}(X)$ is defined for such spaces.

The objects here are precisely those described by Definition 6.0.1. For finitely-supported measures, $\nu \ll \mu$ if and only if $\operatorname{supp}(\nu) \subseteq \operatorname{supp}(\mu)$. Therefore the morphisms are precisely functions $f\colon X \to Y$ between metric spaces such that the induced map $f\#\colon \mathcal{K}^{\mathrm{m}} \to \mathcal{P}(Y)$ has its image contained in $\mathcal{L}^{\mathrm{m}}$. This holds for any of the variants $\mathsf{Met}^*$ of categories of metric spaces, though in the only $\mathsf{Met}$ and $\mathsf{pMet}$ with short maps are used.

Unsurprisingly, the Vietoris–Rips complex provides a natural example of the construction of simplicial thickenings. The definition of the metric realization gives a new description of the Vietoris–Rips complex and metric thickening as functors.

**Definition 6.3.2.** *Let* $r \in [0, +\infty]$. *The* ***Vietoris–Rips functor*** $\mathrm{VR}(\square; r) : \mathsf{Met} \to \mathsf{MetTh}$ *is defined by*

$$\mathrm{VR}(\square; r) : \mathsf{Met} \ni X \mapsto (X, \mathrm{VR}(X; r), \mathrm{id})$$
$$f : X \to Y \mapsto (f, f)$$

This is well-defined because $f$ is a short map and therefore sends any simplex $\sigma$ to a set of points with no larger diameter. The **Vietoris–Rips simplicial thickening** is the composition of functors $\mathrm{VR}(\square; r)^{\mathrm{m}}$. For consistency with previous chapters this will be stylized $\mathscr{V}\mathscr{R}(\square; r)$.

**Definition 6.3.3.** *The* ***Čech complex functor*** $\check{\mathrm{C}}(\square; r) : \mathsf{Met} \to \mathsf{MetTh}$ *is defined by*

$$\check{\mathrm{C}}(\square; r) : \mathsf{Met} \ni X \mapsto (X, \check{\mathrm{C}}(X; r), \mathrm{id})$$
$$f : X \to Y \mapsto (f, f)$$

Again, the **Čech simplicial thickening** is the composition $\check{\mathrm{C}}(\square; r)^{\mathrm{m}} = \check{\mathscr{C}}(\square; r)$.

## 6.4 Homotopy Types and (Co)Limit Operations

Vietoris–Rips and Čech simplicial complexes preserve certain homotopy properties under products and wedge sums. The case of ($L^\infty$) products is given in [18, Proposition 10.2], and the case of wedge sums is given in [21, Proposition 4] and [62].

This section gives categorical proofs of similar results for metric thickenings. In particular, if $\mathsf{Met}$ and $\mathsf{sCpx}$ have (co)limits of a certain shape, then so does $\mathsf{MetTh}$, by Proposition 6.2.5. These (co)limits are not generally preserved by the metric thickening functors $\square^{\mathrm{m}}$, $\mathscr{V}\mathscr{R}(\square; r)$, and $\check{\mathscr{C}}(\square; r)$, but certain ones are preserved up to homotopy type.

### 6.4.1 Metric Thickenings of Products

The simples limit operation to consider is the product. Let $\times$ denote the product in $\mathsf{Met}$ and $\mathsf{sCpx}$, and $\prod$ the product in $\mathsf{MetTh}$. Since products exist in both $\mathsf{Met}$ and $\mathsf{sCpx}$, they exist in $\mathsf{MetTh}$ by Proposition 6.2.5. Explicitly, the product of $\mathscr{M} = (X, K, \phi)$ and $\mathscr{N} = (Y, L, \psi)$ is $\mathscr{M} \prod \mathscr{N} = (X \times Y, K \times L, \phi \times \psi)$.

**Proposition 6.4.1.** *For any simplicial metric thickenings $\mathscr{M}$ and $\mathscr{N}$, the metric realization factors over the product up to homotopy:*

$$\mathscr{M}^{\mathrm{m}} \times \mathscr{N}^{\mathrm{m}} \simeq (\mathscr{M} \prod \mathscr{N})^{\mathrm{m}}.$$

*Proof.* Let $\mathscr{M} = (X, K, \phi)$ and $\mathscr{N} = (Y, L, \psi)$. Elements of $\mathscr{M}^{\mathrm{m}}$ are finitely-supported measures of the form $\mu = \sum_i \lambda_i \delta[x_i]$ with $x_i \in X$ and $\mathrm{supp}(\mu) \in K$. Likewise elements of $\mathscr{N}^{\mathrm{m}}$ have the form $\nu = \sum_j \xi_j \delta[y_j]$ with $y_j \in Y$ and $\mathrm{supp}(\nu) \in L$. Thus elements of $\mathscr{M}^{\mathrm{m}} \times \mathscr{N}^{\mathrm{m}}$ are pairs $(\mu, \nu) = (\sum_i \lambda_i \delta[x_i], \sum_j \xi_j \delta[y_j])$ with $\mathrm{supp}(\mu) \times \mathrm{supp}(\nu) \in K \times L$, i.e. $\mathrm{supp}(\mu) \in K$ and $\mathrm{supp}(\nu) \in L$. On the other hand, elements of $(\mathscr{M} \prod \mathscr{N})^{\mathrm{m}}$ are measures on $X \times Y$ of the form $\sum_k \zeta_k \delta[(x, y)_k]$.

With this in mind, there is is an obvious injection $\iota\colon \mathscr{M}^{\mathrm{m}} \times \mathscr{N}^{\mathrm{m}} \hookrightarrow (\mathscr{M} \prod \mathscr{N})^{\mathrm{m}}$ via

$$\left(\sum_i \lambda_i \delta[x_i], \sum_j \xi_j \delta[y_j]\right) \mapsto \sum_{i,j} \lambda_i \xi_j \delta[(x_i, y_j)].$$

Concretely, $\iota$ sends a pair of measures on $X$ and $Y$ to their product measure on $X \times Y$.

There is also a surjection $\rho\colon (\mathscr{M} \prod \mathscr{N})^{\mathrm{m}} \twoheadrightarrow \mathscr{M}^{\mathrm{m}} \times \mathscr{N}^{\mathrm{m}}$ given by taking the marginals of the joint distribution:

$$\sum_{i,j} \zeta_{i,j} \delta[(x_i, y_j)] \mapsto \left(\sum_i \left(\sum_j \zeta_{i,j}\right) \delta[x_i], \sum_j \left(\sum_i \zeta_{i,j}\right) \delta[y_j]\right).$$

Finally, $\iota$ and $\rho$ are homotopy inverses. By construction $\rho \circ \iota = \mathrm{id}$. The composition $\iota \circ \rho$ gives the map

$$\sum_{i,j} \zeta_{i,j} \delta[(x_i, y_j)] \mapsto \sum_{i,j} \left(\sum_i \zeta_{i,j}\right) \left(\sum_j \zeta_{i,j}\right) \delta[(x_i, y_j)].$$

This is homotopic to the identity on $(\mathscr{M} \prod \mathscr{N})^{\mathrm{m}}$ via the straight-line homotopy $H\colon (\mathscr{M} \prod \mathscr{N})^{\mathrm{m}} \times I \to (\mathscr{M} \prod \mathscr{N})$ where

$$H(t, \mu) = t\,\mathrm{id}(\mu) + (1 - t)\iota \circ \rho(\mu).$$

This is clearly well-defined as a map to $\mathscr{P}((X \times Y))$. To see that the image of $H$ is in $(\mathscr{M} \prod \mathscr{N})^{\mathrm{m}}$, note that $\mathrm{supp}(\iota \circ \rho(\mu)) \subseteq \mathrm{supp}(\mu)$, so the entire homotopy takes place within a simplex of $K \times L$. It then follows from [3, Lemma 3.9] that the homotopy $H$ is continuous. □

**Proposition 6.4.2.** *As functors* $\mathsf{Met} \to \mathsf{MetTh}$, *both* $\mathrm{VR}(\square; r)$ *and* $\check{\mathrm{C}}(\square; r)$ *preserve products.*

*Proof.* As simplicial complexes $\mathrm{VR}(X \times Y; r) \cong \mathrm{VR}(X; r) \prod \mathrm{VR}(Y; r)$ since, with the $L^\infty$ metric, a subset of $X \times Y$ has diameter equal to the maximum of the diameters of its coordinate projections. Similarly, $\check{\mathrm{C}}(X \times Y; r) \cong \check{\mathrm{C}}(X; r) \prod \check{\mathrm{C}}(Y; r)$ since a collection of $L^\infty$ balls intersect if and only if their projections onto both factors intersect. □

**Corollary 6.4.3.** *For any metric spaces $X$ and $Y$, the product operation factors through the metric Vietoris–Rips and Čech metric thickenings up to homotopy:*

$$\mathscr{VR}(X \times Y; r) \simeq \mathscr{VR}(X; r) \times \mathscr{VR}(Y; r)$$

$$\check{\mathrm{C}}(X \times Y; r) \simeq \check{\mathrm{C}}(X; r) \times \check{\mathrm{C}}(Y; r).$$

*Proof.* Apply Proposition 6.4.1 and Proposition 6.4.2. □

**Remark 6.4.4.** The same result holds true for coproducts; however, the coproduct is not interesting. Recall the coproduct $X \coprod Y$ of metric spaces $X$ and $Y$ has $d(x, y) = +\infty$ for $x \in X$ and $y \in Y$, and so is only defined in $\mathsf{pMet}$. Therefore $\mathrm{VR}(X \coprod Y; r) \cong \mathrm{VR}(X; r) \coprod \mathrm{VR}(Y; r)$ and $\check{\mathrm{C}}(X \coprod Y; r) \cong \check{\mathrm{C}}(X; r) \coprod \mathrm{VR}(Y; r)$ trivially. Similarly, $\mathscr{M}^{\mathrm{m}} \coprod \mathscr{N}^{\mathrm{m}} \cong (\mathscr{M} \coprod \mathscr{N})^{\mathrm{m}}$ for trivial reasons.

### 6.4.2 Metric Thickenings of Gluings

The wedge sum is a natural colimit operation to consider. Recall that a terminal object in a category $\mathsf{C}$ is the (unique up to isomorphism) object $\bullet \in \mathsf{C}$ such that there is a unique morphism $\bullet_A \colon \bullet \to A$ for every $A \in \mathsf{C}$.

**Definition 6.4.5.** *Let $\bullet$ be the terminal object in a category $\mathsf{C}$. Let $A, B \in \mathsf{C}$. The **wedge sum** of $A$ and $B$ is the pushout of $\bullet_A$ and $\bullet_B$:*

$$\begin{tikzcd} \bullet \arrow[r, "\bullet_B"] \arrow[d, "\bullet_A"'] & B \arrow[d, "\iota_B"] \\ A \arrow[r, "\iota_A"] & A \vee B \end{tikzcd}$$

**Proposition 6.4.6.** *Wedge sums exist in* $\mathsf{Met}$*,* $\mathsf{sCpx}$*, and* $\mathsf{MetTh}$*.*

*Proof.* The description of the wedge sum in each category is essentially the same. The terminal object in $\mathsf{Met}$ is the metric space with a single point. The wedge sum $X \vee Y$ is the metric space

$$X \sqcup Y / (\bullet_X \sim \bullet_Y),$$

that is, $X$ and $Y$ are "glued together" at the points $\bullet_X$ and $\bullet_Y$. This common basepoint in $X \vee Y$ will be represented by $\bullet$. The metric on $X \vee Y$ is given by $d(x, y) = d(x, \bullet) + d(\bullet, y)$ for $x \in X$ and $y \in Y$, while distances within $X$ and $Y$ are unchanged. One can check that with this metric $X \vee Y$ satisfies the appropriate universal property.

The terminal object in $\mathsf{sCpx}$ is the simplicial complex with a single vertex. The wedge sum $K \vee L$ is the simplicial complex

$$K \sqcup L / (\bullet_K \sim \bullet_L).$$

Since wedge sums exist in both $\mathsf{Met}$ and $\mathsf{sCpx}$, they exist in $\mathsf{MetTh}$ by Proposition 6.2.5. Explicitly, the wedge sum of $\mathscr{M} = (X, K, \phi)$ and $\mathscr{N} = (Y, L, \psi)$ is $\mathscr{M} \vee \mathscr{N} = (X \vee Y, K \vee L, \phi \vee \psi)$. □

**Remark 6.4.7.** For any simplicial metric thickenings $\mathscr{M}$ and $\mathscr{N}$, the metric realization factors over the wedge sum. Indeed, we have $\mathscr{K}^{\mathrm{m}} \vee \mathscr{L}^{\mathrm{m}} = (\mathscr{K} \vee \mathscr{L})^{\mathrm{m}}$. However, if $F\colon \mathsf{Met} \to \mathsf{MetTh}$, it is too much to expect that $F(\mathscr{M} \vee \mathscr{N}) \cong F(\mathscr{M}) \vee F(\mathscr{N})$ (this fails for the Vietoris–Rips functor, for example). Therefore proving that the metric thickening behaves well with respect to wedge sums is more delicate than the product case.

**Proposition 6.4.8.** *Let $\mathcal{M} = (X,K,\phi)$ and $\mathcal{N} = (Y,L,\psi)$ be simplicial thickenings. Suppose the simplicial thickening $\mathcal{V} = (X \vee Y, S, \phi)$ has the property that $S \supseteq K \vee L$, and if $\sigma \in S$ is a subset of neither $X$ nor $Y$, then $\sigma \cup \bullet$ is also a simplex in $S$. Then $\mathcal{V}^{\mathrm{m}} \simeq (\mathcal{M} \vee \mathcal{N})^{\mathrm{m}}$.*

*Proof.* Elements of both $\mathcal{V}^{\mathrm{m}}$ and $(\mathcal{M} \vee \mathcal{N})^{\mathrm{m}}$ have the form

$$\sum_i \lambda_i \delta[x_i] + \sum_j \xi_j \delta[y_j] + \omega\delta[\bullet].$$

where $x_i \in X$ and $y_j \in Y$. Define $\lambda = \sum_i \lambda_i$ and $\xi = \sum_j \xi_j$, so $\omega + \lambda + \xi = 1$. Elements of $(\mathcal{M} \vee \mathcal{N})^{\mathrm{m}}$ must satisfy $\lambda = 0$ or $\xi = 0$. Since $S \supseteq K \vee L$, there is an inclusion $\iota\colon (\mathcal{M} \vee \mathcal{N})^{\mathrm{m}} \hookrightarrow \mathcal{V}^{\mathrm{m}}$.

Define $\rho\colon \mathcal{V}^{\mathrm{m}} \twoheadrightarrow (\mathcal{M} \vee \mathcal{N})^{\mathrm{m}}$ by

$$\omega\delta[\bullet] + \sum_i \lambda_i \delta[x_i] + \sum_j \xi_j \delta[y_j] \mapsto \begin{cases} (2\xi + \omega)\delta[\bullet] + \left(1 - \frac{\xi}{\lambda}\right)\sum_i \lambda_i \delta[x_i] & \text{if } \lambda \geq \xi \\ (2\lambda + \omega)\delta[\bullet] + \left(1 - \frac{\lambda}{\xi}\right)\sum_j \xi_j \delta[y_j] & \text{if } \xi \geq \lambda, \end{cases}$$

setting $\frac{\xi}{\lambda} = 1$ if $\lambda = 0$ and $\frac{\lambda}{\xi} = 1$ in the case that $\xi = 0$. To see that $\rho$ is continuous, note that the two piecewise formulas agree when $\lambda = \xi$ (in which case the image of $\rho$ is $\bullet$). By construction the image of $\rho$ is in $(\mathcal{M} \vee \mathcal{V})^{\mathrm{m}}$, and $\rho$ is in fact a deformation retract, so $\rho \circ \iota = \mathrm{id}$.

To complete the proof, $\iota \circ \rho$ is homotopic to the identity via

$$H(t,\mu) = t\,\mathrm{id}(\mu) + (1-t)\iota \circ \rho(\mu).$$

Two cases are required to show that the image of $H$ is indeed $\mathcal{V}^{\mathrm{m}}$. If $\mathrm{supp}(\mu) \subseteq X$ or $\mathrm{supp}(\mu) \subseteq Y$, then $\mathrm{supp}(\iota \circ \rho(\mu)) = \mathrm{supp}(\mu)$. Otherwise $\mathrm{supp}(\iota \circ \rho(\mu)) = \mathrm{supp}(\mu) \cup \bullet$. Regardless, $\mathrm{supp}(\mu) \cup \mathrm{supp}(\iota \circ \rho(\mu))$ is a simplex in $S$ by assumption. It then follows from [3, Lemma 3.9] that the homotopy $H$ is continuous. □

**Corollary 6.4.9.** *For any metric spaces $X$ and $Y$, the wedge sum factors through the Vietoris–Rips and Čech metric thickenings:*

$$\mathcal{VR}(X \vee Y; r) \simeq \mathcal{VR}(X; r) \vee \mathcal{VR}(Y; r)$$

$$\check{\mathcal{C}}(X \vee Y; r) \simeq \check{\mathcal{C}}(X; r) \vee \check{\mathcal{C}}(Y; r).$$

*Proof.* The Vietoris–Rips case follows since $\mathrm{VR}(X \vee Y; r) \supseteq \mathrm{VR}(X; r) \vee \mathrm{VR}(Y; r)$, and since if $\sigma \in \mathrm{VR}(X \vee Y; r)$ is not a subset of either $X$ or $Y$, then $\sigma \cup \bullet \in \mathrm{VR}(X \vee Y; r)$. The Čech case is analogous. □

## 6.5 Dowker's Theorem

A classical result about simplicial complexes is Dowker's theorem. This section presents a direct proof of the analogous result for simplicial metric thickenings.

If $X$ and $Y$ are sets, a **relation**, $R$, is a subset of $X \times Y$ such that $\pi_1(R) = X$ and $\pi_2(R) = Y$. If $(x, y) \in R$, then $x$ is related to $y$, written $xRy$. Let $R(y) = \pi_1(\pi_2^{-1}(y))$, the set of points in $x$ to which $y$ is related, and define $R(x)$ analogously.

An relation defines two simplicial complexes, $D_X$ and $D_Y$, as follows: the vertex set of $D_X$ is $X$, and a finite $\sigma \subseteq X$ is in $D_X$ if $\sigma \subseteq R(y)$ for some $y \in Y$. In parallel, $D_Y$ has vertex set $Y$ and a simplex $\sigma$ whenever $\sigma \subseteq R(x)$.

**Theorem 6.5.1** (Dowker)**.** *There is a homotopy equivalence* $D_X \simeq D_Y$.

The original proof is in [63].

The motivation for Dowker's theorem was to explain why several *a prior* different homology theories gave the same result. In fact, both the Vietoris–Rips and Čech complexes can be defined as Dowker complexes.

**Example 6.5.2.** Chapter 2 noted that the Čech complex admits two definitions—as a nerve complex, or by containment in open balls. These are the two Dowker complexes of a relation. Let $X$ be a metric space, and let $B = \{\mathrm{B}_r(x) \mid x \in X\}$. Define a relation between $X$ and $B$ by $(x, b) \in R$ if $x \in b$. Then $D_X$ contains a simplex for every finite set of points $\sigma$ for which $\sigma \subseteq \mathrm{B}_r(x)$ for some $x \in X$. Alternatively, $D_B$ contains a simplex for every finite collection of balls which all contain some $x \in X$, i.e. for which the intersection is nonempty. Both $D_X$ and $D_B$ are called the Čech complex as they happen to be not only homotopy equivalent but equal.

By using more general open sets two complexes can be defined by the same relation which are not equal, but must still be homotopy equivalent, per Theorem 6.5.1. Replacing open balls with all sets of diameter at most $r$ gives the Vietoris–Rips complex as a Dowker complex. Since the set of all sets of bounded diameter unwieldy, usually only $D_X$ is described as the Vietoris–Rips complex. ♦

A relation is **symmetrically finite** if $R(y)$ is finite for all $y$ and $R(x)$ is finite for all $x$. A symmetrically finite relation defines locally-finite simplicial complexes, and so any simplicial metric thickenings $\mathscr{X} = (X, D_X, \mathrm{id})$ and $\mathscr{Y} = (Y, D_Y, \mathrm{id})$ have $\mathscr{X}^{\mathrm{m}} \cong |D_X|$ and $\mathscr{Y}^{\mathrm{m}} \cong |D_Y|$. Then Dowker's theorem says that $\mathscr{X}^{\mathrm{m}} \simeq \mathscr{Y}^{\mathrm{m}}$. The following gives a direct proof of this result without appealing to either Dowker's theorem or the homeomorphism between locally-finite simplicial complexes and simplicial metric thickenings. The proof proceeds by working at the level of metric realizations.

**Definition 6.5.3.** *Let $\mathscr{X}$ and $\mathscr{Y}$ be simplicial metric thickenings. Two maps $f, g\colon \mathscr{X}^{\mathrm{m}} \to \mathscr{Y}^{\mathrm{m}}$ are **contiguous** if for all $\mu \in \mathscr{X}^{\mathrm{m}}$ and some $\varepsilon \in (0,1)$, the convex linear combination $(1-\varepsilon)f(\mu)+\varepsilon g(\mu)$ is in $\mathscr{Y}^{\mathrm{m}}$.*

**Lemma 6.5.4.** *If $f$ and $g$ are continuous and contiguous, they are homotopic.*

*Proof.* Note that if the convex combination requirement holds for some $\varepsilon \in (0,1)$, then it necessarily holds for all because simplicial metric thickenings are closed under absolute continuity. Therefore $H(\mu, t) = (1-t)f(\mu) + tg(\mu)$ is a well-defined map $\mathscr{X}^{\mathrm{m}} \times I \to \mathscr{Y}^{\mathrm{m}}$.

Further, $H$ is continuous as follows: choose a point $\mu \in \mathscr{X}^{\mathrm{m}}$ and let $N = W_2(f(\mu), g(\mu))$. Let $\nu$ be another measure in $\mathscr{X}^{\mathrm{m}}$. Let $\gamma_f$ be the optimal transport plan from $f(\mu)$ to $f(\nu)$ and $\gamma_g$ the optimal plan from $g(\mu)$ to $g(\nu)$. By the triangle inequality, the optimal transport plan from $f(\mu)$ to $g(\nu)$ has cost at most $N + \mathrm{cost}(\gamma_g)$. Assume that $s > t$. Then a transport plan from

$$H(\mu, t) = (1-t)f(\mu) + tg(\mu) \text{ to } H(\nu, s) = (1-s)f(\mu) + sg(\mu)$$

is given by

$$\alpha := (1-s)\gamma_f + t\gamma_g + (s-t)\gamma_{fg}.$$

Then $\mathrm{cost}(\alpha) \le (1-s)W(f(\mu), f(\nu)) + tW(g(\mu), g(\nu)) + (s-t)N$.

Let $\varepsilon > 0$. By the continuity of $f$ and $g$ there exists a $\delta > 0$ such that each of $W_2(f(\mu), f(\nu))$, $W_2(g(\mu), g(\nu))$, and $|s-t|$ are less than $\varepsilon/(1-s+t+N)$. This establishes the inequality

$$\mathrm{cost}(\alpha) < \frac{(1-s)\varepsilon}{(1-s+t+N)} + \frac{t\varepsilon}{(1-s+t+N)} + \frac{N\varepsilon}{(1-s+t+N)} = \varepsilon.$$

A similar argument holds assuming $s \le t$, giving the continuity of $H$. □

Let $\mathscr{B} \subseteq \mathscr{X}^{\mathrm{m}}$ be the space of all uniform distributions in $\mathscr{X}^{\mathrm{m}}$. (Since $\mathscr{X}^{\mathrm{m}}$ only contains finitely-supported measures, this is the set of measures of the form $\mu = \frac{1}{n+1}\sum_{i=0}^{n} \delta[x_i]$.)

Since $\mathscr{P}(X)$ is a metric space, one can construct the space $\mathscr{P}(\mathscr{P}(X))$, that is the Wasserstein space of measures *of measures*. There is a natural map $\mathscr{A} : \mathscr{P}(\mathscr{P}(X)) \to \mathscr{P}(X)$ given by averaging: for any Borel set $S \subseteq X$,

$$\mathscr{A}(\mu)(S) := \int_{\mathscr{P}(X)} \nu(S)\, \mathrm{d}\mu(\nu). \tag{6.1}$$

(This map is the unit in the monad structure on the category of metric spaces discussed in [60].)

**Lemma 6.5.5.** *Let $\mu = \sum_{i=1}^{n} \lambda_i \delta[x_i] \in \mathscr{P}(X)$. Then there exists a unique measure $\phi(\mu) \in \mathscr{P}(\mathscr{B})$ such that $\mathscr{A}(\phi(\mu)) = \mu$ and $\phi(\mu)$ is supported on a sequence of measures $\nu_0, \ldots, \nu_k$ with*

$$\mathrm{supp}(\nu_0) \supset \mathrm{supp}(\nu_1) \supset \cdots \supset \mathrm{supp}(\nu_k).$$

*Proof.* Define $\widetilde{\phi}$ on finite *sets* by

$$\widetilde{\phi}(\{\mathrm{x}_0, \ldots, \mathrm{x}_\mathrm{n}\}) = \delta\left[\frac{1}{n+1}\sum_{i=0}^{n} \delta[x_i]\right].$$

Suppose that $\mu = \sum_{i=1}^{n} \lambda_i \delta[x_i]$ is ordered so that $\lambda_i \le \lambda_j$ whenever $i < j$. Define

$$\phi(\mu) = \sum_{i=1}^{n} a_i \widetilde{\phi}(\{\mathrm{x_i}, \ldots, \mathrm{x_n}\})$$

where

$$a_1 = n\lambda_1 \text{ and } a_i = (\lambda_i - \lambda_{i-1})(n - i + 1).$$

Then $\phi(\mu)$ satisfies both conditions by construction. □

Intuitively $\phi$ changes the measure $\mu$ into barycentric coordinates. When $\mu$ is a uniform distribution, $\phi(\mu) = \delta[\mu]$. Consequently, call $\phi(\mathscr{X}^{\mathrm{m}})$ the **metric barycentric subdivision** of $\mathscr{X}^{\mathrm{m}}$.

**Lemma 6.5.6.** *The barycentric subdivision of $\mathscr{X}^{\mathrm{m}}$ is isometric to $\mathscr{X}^{\mathrm{m}}$.*

*Proof.* The map $\phi$ is an isometry, since for any optimal transference plan $\gamma$, $\phi(\gamma)$ is an optimal transference plan. □

**Definition 6.5.7.** *Let $R$ be a symmetrically finite relation between metric spaces $X$ and $Y$. Define the Dowker simplicial metric thickenings $R_{\mathscr{X}}$ and $R_{\mathscr{Y}}$ of $R$ as follows. For each $y \in Y$, include all measures supported on $R(y)$ in $R_{\mathscr{X}}$, and for each $x \in X$ include all measures supported on $R(x)$ in $R_{\mathscr{Y}}$.*

**Theorem 6.5.8.** *Let $X$ and $Y$ be discrete metric spaces and $R$ a symmetrically finite relation. Then the simplicial metric thickenings $R_{\mathscr{X}}$ and $R_{\mathscr{Y}}$ are homotopy equivalent.*

*Proof.* First, take the spaces of barycenters, $\mathscr{B}_{\mathscr{X}}$ and $\mathscr{B}_{\mathscr{Y}}$ of $R_{\mathscr{X}}$ and $R_{\mathscr{Y}}$. Define $\psi\colon \mathscr{B}_{\mathscr{X}} \to \mathscr{B}_{\mathscr{Y}}$ by

$$\psi(\mu) = \widetilde{\phi}\left(\bigcap_{x \in \mathrm{supp}(\mu)} R(x)\right) \tag{6.2}$$

and similarly $\theta\colon \mathscr{B}_{\mathscr{Y}} \to \mathscr{B}_{\mathscr{X}}$ by

$$\theta(\nu) = \widetilde{\phi}\left(\bigcap_{y \in \mathrm{supp}(\nu)} R(y)\right) \tag{6.3}$$

where $\widetilde{\phi}$ is as in Lemma 6.5.5. Note that $\mathrm{supp}(\phi(\mu)) = \cap_{x \in \mathrm{supp}(\mu)} R(x)$ is non-empty by the construction of the Dowker complex. Extend $\psi$ and $\theta$ linearly across simplices. Continuity is immediate since $\mathscr{B}_{\mathscr{X}}$ and $\mathscr{B}_{\mathscr{Y}}$ are discrete. The compositions $\psi \circ \theta$ and $\theta \circ \psi$ are contiguous to the identity maps by construction.

$\square$

# Chapter 7

# Conclusion

There are many interesting questions about simplicial metric thickenings in general, and Vietoris–Rips metric thickenings in particular, which remain to be answered.

The homotopy type of $\mathrm{VR}_<\left(\mathbb{S}^1; r\right)$ is known for all $r$—is the homotopy type of $\mathcal{VR}_<\left(\mathbb{S}^1; r\right)$ the same? More generally, are $\mathrm{VR}_<(X; r)$ and $\mathcal{VR}_<(X; r)$ always homotopy equivalent? This holds in all known examples, but is not yet proven. In a similar vein, the infinite Vietoris–Rips metric thickening, $\mathcal{VR}^{(\infty)}(X; r)$ was mentioned in Section 4.4. In all known examples, $\mathcal{VR}^{(\infty)}(X; r) \simeq \mathcal{VR}(X; r)$, but it is not known if this holds in general.

The theorems of Chapter 4 show that when $r$ is small the homotopy type of $\mathcal{VR}(X; r)$ can be determined. A significant, but difficult problem is the following: is there a general method to determine how the homotopy type of $\mathcal{VR}(X; r)$ changes once that bound is passed? Perhaps preliminary to this is to determine precisely what the scale parameter causing a change in homotopy type is, in general. All spaces for which this is currently known (for either $\mathcal{VR}(X; r)$ or $\mathrm{VR}(X; r)$) possess a high degree of symmetry. We conjecture that a more precise analysis of local curvature will be important to answering these questions in any form.

The Morse lemma proven in Chapter 5 is at best the beginning of a theory. Classical Morse theory possesses also the second Morse lemma, which states that when a critical point of index $j$ is passed in the sublevel sets, the homotopy type changes by the addition of a $j$-dimensional cell. Is there an analogous statement that can be made in the setting of Theorem 5.2.2? One difficulty is that lower semi-continuous, $\lambda$-convex functions possess more variety of critical points than classical Morse functions. It is almost certainly necessary to restrict to some subset of functions which have "well-behaved" critical points, but the appropriate restriction is far from clear.

The categorical constructions in Chapter 6 leave several paths to explore. A laudable goal is to develop a better understanding of the relation between the homotopy type of a metric realization of a simplicial metric thickening, and the homotopy type of the metric space and simplicial complex which define that structure. A naïve conjecture is that if $\mathscr{M} = (X, K, \phi)$ is a simplicial metric thickening, and $K$ is homotopy equivalent to some other complex $L$ via simplicial collapse, then $\mathscr{M}^{\mathrm{m}}$ should be homotopy equivalent to $|L|$. However, this statement is almost certainly too strong. Additionally, a stronger version of the Dowker's theorem in Section 6.5, in particular one admitting non-finite relations, should be possible. This could lead to a much richer theory of spaces like simplicial metric thickenings, but in which measures of infinite support are permitted, as indeed $\mathscr{V}\mathscr{R}^{(\infty)}(X; r)$ is already an example.

# Bibliography


[1] Mikhail Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in Group Theory*, pages 75–263. Springer New York, New York, NY, 1987.

[2] Jean-Claude Hausmann. On the Vietoris–Rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995.

[3] Michał Adamaszek, Henry Adams, and Florian Frick. Metric reconstruction via optimal transport. *SIAM Journal on Applied Algebra and Geometry*, 2(4):597–619, 2018.

[4] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[5] Paul Alexandroff. Über den allgemeinen Dimensionsbegriff und seine Beziehungen zur elementaren geometrischen Anschauung. *Mathematische Annalen*, 98:617–635, 1928.

[6] Leopold Vietoris. Über den höheren zusammenhang kompakter räume und eine klasse von zusammenhangstreuen abbildungen. *Mathematische Annalen*, 97(1):454–472, 1927.

[7] Heinrich Reitberger. Leopold Vietoris (1891-2002). *Notices of the American Mathematical Society*, 49(10):1232–1236, 2002.

[8] Janko Latschev. Vietoris–Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001.

[9] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[10] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

[11] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009.

[12] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[13] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.

[14] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.

[15] Robert W. Ghrist. *Elementary applied topology*, volume 1. Createspace Seattle, 2014.

[16] Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

[17] Michał Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56(2):251–273, 2016.

[18] Michał Adamaszek and Henry Adams. The Vietoris–Rips complexes of a circle. *Pacific Journal of Mathematics*, 290(1):1–40, 2018.

[19] Michał Adamaszek, Henry Adams, and Samadwara Reddy. On Vietoris–Rips complexes of ellipses. *Journal of Topology and Analysis*, 11(03):661–690, 2019.

[20] Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier. A complete characterization of the one-dimensional intrinsic čech persistence diagrams for metric graphs. In *Research in Computational Topology*, pages 33–56. Springer, 2018.

[21] Michał Adamaszek, Henry Adams, Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier. Vietoris-Rips and Čech complexes of metric gluings. In *34th International Symposium on Computational Geometry (SoCG 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[22] Žiga Virk. 1-dimensional intrinsic persistence of geodesic spaces. *arXiv preprint arXiv:1709.05164*, 2017.

[23] Matthew C.B. Zaremsky. Bestvina-brady discrete morse theory and vietoris-rips complexes. *arXiv preprint arXiv:1812.10976*, 2018.

[24] Sunhyuk Lim, Facundo Memoli, and Osman Berat Okutan. Vietoris-rips persistent homology, injective metric spaces, and the filling radius. *arXiv preprint arXiv:2001.07588*, 2020.

[25] Mikhail Katz. Diameter-extremal subsets of spheres. *Discrete & Computational Geometry*, 4(2):117–137, 1989.

[26] Katsuro Sakai. *Geometric Aspects of General Topology*. Springer Monographs in Mathematics. Springer, 2013.

[27] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley-Interscience, 2 edition, 1992.

[28] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[29] Leonid Vasilevich Kantorovich and Gennady S. Rubinstein. On a space of completely additive functions. *Vestnik Leningrad. Univ*, 13(7):52–59, 1958.

[30] Cédric Villani. *Topics in Optimal Transportation*. Number 58 in Graduate Studies in Mathematics. American Mathematical Society, 2003.

[31] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser, 2015.

[32] Cédric Villani. *Optimal transport: old and new*. Springer, 2008.

[33] Hans G. Kellerer. Duality theorems and probability metrics. In *Proceedings of the Seventh Conference on Probability theory, Braşov, Romania*, pages 211–220, 1985.

[34] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[35] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[36] Robert J. McCann. Polar factorization of maps on riemannian manifolds. *Geometry and Functional Analysis*, 11(3):589–608, 2001.

[37] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential.* Hermann, 1978.

[38] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.

[39] Karsten Grove and Hermann Karcher. How to conjugate $C^1$-close group actions. *Mathematische Zeitschrift*, 132(1):11–20, 1973.

[40] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[41] Jürgen Jost. *Riemannian Geometry and Geometric Analysis.* Universitext. Springer, 7 edition, 2017.

[42] Hermann Karcher. Riemannian center of mass and so called Karcher mean. *arXiv preprint arXiv:1407.2087*, 2014.

[43] Shin-Ichi Ohta. Barycenters in Alexandrov spaces of curvature bounded below. *Advances in Geometry*, 14(4):571–587, 2012.

[44] Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces: Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs: April 16-July 13, 2002, Emile Borel Centre of the Henri Poincaré Institute, Paris, France*, 338:357, 2003.

[45] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

[46] Martin Bridson and André Haefliger. *Metric Spaces of Non-Positive Curvature*. Number 319 in Grundlehren der mathematischen Wissenscchaften. Springer, 1999.

[47] David Groisser. Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Advances in Applied Mathematics*, 33(1):95–135, 2004.

[48] Bijan Afsari. Riemannian $L^p$ center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

[49] Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, 2013.

[50] Michał Adamaszek, Henry Adams, Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier. On homotopy types of Vietoris–Rips complexes of metric gluings. To appear in Journal of Applied and Computational Topology, 2020.

[51] Henry Adams, Johnathan Bush, and Florian Frick. Metric thickenings, Borsuk–Ulam theorems, and orbitopes. *Mathematika*, 66:79–102, 2020.

[52] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[53] Wilfrid Gangbo, Hwa Kil Kim, and Tommaso Pacini. *Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems*. American Mathematical Society, 2010.

[54] Shin-ichi Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *American journal of mathematics*, 131(2):475–516, 2009.

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

[56] John Milnor. *Morse Theory*. Annals of Mathematics Studies. Princeton University Press, 1969.

[57] Karl-Theodor Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *Journal de Mathématiques Pures et Appliquées*, 84(2):149–168, 2005.

[58] Emily Riehl. *Category Theory in Context*. Courier Dover Publications, 2017.

[59] Dmitry N. Kozlov. *Combinatorial Algebraic Topology*, volume 21 of *Algorithms and Computation in Mathematics*. Springer, 2008.

[60] Tobias Fritz and Paolo Perrone. A probability monad as the colimit of spaces of finite samples. *Theory and Applications of Categories*, 34(7):170–220, 2019.

[61] F. William Lawvere. Metric spaces, generalized logic, and closed categories. *Reprints in Theory and Applications of Categories*, 1:1–37, 2002.

[62] Michael Lesnick, Raúl Rabadán, and Daniel I.S. Rosenbloom. Quantifying genetic innovation: Mathematical foundations for the topological study of reticulate evolution. *SIAM Journal on Applied Algebra and Geometry*, 4(1):141–184, 2020.

[63] C.H. Dowker. Topology of metric complexes. *American Journal of Mathematics*, 74(3):555–577, 1952.