Geometry of generalised spaces of persistence diagrams and optimal partial transport for metric pairs

CHE MOGUEL, MAURICIO,ADRIAN

Geometry of generalised spaces of persistence diagrams and optimal partial transport for metric pairs

Mauricio Che Moguel

A thesis presented for the degree of

Doctor of Philosophy

Department of Mathematical Sciences

University of Durham

United Kingdom

2024

Geometry of generalised spaces of persistence diagrams and optimal partial transport for metric pairs

Mauricio Che Moguel

Abstract

In this thesis, we study the geometry of two families of metric spaces that can be defined over a metric pair. We first focus on generalised spaces of persistence diagrams over metric pairs. We prove that the construction of these metric spaces is functorial and preserves certain geometric properties of the underlying space, namely completeness, separability, geodesicity, and non-negative curvature in the sense of Alexandrov. We also study the continuity of these constructions with respect to Gromov–Hausdorff convergence. We then move on to spaces of Radon measures endowed with the optimal partial transport metrics introduced by Figalli and Gigli. We adapt results from Figalli and Gigli's work to the class of proper metric pairs. Furthermore, we prove that when endowed with the L^2-optimal partial transport distance, the resulting space of Radon measures is a non-negatively curved Alexandrov space, whenever the underlying space has the same property. This result is new, even in the Euclidean setting considered by Figalli and Gigli. Finally, in an appendix, we study basic properties of Gromov–Hausdorff convergence for metric pairs. We prove that this convergence is metrisable in the context of proper metric pairs, and present versions of the classical embedding, completeness, and precompactness theorems.

Supervisor: Fernando Galaz-García

A la memoria de mi abuelo, Juan Gualberto Che Lara.

Acknowledgements

It might be impossible for me to thank everyone who has directly or indirectly contributed to this work. I apologise to those I unintentionally leave unmentioned. I hope this work reflects the support and encouragement I have received.

First and foremost, I want to thank my supervisor, Fernando Galaz-García, who has guided me through the world of mathematics and academia to a greater extent than I could have ever imagined. Since the first day we were in contact, his kindness, generosity, and willingness to help have made this whole experience incredibly enriching. He has always gone the extra mile to keep me motivated, which I have needed more often than I would like to admit. He has introduced me to so many people and encouraged me to attend and present my work at several conferences and events, always pushing me to give my best. He would always smile at me and say: *Recuerda ¡Cobra Kai!* to remind me to stand up and keep fighting.

Ni esta tesis, ni la persona que soy, podrían existir sin el apoyo y la presencia incondicionales de mi familia. Mamá y papá: no me alcanzará la vida para agradecerles por ser quienes son, por nunca dejar que me alejara de mis metas y enseñarme la importancia de la constancia, entre tantos otros valores que ahora guían mi vida. A mis hermanos y mis sobrinos, por siempre estar pendientes de mí, y recibirme en casa como si no hubiera pasado el tiempo, cada una de las veces que he tenido la oportunidad de visitarles desde que empecé esta aventura. Y a mi abuelita, por ser un pilar para mi familia.

A mi querida Paula, por tu apoyo y amor constantes. Han sido años con altibajos para ambos, y estar lejos ha sido uno de los retos más difíciles, pero en todo este tiempo nunca has dejado de apoyarme como lo hiciste desde el primer momento. Has sido para mí un lugar seguro y una fuente inagotable de motivación, especialmente cuando más he dudado de mí mismo. No podría estar terminando de escribir esta tesis de una mejor manera que contigo a mi lado, en Durham, finalmente, después de más de tres años de intentar que conocieras este pintoresco pueblito y un poco de lo que aquí he vivido.

Thanks to the staff, PhD students, and postdocs in the Department of Mathematical Sciences who have enriched my academic experience. Special thanks to Martin Kerin, Kohei Suzuki, Wilhelm Klingenberg, Norbert Peyerimhoff, John Parker, Alpár Mészáros, Amit Einav, Mo Dick Wong, Jaime Santos-Rodríguez, Irving Calderón, Mohammad Al Attar, Dan Disney, and Léonie Papon.

A massive thanks to my co-authors, collaborators, mentors and mathematical friends who have supported me and with whom I have had the invaluable opportunity to work and discuss mathematics, either in Durham, in Mexico, or beyond. Thanks to Jesús Núñez-Zimbrón, Oscar Palmas Velasco, José Carlos Gómez Larrañaga, Omar Muñiz, Noé Bárcenas, and Didier Solís Gamboa for all your support in Mexico—especially to José Carlos and Didier for hosting me during my research visits in Merida, and to Jesús for our exciting projects together. Thanks to Luis Guijarro for hosting me in Madrid and for all the enjoyable conversations and advice. Special thanks to Andrés Ahumada Gómez and Manuel Mellado Cuerno for being not only great collaborators, but my mathematical siblings. Also, thanks to Ingrid Amaranta Membrillo-Solis and Motiejus Valiunas, for our paper together and for all your support. Finally, thanks to David Bate, Ana Lucía García Pulido, John Harvey, Catherine Searle, Diego Corro, and Raquel Perales, for various invitations to give talks at different events, as well as for many fruitful conversations.

During these years, Durham has become not only the place where I completed my PhD thesis but also a second home, thanks to the amazing friends I have made here. Special thanks to Los Aluxes del Norte, with whom I had such good times playing the best live music in Durham, and to all my Mexican and not-so-Mexican friends here, for the carnitas asadas, pub crawls, salsa dancing, hiking, sports, formals, random weekends abroad, and an endless list of experiences that made the Durham journey feel like a dream come true.

A mis amigos en Mérida y México por seguirme apoyando durante estos años, muchísimas gracias por tanto y perdón por tan poco. Especialmente gracias a mis *Casi Amigos*, por ser mi segunda familia para la posteridad. A Jhony y Luis, porque en verdad no estaría donde estoy si no hubieran sido mis mentores cuando empecé en la OMM, hace ya más de diez años.

Finally, thanks to the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) for sponsoring my PhD research via the Doctoral Scholarship No. 769708.

Contents

Declaration

The work in this thesis is based on research carried out at the Department of Mathematical Sciences, University of Durham, England. No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is the sole work of the author unless referenced to the contrary in the text.

Some of the work presented in this thesis has been published, accepted in journals, or is available as preprints in public repositories - the relevant publications are listed below.

- Andrés Ahumada Gómez, **Mauricio Che**. *Gromov-Hausdorff convergence of metric pairs and metric tuples.* Differential Geometry and its Applications 94:Paper No. 102135, 29, 2024. DOI:10.1016/j.difgeo.2024.102135.
- **Mauricio Che**. *Optimal partial transport for metric pairs*, 2024. Preprint. arXiv:2406.17674.
- **Mauricio Che**, Fernando Galaz-García, Luis Guijarro, Ingrid Membrillo Solis. *Metric geometry of spaces of persistence diagrams.* Journal of Applied and Computational Topology, 2024. DOI:10.1007/s41468-024-00189-2.
- **Mauricio Che**, Fernando Galaz-García, Luis Guijarro, Ingrid Membrillo Solis, Motiejus Valiunas. *Basic metric geometry of the bottleneck distance.* Proceedings of the American Mathematical Society 152(8):3575–3591, 2024. DOI:10.1090/proc/16776.

Chapter 1

Introduction

Two of the most fundamental concepts in mathematics are those of metric and measure. The interaction between these notions is at the heart of the theory of *optimal transport*.

Originating in the work of Gaspard Monge, optimal transport formalises the idea of comparing distribution of masses by taking into account the geometry of the space (see [15, 54] for a historical summary). This level of abstraction is sufficiently robust to allow for applications in different contexts. For instance, during the 80s and 90s, Brenier [6, 7], Jordan, Kinderlehrer and Otto [35] established connections between optimal transport and partial differential equations modelling physical phenomena such as fluid mechanics and gas dynamics.

Another important application of optimal transport is related to the concept of curvature. Namely, during the late 90s and early 2000s, due to the combined work of Cordero-Erausquin, McCann, Schmuckenschläger, [21, 22], Otto, Villani [45], von Renesse and Sturm [55], it was discovered that there is an equivalence between the existence of lower Ricci curvature bounds on a Riemannian manifold and the convexity of certain entropy functionals on the space of probability measures over that manifold, endowed with a metric induced by an optimal transport problem. This in turn led to the development of the theory of lower Ricci curvature bounds on general metric measure spaces by Lott, Sturm and Villani [39, 50, 51].

One more instance where optimal transport has found applications is in the fast-growing field of Topological Data Analysis (TDA). Namely, optimal transport can be used to define metrics between *persistence diagrams*. We now give a brief description of this setting.

Consider collections of simplicial complexes $\{K_t\}_{t\in\mathbb{R}}$ and maps $\{\iota^{s,t}\colon K_s \to K_t\}_{s\leq t}$ such that $\iota^{s,t} \circ \iota^{r,s} = \iota^{r,t}$ for any $r \leq s \leq t$, and $\iota^{s,s} = \mathrm{id}_{K_s}$. Common simplicial filtrations are given by Vietoris–Rips complexes associated to given point-clouds in $\mathbb{R}^n$, or by looking at sublevel sets of sufficiently nice functions on a topological space (see, for example, [25]). By looking at the homology groups of simplicial complexes in such a filtration, we get a one-parameter family of modules $\{H_\bullet(K_t)\}_{t\in\mathbb{R}}$ and linear maps $\{\iota^{s,t}_{\#}\colon H_\bullet(K_s) \to H_\bullet(K_t)\}_{s\leq t}$, which together define a *persistent module*. Under very general assumptions, it is possible to describe such a persistent module with a multiset of points in $\mathbb{R}^2_{\geq} = \{(x, y) \in \mathbb{R}^2 : x \leq y\}$, which is what we call the *persistence diagram* associated to the filtration (see [23] for details).

For statistical applications, it is useful to consider the set of all persistence diagrams and endow it with a metric (see, for example, [41]). We can define a one-parameter family of metrics between persistence diagrams, motivated by optimal transport, as in the following definition.

Definition 1.0.1. Given two persistence diagrams σ, τ, and some $p \in [1, \infty)$, the *L^p-Wasserstein metric* between σ and τ is

$$d_p(\sigma, \tau) = \inf_{\phi\colon \sigma'\to\tau'} \left(\sum_{x\in\sigma} |x - \phi(x)|^p\right)^{1/p}, \tag{1.1}$$

where $\phi\colon \sigma' \to \tau'$ runs over all possible bijections between $\sigma' \sim \sigma$ and $\tau' \sim \tau$, where $\sigma' \sim \sigma$ means that σ' and σ only differ by points in $\Delta = \{(x, y) \in \mathbb{R}^2 : x = y\}$. The function d_p defines a metric on the set of equivalence classes of persistence diagrams σ satisfying

$$\sum_{x\in\sigma} \mathrm{dist}(x, \Delta)^p < \infty,$$

which defines the *space of p-persistence diagrams*, denoted by $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$.

The definition of d_p is motivated by the fact that usually, in applications, points in a persistence diagram that are close to Δ may be regarded as noise in the corresponding filtration, whereas points far from Δ represent more relevant topological features.

The spaces $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ have been studied by several authors. We know, for instance, that $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ is complete, separable and geodesic for any choice of $p \in [1, \infty)$ [41]. There are also explicit descriptions of geodesics in $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ [20]. Moreover, $\mathcal{D}_2(\mathbb{R}^2_{\geq}, \Delta)$ is an Alexandrov space with non-negative curvature [53]. These properties have been used to address the existence of barycentres of probability measures defined on $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$, as well as to device algorithms converging to such barycentres [41, 53]. Other approaches have been oriented towards the study of embeddings of the space of persistence diagrams into vector spaces, with the hope of more computationally efficient methods [4, 11, 42, 56]. In this direction, it is also important to understand the geometry of the space of persistence diagrams to get information about such maps.

Coming back to optimal transport, an important restriction for the classical formulation is that it is applicable only to measures with identical total mass, and it is therefore interesting to explore ways to define optimal transport between unbalanced measures. Different approaches have been proposed (see, for example, [14, 19, 26, 27, 31, 38] and references therein), and very recently, Savaré and Sodini, in [49], presented a general framework that includes previous approaches in the case of finite Radon measures.

In [27], Figalli and Gigli introduced a generalisation of optimal transport to the setting of positive Radon measures on bounded domains in Euclidean space, motivated by finding solutions to evolution equations with Dirichlet boundary conditions, by interpreting such equations as gradient flows of certain energy functionals defined on spaces of positive Radon measures endowed with optimal partial transport metrics, in the spirit of Jordan–Kinderlehrer-Otto scheme [35]. The idea behind optimal partial transport on a bounded domain $\Omega \subset \mathbb{R}^n$ is that we can use the boundary $\partial\Omega$ as an infinite source of mass that we can freely use to compensate the difference in the total mass of the measures we are comparing, provided that we pay the extra cost of transporting mass from and to it.

More recently, in [24], Divol and Lacombe extended this theory to unbounded domains $\Omega \subset \mathbb{R}^2$, obtaining spaces of Radon measures that we will denote by $\mathcal{M}_p(\overline{\Omega}, \partial\Omega)$ (see Chapter 4 for precise definitions), and proved that the space of persistence diagrams $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ can be isometrically embedded into the space of measures $\mathcal{M}_p(\mathbb{R}^2_{\geq}, \Delta)$.

Our contributions

In the first part of this thesis, which is based on the articles [17, 18], we study the following generalisation of the spaces $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$. Consider a *metric pair* (X, A), i.e. an ordered pair consisting of a metric space X and a closed, non-empty $A \subset X$. Then we can define persistence diagrams in the metric pair (X, A) as finite or countably infinite multisets of points in X, and define, for any $p \in [1, \infty)$, the L^p-Wasserstein metric between persistence diagrams in analogy to equation (1.1). This defines a one-parameter family of metric spaces $\mathcal{D}_p(X, A)$. We study properties of $\mathcal{D}_p(X, A)$ inherited from the underlying metric pair (X, A). The following theorem generalises known results for the spaces $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ and its proof follows similar ideas to those in [20, 41, 52, 53]. Also note the overlap with [10], which is independent from [17, 18] and came out after earlier versions of the latter articles.

Theorem A. *Let (X, A) be a metric pair and $p \in [1, \infty)$. Then the following conditions hold:*

1. *$\mathcal{D}_p(X, A)$ is complete whenever X is complete.*

2. *$\mathcal{D}_p(X, A)$ is separable whenever X is separable.*

3. *$\mathcal{D}_p(X, A)$ is geodesic whenever X is proper and geodesic.*

4. *$\mathcal{D}_2(X, A)$ is a non-negatively curved Alexandrov space whenever X is a proper non-negatively curved Alexandrov space.*

A particularly interesting consequence of Theorem A is that $\mathcal{D}_2(X, A)$ inherits the structure of Alexandrov space from the underlying space X, at least in the non-negatively curved case. Alexandrov spaces are generalisations of Riemannian manifolds with sectional curvature uniformly bounded from below. One of the aspects in which this spaces are similar to Riemannian manifolds is that it is possible to define the space of directions at any given point, which is a natural generalisation of the set of unit tangent vectors at a given point in a Riemannian manifold. We discuss the following result relating spaces of directions of the space $\mathcal{D}_2(X, A)$ with the spaces of directions of the underlying space X.

Theorem B. *Let (X, A) be a metric pair such that X is a proper, non-negatively curved Alexandrov space. Let $\sigma, \sigma' \in \mathcal{D}_2(X, A)$ and $\xi, \xi' \in \mathrm{Geo}(\mathcal{D}_2(X, A))$ be such that $\xi_0 = \xi'_0 = \sigma_A$, $\xi_1 = \sigma$, $\xi'_1 = \sigma'$, and such that $\xi_t = \{\{\xi_t^x : x \in \sigma\}\}$ and $\xi'_t = \left\{\left\{\xi_t^{x'} : x' \in \sigma'\right\}\right\}$ for some $\{\xi^x\}_{x\in\sigma}$, $\{\xi^{x'}\}_{x'\in\sigma'} \subset \mathrm{Geo}(X)$. Then*

$$d_2(\sigma, \sigma_A) d_2(\sigma', \sigma_A) \cos \measuredangle(\xi, \xi') = \sum_{x\in\tau} d(x, A) d(\phi(x), A) \cos \measuredangle(\xi^x, \xi^{\phi(x)}),$$

for some bijection $\phi \colon \tau \to \tau'$ between $\tau \subset \sigma$ and $\tau' \subset \sigma'$, and such that $\xi_0^x = \xi_0^{\phi(x)}$ for all $x \in \tau$.

It is also natural to consider $\mathcal{D}_\infty(X, A)$, the set of persistence diagrams such that $\sup_{x\in\sigma} \mathrm{dist}(x, A) < \infty$, endowed with the *bottleneck distance*, i.e. the function we obtain in equation (1.1) when we replace the L^p-norm of $\{d(x, \phi(x))\}_{x\in\sigma}$ by the corresponding L^∞-norm. In general, $\mathcal{D}_\infty(X, A)$ is not a metric space, but a pseudometric space. This makes the proofs of some of its geometric properties more technical, and we do not carry out those details in this thesis. Such discussion can be found in [18]. However, we address the continuity of the maps $(X, A) \mapsto \mathcal{D}_p(X, A)$, for $p \in [1, \infty]$, with respect to the Gromov–Hausdorff convergence. We obtain the following result.

Theorem C. *Let* $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ *be a sequence of metric pairs converging Gromov–Hausdorff to a metric pair* (X, A) *(see Definition 3.7.1). Then the sequence* $\{(\mathcal{D}_\infty(X_i, A_i), \sigma_{A_i})\}_{i\in\mathbb{N}}$ *is pointed Gromov–Hausdorff convergent to* $(\mathcal{D}_\infty(X, A), \sigma_A)$*. On the other hand, for any* $p \in [1, \infty)$ *there are convergent sequences* $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ *such that the sequence* $\{(\mathcal{D}_p(X_i, A_i), \sigma_{A_i})\}_{i\in\mathbb{N}}$ *is not convergent. In particular, when restricted to proper metric pairs (i.e.* (X, A) *such that* X *is proper),* $\mathcal{D}_p$ *is continuous if and only if* $p = \infty$*.*

The notion of Gromov–Hausdorff convergence for metric pairs mentioned in Theorem C was introduced in [17]. This concept, which is of interest by itself, is further studied for the sake of completeness in Appendix A, based on [1]. The main result in the Appendix is the metrisability of the Gromov–Hausdorff convergence of proper metric pairs, which proves that the map $(X, A) \mapsto (\mathcal{D}_\infty(X, A), \sigma_A)$ is not only sequentially continuous, but continuous, when restricted to proper metric pairs. More precisely, we prove the following.

Theorem D. *There is a metric* d_{GH} *in the class of proper metric pairs which induces the corresponding Gromov–Hausdorff convergence.*

In the second part of this thesis, which is based on the manuscript [16], we consider a natural extension of the optimal partial transport problem introduced by Figalli and Gigli in [27, Problem 1.1] to the setting of proper metric pairs. Namely, given a proper metric pair (X, A), and a parameter $p \in [1, \infty)$, the space $\mathcal{M}_p(X, A)$ of positive Radon measures μ on $X \setminus A$ satisfying

$$\int_{X\setminus A} d(x, A)^p \, d\mu(x) < \infty$$

can be endowed with the metric

$$\mathrm{Wb}_p(\mu, \nu) = \inf_{\gamma\in \mathrm{Adm}(\mu,\nu)} \left(\int d(x, y)^p \, d\gamma(x, y)\right)^{1/p}, \tag{1.2}$$

where $\mathrm{Adm}(\mu, \nu)$ is the set of partial transport plans between μ and ν (see Section 4.2 for precise definitions).

We adapt statements and proofs from [2, 24, 27] to obtain a self-contained exposition about basic aspects of optimal partial transport for proper metric pairs.

Theorem E. *Let (X, A) be a proper metric pair, and fix $p \in [1, \infty)$. Then for any $\mu, \nu \in \mathcal{M}_p(X, A)$ there exists an optimal partial transport plan, i.e. a minimiser in Equation (1.2). Moreover,* Wb_p *defines a metric on $\mathcal{M}_p(X, A)$.*

In the proof of Theorem E, we fix an oversight in the argument for [24, Proposition 3.2], following ideas from [2] (see the proofs of Theorem 4.2.7 and Theorem 4.2.10 for details).

We also prove a characterisation of optimal partial transport plans in terms of cyclical monotonicity and the existence of Kantorovich potentials, in analogy to the characterisation of classical optimal transport, adapting arguments from [2, 27] (see Theorem 4.3.1).

Based on [2, 27], we obtain that the spaces $\mathcal{M}_p(X, A)$ inherit properties of the metric pair (X, A), namely completeness, separability, geodesicity, the non-branching property, and non-negative curvature in the sense of Alexandrov (see Theorems 4.4.1, 4.5.1, 4.5.7, and 4.6.1 for details). We point out that item 4 in Theorem F below is new, even in the Euclidean setting considered by Figalli and Gigli in [27], although the argument is an adaptation of the proof of [2, Theorem 2.20].

Theorem F. *Let (X, A) be a proper metric pair, and fix $p \in [1, \infty)$. Then the following conditions hold:*

1. *$\mathcal{M}_p(X, A)$ is complete and separable.*
2. *If X is geodesic, then $\mathcal{M}_p(X, A)$ is geodesic.*
3. *If X is geodesic and non-branching, then $\mathcal{M}_p(X, A)$ is non-branching.*
4. *If X is a non-negatively curved Alexandrov space, then $\mathcal{M}_2(X, A)$ is a non-negatively curved Alexandrov space.*

We adapt [24, Proposition 3.5] to the setting of proper metric pairs, which yields that the generalised spaces of persistence diagrams are isometrically embedded into the spaces of optimal partial transport with the corresponding metrics (see Theorem 4.7.2 for details).

Theorem G. *The space $\mathcal{D}_p(X, A)$ can be isometrically embedded into $\mathcal{M}_p(X, A)$ for any metric pair (X, A) such that X is proper, and any $p \in [1, \infty)$.*

Observe that item 3 in Theorem F combined with Theorem G implies that $\mathcal{D}_p(X, A)$ is non-branching for any $p \in (1, \infty)$, whenever X itself is non-branching. This result is new in the general setting of proper metric pairs. Also observe that item 4 in Theorem F combined with Theorem G gives an alternative proof of item 4 in Theorem A.

Finally, as another consequence of Theorem G, we get that $\mathcal{M}_p(X, A)$ has infinite Hausdorff, covering, asymptotic, Assouad and Assouad–Nagata dimensions for any $p \in [1, \infty)$, whenever (X, A) is a proper metric pair satisfying the hypotheses of [17, Proposition 7.3]. In particular, item 4 in Theorem F, applied to proper metric pairs satisfying the hypotheses of [17, Proposition 7.3], yields a new, systematic way to construct infinite-dimensional Alexandrov spaces, which is a class of spaces that is not yet well-understood (see, for example, [43, 47, 57, 58]).

Chapter 2

Preliminaries

In this chapter we fix notation and terminology that we use in the rest of the thesis, following [2, 12, 13, 54] as the main references. In section 2.1 we describe general notions related to metric spaces. In section 2.2 we review Alexandrov spaces and related notions. In section 2.3 we recall the definition and basic properties of the Gromov–Hausdorff distance between metric spaces. Finally, in section 2.4 we deal with notation and basic results about classical optimal transport.

2.1 Basics of metric spaces

Let us recall some basic definitions and fix notation related to metric spaces.

Definition 2.1.1. Let X be a set. A map $d\colon X \times X \to [0, \infty]$ is an *extended pseudometric* on X if it is symmetric, satisfies the triangle inequality and $d(x, x) = 0$ for any $x \in X$. A *pseudometric* on X is an extended pseudometric d satisfying $d(x, y) < \infty$ for any $x, y \in X$. An *extended metric* on X is an extended pseudometric d such that $d(x, y) = 0$ if only if $x = y$. A *metric* on X is an extended metric such that $d(x, y) < \infty$ for any $x, y \in X$. We call (X, d) an *(extended, pseudo) metric space* if d is a (extended, pseudo) metric on X. Whenever d is clear from the context, we simply write X instead of (X, d).

The *open ball* of radius r around x is the set $B^d_r(x) = \{y \in X : d(x, y) < r\}$, whereas

$\overline{B}^d_r(x)$ denotes the corresponding *closed ball*. Moreover, for any $A \subset X$, we denote by $B^d_r(A) = \{x \in X : d(x, A) < r\}$ the *open neighbourhood* of radius r around A, whereas $\overline{B}^d_r(A)$ denotes the corresponding *closed neighbourhood*. Again, whenever d is clear from the context, we drop the dependence on d from the symbols above.

We say that a metric space X is *proper* if the closed ball $\overline{B}_r(p)$ is compact for any $r \geq 0$ and any $p \in X$.

Definition 2.1.2. We denote by Met the category whose objects are metric spaces and whose morphisms are Lipschitz maps. On the other hand, $\mathsf{Met}_{\mathsf{Pair}}$ is the category of *metric pairs*, i.e. ordered pairs (X, A) where X is a metric space and $A \subset X$ is closed and non-empty, and whose morphisms are *relative Lipschitz maps*, i.e. Lipschitz maps $f\colon X \to Y$ such that $f(A) \subset B$, where $(X, A), (Y, B) \in \mathsf{Met}_{\mathsf{Pair}}$. If we restrict our attention to metric pair (X, A) where A is a singleton, we talk about *pointed metric spaces* and *pointed Lipschitz maps*. We denote the category of pointed metric spaces by Met_*. Similarly, we define the categories $\mathsf{PMet}_{\mathsf{Pair}}$ and PMet_* of pseudometric pairs and of pointed pseudometric spaces, respectively.

Definition 2.1.3. Let X be a metric space. We denote by $\mathcal{C}([a, b], X)$ the space of continuous curves $\xi\colon [a, b] \to X$, endowed with the uniform metric. For any $t \in [a, b]$, $e_t\colon \mathcal{C}([a, b], X) \to X$ is the *evaluation map* given by $e_t(\xi) = \xi_t = \xi(t)$. The *length* of a continuous curve $\xi \in \mathcal{C}([a, b], X)$ is

$$\mathcal{L}(\gamma) = \sup \left\{ \sum_{i=0}^{n-1} d(\gamma(t_i), \gamma(t_{i+1})) \right\},$$

where the supremum is taken over all finite partitions $a = t_0 \leq t_1 \leq \cdots \leq t_n = b$ of the interval $[a, b]$. We say that X is a *length space* if

$$d(x, y) = \inf \{\mathcal{L}(\xi) : \xi \in \mathcal{C}([a, b], X),\ \xi_a = x,\ \xi_b = y\}. \tag{2.1}$$

Any continuous curve $\xi\colon [a, b] \to X$ with finite length can be re-parameterised with constant speed (see [12, Proposition 2.5.9]). This means that, in that case, we can assume that $[a, b] = [0, 1]$ and that there is some $v > 0$ such that, for any $s, t \in [0, 1]$

we have

$$\mathcal{L}(\xi|_{[s,t]}) = v|s-t|.$$

Definition 2.1.4. A *constant speed geodesic*, or simply a *geodesic*, is a continuous curve $\xi \in \mathcal{C}([0,1], X)$ such that

$$d(\xi_s, \xi_t) = d(\xi_0, \xi_1)|s-t|$$

for any $s, t \in [0,1]$. We denote by $\mathrm{Geo}(X)$ the space of geodesics in X, endowed with the uniform metric. We say that X is a *geodesic space* if for any $x, y \in X$ there exists $\xi \in \mathrm{Geo}(X)$ such that $\xi_0 = x$ and $\xi_1 = y$.

It is known that if X is a complete, separable and geodesic space, then $\mathrm{Geo}(X)$, endowed with the uniform metric, is complete and separable (see [2, Section 3.2]). Moreover, if X is a proper space then $\mathrm{Geo}(X)$ is proper as well, by the Arzelà-Ascoli theorem.

Definition 2.1.5. We say that a geodesic space X is *non-branching* if for any $t \in (0,1)$ the map $(e_0, e_t)\colon \mathrm{Geo}(X) \to X \times X$ is injective.

We conclude this section by briefly recalling the definition of the Hausdorff dimension (see [12, Section 1.7] for further details).

Let X be a metric space and denote the diameter of a subset $S \subset X$ by $\mathrm{diam}(S)$. For any $\delta \geq 0$, $\varepsilon > 0$ and $A \subset X$, let

$$\mathcal{H}^\delta_\varepsilon(A) = \inf\left\{\sum_{i\in\mathbb{N}} (\mathrm{diam}(S_i))^\delta : A \subset \bigcup_{i\in\mathbb{N}} S_i \text{ and } \mathrm{diam}(S_i) < \varepsilon\right\}.$$

If no such covering exists, then by convention $\mathcal{H}^\delta_\varepsilon(A) = \infty$.

Definition 2.1.6. The *δ-dimensional Hausdorff measure of A* is given by

$$\mathcal{H}^\delta(A) = \omega_\delta \cdot \lim_{\varepsilon\searrow 0} \mathcal{H}^\delta_\varepsilon(A),$$

where $\omega_\delta > 0$ is a normalisation constant such that, if δ is an integer n, the n-dimensional Hausdorff measure of the unit cube in n-dimensional Euclidean space

$\mathbb{R}^n$ is 1. The *Hausdorff dimension of* X is the number

$$\begin{aligned}\dim_{\mathcal{H}}(X) &= \sup\{\,\delta : \mathcal{H}^\delta(X) > 0\,\} = \sup\{\,\delta : \mathcal{H}^\delta(X) = \infty\,\}\\ &= \inf\{\,\delta : \mathcal{H}^\delta(X) = 0\,\} = \inf\{\,\delta : \mathcal{H}^\delta(X) < \infty\,\}.\end{aligned}$$

2.2 Alexandrov spaces

Alexandrov spaces are synthetic generalisations of Riemannian manifolds with sectional curvature bounded from below. This generalisation comes from the classical Toponogov's comparison theorem in Riemannian geometry (see [29, Section 6.4] and [40, Theorem 2.2]).

More precisely, the n-dimensional *model space* with constant sectional curvature κ is given by

$$\mathbb{M}^n_\kappa = \begin{cases} \mathbb{S}^n_\kappa, & \text{if } \kappa > 0,\\ \mathbb{R}^n, & \text{if } \kappa = 0,\\ \mathbb{H}^n_\kappa, & \text{if } \kappa < 0,\end{cases}$$

where $\mathbb{S}^n_\kappa$ and $\mathbb{H}^n_\kappa$ are the sphere and the hyperbolic space with their canonical metrics re-scaled by $1/\sqrt{|\kappa|}$.

Definition 2.2.1. A *geodesic triangle* $\triangle pqr$ in X consists of three points $p, q, r \in X$ and three minimising geodesics $[pq]$, $[qr]$, $[rp]$ between those points. A *comparison triangle* in $\mathbb{M}^2_\kappa$ for $\triangle pqr$ is a geodesic triangle $\widetilde{\triangle}_\kappa pqr = \triangle \widetilde{p}\widetilde{q}\widetilde{r}$ in $\mathbb{M}^2_\kappa$ such that

$$d(\widetilde{p}, \widetilde{q}) = d(p, q),\ \ d(\widetilde{q}, \widetilde{r}) = d(q, r),\ \ d(\widetilde{r}, \widetilde{p}) = d(r, p).$$

Definition 2.2.2. We say that X is an *Alexandrov space with curvature bounded below by* κ, and denote it by $\mathrm{curv}(X) \geq \kappa$, if X is complete, geodesic and satisfies the following condition (see Figure 2.1):

(T_κ) For any geodesic triangle $\triangle pqr$, any comparison triangle $\widetilde{\triangle}_\kappa pqr$ in $\mathbb{M}^2_\kappa$ and any point $x \in [qr]$, the corresponding point $\widetilde{x} \in [\widetilde{q}\widetilde{r}]$ such that $d(\widetilde{q}, \widetilde{x}) = d(q, x)$

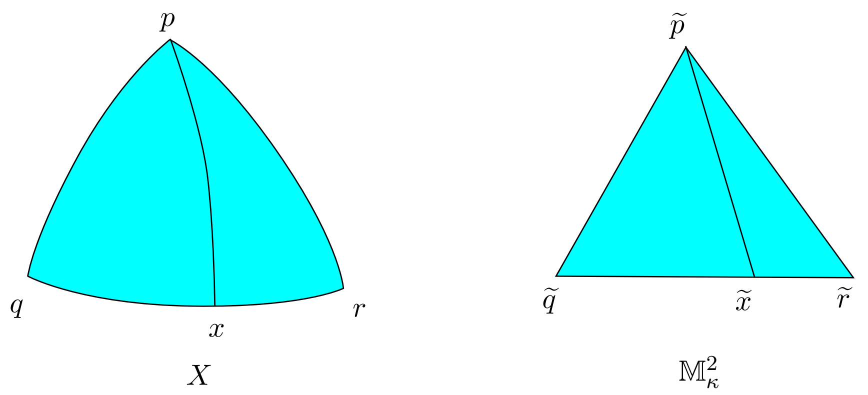

Figure 2.1: $\triangle\widetilde{p}\widetilde{q}\widetilde{r}$ is a comparison triangle for $\triangle pqr$ in $\mathbb{M}^2_\kappa$. Condition (T_κ) means that $d(p,x) \geq d(\widetilde{p},\widetilde{x})$, for any $x \in [qr]$ and the corresponding point $\widetilde{x} \in [\widetilde{q}\widetilde{r}]$ such that $d(q,x) = d(\widetilde{q},\widetilde{x})$.

satisfies

$$d(p,x) \geq d(\widetilde{p},\widetilde{x}).$$

Condition (T_κ) is equivalent to the following:

(A_κ) For any $p \in X$ and any $\xi^1, \xi^2 \in \mathrm{Geo}(X)$ such that $\xi^1_0 = \xi^2_0 = p$, the function $(s,t) \mapsto \widetilde{\measuredangle}_\kappa \xi^1_s p \xi^2_t$ is non-increasing in both s and t, where $\widetilde{\measuredangle}_\kappa \xi^1_s p \xi^2_t$ denotes the angle at $\widetilde{p}$ in the comparison triangle $\widetilde{\triangle}_\kappa \xi^1_s p \xi^2_t$.

Remark 2.2.3. Observe that condition (T_0) can be formulated as follows: for any geodesic triangle $p,q,r \in X$, any geodesic $\xi \in \mathrm{Geo}(X)$ with $\xi_0 = q$ and $\xi_1 = r$,

$$d(p,\xi_t)^2 \geq (1-t)d(p,q)^2 + td(p,r)^2 - (1-t)td(q,r)^2 \tag{2.2}$$

holds for any $t \in [0,1]$. This is due to the right hand side of the inequality above being the square of $|\widetilde{p} - \widetilde{\xi}_t|$ in the comparison triangle $\widetilde{\triangle}_0 pqr$.

Remark 2.2.4. Condition (A_κ) implies that the *angle* between $\xi^1, \xi^2 \in \mathrm{Geo}(X)$ with $\xi^1_0 = \xi^2_0$, given by

$$\measuredangle(\xi^1,\xi^2) = \lim_{s,t\to 0} \widetilde{\measuredangle}_\kappa \xi^1_s p \xi^2_t$$

is well-defined. Geodesics that make an angle zero determine an equivalence class called *geodesic direction*. The set of geodesic directions at a point $p \in X$ is denoted by Σ'_p. When equipped with the angle metric $\measuredangle$, the set Σ'_p is a metric space. The completion of $\left(\Sigma'_p, \measuredangle\right)$ is called the *space of directions of* X *at* p, and is denoted by Σ_p. Note that in a closed Riemannian manifold the space of directions at any point is isometric to the unit sphere in the tangent space to the manifold at the given point.

2.3 Gromov–Hausdorff convergence

Let us also recall the definition of the Hausdorff distance between subsets of a metric space.

Definition 2.3.1. For subsets A and B of a metric space (X, d), the *Hausdorff distance* of A and B is defined as

$$d^d_{\mathsf{H}}(A, B) = \inf \{\varepsilon > 0 : A \subset B_\varepsilon(B) \text{ and } B \subset B_\varepsilon(A)\}.$$

Here, we use the convention that the infimum of an empty set is ∞.

Definition 2.3.2. Let (X, d_X) and (Y, d_Y) be metric spaces. Then the Gromov–Hausdorff distance between (X, d_X) and (Y, d_Y) is defined as

$$d_{\mathsf{GH}}(X, Y) = \inf \left\{ d^d_{\mathsf{H}}(X, Y) \right\}$$

where d runs over all admissible metrics on $X \sqcup Y$, i.e. metrics such that

$$d|_{X\times X} = d_X,$$
$$d|_{Y\times Y} = d_Y.$$

The previous definition allows for spaces with infinite Gromov–Hausdorff distance. However, when restricted to compact spaces, d_{GH} defines a metric. When dealing with non-compact spaces, though, it is customary to use the pointed Gromov–Hausdorff convergence (see, for example, [12, Chapter 8] or [32]).

Definition 2.3.3. A sequence $\{(X_i, a_i)\}_{i\in\mathbb{N}} \subset \mathsf{Met}_*$ is *pointed Gromov–Hausdorff convergent* to $(X, a) \in \mathsf{Met}_*$ if there exist sequences $\{\varepsilon_i\}_{i\in\mathbb{N}}$ and $\{R_i\}_{i\in\mathbb{N}}$ of positive numbers with $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$, and maps $\phi_i \colon \overline{B}_{R_i}(a_i) \to X$ satisfying the following:

1. $|d_{X_i}(x, y) - d_X(\phi_i(x), \phi_i(y)| \leq \varepsilon_i$ for any $x, y \in \overline{B}_{R_i}(a_i)$;
2. $d_{\mathsf{H}}^{d_X}(\phi_i(a_i), a) \leq \varepsilon_i$;
3. $\overline{B}_{R_i}(a) \subset \overline{B}_{\varepsilon_i}(\phi_i(\overline{B}_{R_i}(a_i)))$.

We denote the pointed Gromov–Hausdorff convergence by

$$(X_i, a_i) \xrightarrow{\mathsf{GH}} (X, a).$$

2.4 Optimal transport and measure theory

Let us now recall basic notions of optimal transport that we will need in Chapter 4. Let X be a metric space, $\mathcal{B}(X)$ be the set of Borel measures on X, and $\mathcal{P}(X)$ be the set of Borel probability measures on X, i.e. the set of measures $\mu \in \mathcal{B}(X)$ such that $\mu(X) = 1$.

Definition 2.4.1. For any Borel measurable map $T \colon X \to Y$ between metric spaces, the *push-forward map* $T_\# \colon \mathcal{P}(X) \to \mathcal{P}(Y)$ is given by

$$T_\#\mu(E) = \mu(T^{-1}(E))$$

for any Borel set $E \subset Y$.

Definition 2.4.2. The *support* of a Borel measure μ, denoted by $\operatorname{supp}(\mu)$, is the smallest closed set $E \subset X$ such that $\mu(X \setminus E) = 0$. We also say that μ is *concentrated on* a Borel measurable set $E \subset X$ if $\mu(X \setminus E) = 0$. In particular, $\operatorname{supp}(\mu)$ can also be defined as the smallest closed set where μ is concentrated.

Let us recall the following Borel measurable selection principle (see, for example, [3, Theorem 1]).

Theorem 2.4.3. *Let X and Y be complete and separable metric spaces, and E a closed, σ-compact (i.e. E can be covered with countably many compact sets) subset of $X \times Y$. If $\pi^1 \colon X \times Y \to X$ is the projection onto the first factor, then $\pi^1(E)$ is a Borel set in X and there exists a Borel measurable map $\phi \colon \pi^1(E) \to Y$ whose graph is contained in E.*

Definition 2.4.4. Given $\mu, \nu \in \mathcal{P}(X)$, we say that $\gamma \in \mathcal{P}(X \times X)$ is a *transport plan* between μ and ν if $\pi^1_{\#}\gamma = \mu$ and $\pi^2_{\#}\gamma = \nu$, where $\pi^1,\ \pi^2 \colon X \times X \to X$ are the coordinate maps. The set of transport plans between μ and ν is denoted by $\mathrm{Adm}(\mu, \nu)$.

Definition 2.4.5. Fix $p \in [1, \infty)$ and $\mu, \nu \in \mathcal{P}(X)$. The *optimal transport problem* for $\mu, \nu \in \mathcal{P}(X)$ with cost function $c(x, y) = d(x, y)^p$ is the following:

$$\text{To minimise the total cost } \int d(x,y)^p \, d\gamma(x,y) \text{ over } \gamma \in \mathrm{Adm}(\mu,\nu). \qquad \text{(OT)}$$

Any solution γ for (OT) is an *optimal plan* between μ and ν. The set of optimal plans between μ and ν is denoted by $\mathrm{Opt}(\mu, \nu)$.

Under very general assumptions on the space X, and the probability measures μ and ν, problem (OT) can be solved (see, for example, [54, Theorem 4.1] for a more general result).

Theorem 2.4.6. *Let X be a complete and separable metric space, and $\mu, \nu \in \mathcal{P}(X)$ such that*

$$\int_X d(x, x_0)^p \, d\mu(x), \int_X d(y, x_0)^p \, d\nu(y) < \infty \tag{2.3}$$

for some (and therefore any) $x_0 \in X$. Then $\mathrm{Opt}(\mu, \nu) \neq \varnothing$. Moreover, in that case,

$$\mathrm{W}_p(\mu,\nu) = \min_{\gamma \in \mathrm{Adm}(\mu,\nu)} \left(\int d(x,y)^p \, d\gamma(x,y) \right)^{1/p}$$

defines a metric on $\mathcal{P}_p(X)$, the set of measures in $\mathcal{P}(X)$ satisfying (2.3).

The following result, commonly known in the optimal transport jargon as the Gluing Lemma, plays a role in the proof of Theorem 2.4.6 (see, for example, [54, p. 11] or [2, Lemma 2.1] for details).

Theorem 2.4.7. *Let $\mu^1, \mu^2, \mu^3 \in \mathcal{P}(X)$, and consider $\gamma^{12} \in \mathrm{Adm}(\mu^1, \mu^2)$, and $\gamma^{23} \in \mathrm{Adm}(\mu^2, \mu^3)$. Then there exists $\gamma^{123} \in \mathcal{P}(X \times X \times X)$ such that*

$$\pi^{12}_{\#}(\gamma^{123}) = \gamma^{12},$$
$$\pi^{23}_{\#}(\gamma^{123}) = \gamma^{23}.$$

More generally, if $\Gamma^1 \in \mathcal{P}(\mathcal{X}^1)$, $\Gamma^2 \in \mathcal{P}(\mathcal{X}^2)$, and $F^i\colon \mathcal{X}^i \to \mathcal{X}$, $i = 1, 2$, are measurable maps such that $F^1_{\#}\Gamma^1 = F^2_{\#}\Gamma^2$, then there exists

$$\hat{\Gamma} \in \mathcal{P}(\{(x_1, x_2) \in \mathcal{X}^1 \times \mathcal{X}^2 : F^1(x_1) = F^2(x_2)\})$$

such that $\pi^i_{\#}\hat{\Gamma} = \Gamma^i$, $i = 1, 2$.

It is also possible to characterise optimal plans in terms of cyclical monotonicity and Kantorovich potentials.

Definition 2.4.8. Let $c\colon X \times X \to \mathbb{R}$ be given by $c(x, y) = d(x, y)^p$ for a fixed $p \in [1, \infty)$. We say that a set $\Gamma \subset X \times X$ is *c-cyclically monotone* if, for any $n \in \mathbb{N}$, any $\{(x_i, y_i)\}_{i=1}^n \subset \Gamma$ and any permutation σ of $\{1, \dots, n\}$,

$$\sum_{i=1}^{n} d(x_i, y_i)^p \le \sum_{i=1}^{n} d(x_i, y_{\sigma(i)})^p$$

holds. The *c-transform* of a function $\phi\colon X \to \mathbb{R} \cup \{-\infty\}$ is given by

$$\phi^c(y) = \inf_{x \in X} \{d(x, y)^p - \phi(x)\}.$$

We say that a function $\phi\colon X \to \mathbb{R} \cup \{-\infty\}$ is *c-concave* if there exists a function $\psi\colon X \to \mathbb{R} \cup \{-\infty\}$ such that $\phi(x) = \psi^c(x)$.

The *c-superdifferential* of a c-concave function ϕ is the set

$$\partial^c_+ \phi = \{(x, y) \in X \times X : c(x, y) = \phi(x) + \phi^c(y)\}.$$

We have the following characterisation of optimal plans, which is a particular case of [54, Theorem 5.10].

Theorem 2.4.9. *Let X be a complete and separable metric space, $\mu, \nu \in \mathcal{P}_p(X)$, and $\gamma \in \mathrm{Adm}(\mu, \nu)$. Then the following conditions are equivalent:*

1. $\gamma \in \mathrm{Opt}(\mu, \nu)$*;*
2. $\mathrm{supp}(\gamma)$ *is c-cyclically monotone;*
3. *There is a c-concave function ϕ (known as a* Kantorovich potential *of γ) with* $\max\{0, \phi\} \in L^1(\mu)$ *and* $\mathrm{supp}(\gamma) \subset \partial_c^+ \phi$*.*

We will also need to deal with different kinds of convergence of measures.

Definition 2.4.10. A sequence $\{\mu\}_{n \in \mathbb{N}}$ of bounded Borel measures on X is *weakly convergent* to μ, and we write $\mu_n \xrightarrow{w} \mu$, if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu \quad \text{for any } f \in \mathcal{C}_b(X),$$

where $\mathcal{C}_b(X)$ is the set of continuous and bounded functions on X.

The following result is a particular case of the classical Prokhorov's theorem (see [37, Theorem 4.2]).

Theorem 2.4.11. *Let $\mathcal{F}$ be a set of bounded Borel measures on a proper metric space X. Then $\mathcal{F}$ is weakly relatively compact (i.e. every sequence in $\mathcal{F}$ has a weakly convergent subsequence) if and only if the following conditions hold:*

1. $\mathcal{F}$ *has* uniformly bounded total variation*, i.e.* $\sup\{\mu(X) : \mu \in \mathcal{F}\} < \infty$*,*
2. $\mathcal{F}$ *is* tight*, i.e. for any $\varepsilon > 0$ there exist a compact set $K \subset X$ such that* $\sup\{\mu(X \setminus K) : \mu \in \mathcal{F}\} < \varepsilon$*.*

An even weaker notion of convergence is that of vague convergence of Radon measures.

Definition 2.4.12. A *Radon measure* μ on X is a Borel measure that is both *finite on compact sets* (i.e. $\mu(K) < \infty$ for any compact set $K \subset X$) and *inner regular* (i.e.

for any Borel measurable set $E \subset X$, $\mu(E)$ can be approximated from below by the measures of compact sets $K \subset E$). We denote by $\mathcal{M}(X)$ the set of Radon measures on X.

Remark 2.4.13. If X is a separable and locally compact metric space, any Borel measure on X that is finite on compact sets is also inner regular (see, for example, [28, Theorem 7.8]). Throughout Chapter 4, where we will focus on proper metric spaces (which, in particular, are separable and locally compact), we will therefore only need to verify finiteness on compact sets in order to prove that a Borel measure is Radon.

Definition 2.4.14. A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of Radon measures on X is *vaguely convergent* to μ, and we write $\mu_n \xrightarrow{v} \mu$, if

$$\int_X f \, d\mu_n \to \int_X f \, d\mu \quad \text{for any } f \in \mathcal{C}_c(X), \tag{2.4}$$

where $\mathcal{C}_c(X)$ is the set of compactly supported continuous functions on X.

Lemmas 2.4.15 and 2.4.16 below are adaptations of known results about the vague topology in the setting of separable, locally compact metric spaces (see [36, Section 15.7] for details). In particular, these results are applicable to proper metric spaces.

Lemma 2.4.15. *Let X be a separable, locally compact metric space, and let $\mathcal{F} \subset \mathcal{M}(X)$. Then $\mathcal{F}$ is vaguely relatively compact if and only if*

$$\sup\{\gamma(K) : \gamma \in \mathcal{F}\} < \infty$$

for any compact $K \subset X$.

Lemma 2.4.16. *Let X be a separable, locally compact metric space, and let $\{\gamma_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(X)$. Then the following are equivalent:*

1. *$\gamma_n \xrightarrow{v} \gamma$ for some $\gamma \in \mathcal{M}(X)$.*

2. *For any bounded open $U \subset X$ and any bounded closed $F \subset X$,*

$$\gamma(U) \leq \liminf_{n\to\infty} \gamma_n(U)$$

and

$$\gamma(F) \geq \limsup_{n\to\infty} \gamma_n(F).$$

Chapter 3

Generalised spaces of persistence diagrams

3.1 Introduction

In this chapter we study the geometry of spaces of generalised persistence diagrams, based on the articles [17, 18]. These spaces provide a general construction over metric pairs that includes the spaces $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$ as defined in the Introduction (Definition 1.0.1). Related spaces have been studied independently in [9, 10].

In section 3.2 we define the generalised spaces of persistence diagrams $\mathcal{D}_p(X, A)$ and the L^p-Wasserstein metrics on them; we also prove that the maps $(X, A) \mapsto \mathcal{D}_p(X, A)$ are functorial. In section 3.3 we prove the existence of optimal bijections between generalised persistence diagrams in the setting of proper metric spaces. In sections 3.4, 3.5 and 3.6 we discuss completeness, separability, and geodesics of the space $\mathcal{D}_p(X, A)$, as well as the structure of Alexandrov space of $\mathcal{D}_2(X, A)$, proving Theorems A and B along the way. Finally, in section 3.7, we discuss the continuity of the maps $(X, A) \mapsto (\mathcal{D}_p(X, A), \sigma_A)$ with respect to the Gromov–Hausdorff convergence, proving Theorem C.

3.2 Spaces of persistence diagrams

Let (X, d) be a metric space and fix $p \in [1, \infty)$. Let $\widetilde{\mathcal{D}}(X)$ be the set of finite or countably infinite multisets of points in X and let $\widetilde{d}_p \colon \widetilde{\mathcal{D}}(X) \times \widetilde{\mathcal{D}}(X) \to [0, \infty]$ be given by

$$\widetilde{d}_p(\widetilde{\sigma}, \widetilde{\tau})^p = \inf_{\phi\colon \widetilde{\sigma} \to \widetilde{\tau}} \sum_{x \in \widetilde{\sigma}} d(x, \phi(x))^p, \tag{3.1}$$

where ϕ ranges over all bijections between $\widetilde{\sigma}, \widetilde{\tau} \in \widetilde{\mathcal{D}}(X)$. Here, by convention, the infimum of an empty set is infinite. Therefore, we have $\widetilde{d}_p(\widetilde{\sigma}, \widetilde{\tau}) = \infty$ whenever $\widetilde{\sigma}$ and $\widetilde{\tau}$ do not have the same cardinality.

Lemma 3.2.1. *The function $\widetilde{d}_p$ is an extended pseudometric on $\widetilde{\mathcal{D}}(X)$.*

Proof. It is clear that $\widetilde{d}_p$ is non-negative and symmetric. The triangle inequality may be proved as follows: if $\widetilde{\rho}, \widetilde{\sigma}, \widetilde{\tau} \in \widetilde{\mathcal{D}}(X)$ have the same cardinality and $\phi\colon \widetilde{\rho} \to \widetilde{\sigma}$ and $\psi\colon \widetilde{\sigma} \to \widetilde{\tau}$ are bijections, then $\psi \circ \phi\colon \widetilde{\rho} \to \widetilde{\tau}$ is also a bijection, therefore

$$\begin{aligned}
\widetilde{d}_p(\widetilde{\rho}, \widetilde{\tau}) &\le \left(\sum_{x \in \widetilde{\rho}} d(x, \psi \circ \phi(x))^p \right)^{1/p} \\
&\le \left(\sum_{x \in \widetilde{\rho}} (d(x, \phi(x)) + d(\phi(x), \psi \circ \phi(x)))^p \right)^{1/p} \\
&\le \left(\sum_{x \in \widetilde{\rho}} d(x, \phi(x))^p \right)^{1/p} + \left(\sum_{x \in \widetilde{\rho}} d(\phi(x), \psi \circ \phi(x))^p \right)^{1/p} \\
&= \left(\sum_{x \in \widetilde{\rho}} d(x, \phi(x))^p \right)^{1/p} + \left(\sum_{y \in \widetilde{\sigma}} d(y, \psi(y))^p \right)^{1/p}.
\end{aligned}$$

Taking the infimum over bijections ϕ and ψ we get the claim. If the cardinalities of $\widetilde{\rho}, \widetilde{\sigma}, \widetilde{\tau}$ are not the same, the inequality is trivial, since both sides or just the right-hand side would be infinite. ◻

Given two multisets $\widetilde{\sigma}$ and $\widetilde{\tau}$, we define their *sum* $\widetilde{\sigma} + \widetilde{\tau}$ to be their disjoint union. We can make $\widetilde{\mathcal{D}}(X)$ into a commutative monoid with monoid operation given by taking sums of multisets, and with identity the empty multiset. It is easy to check that $\widetilde{d}_p$ satisfies

$$\widetilde{d}_p(\widetilde{\sigma}, \widetilde{\tau}) \geq \widetilde{d}_p(\widetilde{\rho} + \widetilde{\sigma}, \widetilde{\rho} + \widetilde{\tau}) \tag{3.2}$$

for all $\widetilde{\sigma}, \widetilde{\tau}, \widetilde{\rho} \in \widetilde{\mathcal{D}}(X)$. Indeed, any bijection $\phi \colon \widetilde{\sigma} \to \widetilde{\tau}$ can be extended into a bijection $\phi' \colon \widetilde{\sigma} + \widetilde{\rho} \to \widetilde{\tau} + \widetilde{\rho}$ by defining $\phi'(x) = \phi(x)$ whenever $x \in \widetilde{\sigma}$ and $\phi'(x) = x$ whenever $x \in \widetilde{\rho}$, and clearly

$$\sum_{x \in \widetilde{\sigma}} d(x, \phi(x))^p = \sum_{x \in \widetilde{\sigma} + \widetilde{\rho}} d(x, \phi'(x))^p \geq d_p(\widetilde{\sigma} + \widetilde{\rho}, \widetilde{\tau} + \widetilde{\rho})^p.$$

From now on, let $(X, A) \in \mathsf{Met_{Pair}}$. Given $\widetilde{\sigma}, \widetilde{\tau} \in \widetilde{\mathcal{D}}(X)$, we write $\widetilde{\sigma} \sim_A \widetilde{\tau}$ if there exist $\widetilde{\alpha}, \widetilde{\beta} \in \widetilde{\mathcal{D}}(A)$ such that $\widetilde{\sigma} + \widetilde{\alpha} = \widetilde{\tau} + \widetilde{\beta}$. It is easy to verify that $\sim_A$ defines an equivalence relation on $\widetilde{\mathcal{D}}(X)$ such that, if $\widetilde{\alpha}_1 \sim_A \widetilde{\alpha}_2$ and $\widetilde{\beta}_1 \sim_A \widetilde{\beta}_2$, then $\widetilde{\alpha}_1 + \widetilde{\beta}_1 \sim_A \widetilde{\alpha}_2 + \widetilde{\beta}_2$, i.e. $\sim_A$ is a congruence relation on $\widetilde{\mathcal{D}}(X)$ (see, for example, [33, p. 27] for further details on congruence relations). We denote by $\mathcal{D}(X, A)$ the quotient set $\widetilde{\mathcal{D}}(X)/{\sim_A}$. Given $\widetilde{\sigma} \in \widetilde{\mathcal{D}}(X)$, we write σ for the equivalence class of $\widetilde{\sigma}$ in $\mathcal{D}(X, A)$. Note that $\widetilde{\sigma} \sim_A \widetilde{\tau}$ if and only if $\widetilde{\sigma}|_{X \setminus A} = \widetilde{\tau}|_{X \setminus A}$, that is, $\widetilde{\sigma}$ and $\widetilde{\tau}$ share the same points with the same multiplicities outside A. The monoid operation on $\widetilde{\mathcal{D}}(X)$ induces a monoid operation on $\mathcal{D}(X, A)$ by defining $\sigma + \tau$ as the congruence class corresponding to $\widetilde{\sigma} + \widetilde{\tau}$. Moreover, we denote by σ_A the equivalence class of the empty multiset.

The function $\widetilde{d}_p$ induces a function $d_p \colon \mathcal{D}(X, A) \times \mathcal{D}(X, A) \to [0, \infty]$ defined by

$$d_p(\sigma, \tau) = \inf_{\widetilde{\alpha}, \widetilde{\beta} \in \widetilde{\mathcal{D}}(A)} \widetilde{d}_p(\widetilde{\sigma} + \widetilde{\alpha}, \widetilde{\tau} + \widetilde{\beta}). \tag{3.3}$$

Note that d_p satisfies the inequality

$$d_p(\sigma, \tau) \geq d_p(\rho + \sigma, \rho + \tau) \tag{3.4}$$

for all $\sigma, \tau, \rho \in \mathcal{D}(X, A)$, analogous to (3.2).

Let $\mathcal{D}_p(X, A)$, is the set of all $\sigma \in \mathcal{D}(X, A)$ such that $d_p(\sigma, \sigma_A) < \infty$.

Lemma 3.2.2. *If $\widetilde{\sigma} \in \widetilde{\mathcal{D}}(X)$ is a finite multiset, then $\sigma \in \mathcal{D}_p(X, A)$.*

Proof. Let $\widetilde{\sigma} \in \widetilde{\mathcal{D}}(X)$ be a multiset of cardinality $k < \infty$. Since $A \subset X$ is non-empty, we can pick an element $a \in A$, and consider the multiset $k\,\{\{a\}\} = \{\{a, \ldots, a\}\} \in \widetilde{\mathcal{D}}(A)$ of cardinality k. Since the finite multisets $\widetilde{\sigma}$ and $k\,\{\{a\}\}$ have the same cardinality, it follows that $d_p(\sigma, \sigma_A) \leq \widetilde{d}_p(\widetilde{\sigma}, k\,\{\{a\}\}) < \infty$. □

Theorem 3.2.3. *The function d_p is an extended metric on $\mathcal{D}(X, A)$ and a metric on $\mathcal{D}_p(X, A)$.*

Proof. We show first that d_p is an extended pseudometric on $\mathcal{D}(X, A)$. Indeed, it is clear that, for all $p \in [1, \infty)$, the function d_p is symmetric, non-negative, and $d_p(\sigma, \sigma) = 0$ for all $\sigma \in \mathcal{D}(X, A)$. For the triangle inequality, let $\widetilde{\rho}, \widetilde{\sigma}, \widetilde{\tau} \in \widetilde{\mathcal{D}}(X)$ and $\varepsilon > 0$. By the definition of d_p, there exist $\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}, \widetilde{\delta} \in \widetilde{\mathcal{D}}(A)$ such that

$$\widetilde{d}_p(\widetilde{\rho} + \widetilde{\alpha}, \widetilde{\sigma} + \widetilde{\beta}) \leq d_p(\rho, \sigma) + \varepsilon,$$
$$\widetilde{d}_p(\widetilde{\sigma} + \widetilde{\gamma}, \widetilde{\tau} + \widetilde{\delta}) \leq d_p(\sigma, \tau) + \varepsilon.$$

Using the commutativity of $\widetilde{\mathcal{D}}(X)$, the contractivity of $\widetilde{d}_p$, and the triangle inequality for $\widetilde{d}_p$, we get

$$\begin{aligned} d_p(\rho, \tau) &\leq \widetilde{d}_p(\widetilde{\rho} + \widetilde{\alpha} + \widetilde{\gamma}, \widetilde{\tau} + \widetilde{\beta} + \widetilde{\delta}) \\ &\leq \widetilde{d}_p(\widetilde{\rho} + \widetilde{\alpha} + \widetilde{\gamma}, \widetilde{\sigma} + \widetilde{\beta} + \widetilde{\gamma}) + \widetilde{d}_p(\widetilde{\sigma} + \widetilde{\beta} + \widetilde{\gamma}, \widetilde{\tau} + \widetilde{\beta} + \widetilde{\delta}) \\ &\leq \widetilde{d}_p(\widetilde{\rho} + \widetilde{\alpha}, \widetilde{\sigma} + \widetilde{\beta}) + \widetilde{d}_p(\widetilde{\sigma} + \widetilde{\gamma}, \widetilde{\tau} + \widetilde{\delta}) \\ &\leq d_p(\rho, \sigma) + d_p(\sigma, \tau) + 2\varepsilon. \end{aligned}$$

Our choice of $\varepsilon > 0$ was arbitrary, implying that $d_p(\rho, \tau) \leq d_p(\rho, \sigma) + d_p(\sigma, \tau)$, as required. Hence, d_p is an extended pseudometric on $\mathcal{D}(X, A)$.

Moreover, by the triangle inequality, d_p is a pseudometric on $\mathcal{D}_p(X, A)$. Indeed, if $\sigma, \tau \in \mathcal{D}_p(X, A)$, then $d_p(\sigma, \tau) \leq d_p(\sigma, \sigma_A) + d_p(\tau, \sigma_A) < \infty$.

Finally, we prove that $d_p(\sigma, \tau) = 0$ if and only if $\sigma = \tau$. For this, let $\widetilde{\sigma}, \widetilde{\tau} \in \widetilde{\mathcal{D}}(X)$ be multisets such that $\sigma \neq \tau$. It then follows that there exists a point $u \in X \setminus A$ which appears in $\widetilde{\sigma}$ and $\widetilde{\tau}$ with different multiplicities (which includes the case when it has multiplicity 0 in one of the diagrams and positive multiplicity in the other).

Without loss of generality, suppose that u appears with higher multiplicity in $\widetilde{\sigma}$. Now let $\varepsilon_1 = \inf\{d(u,v) : v \in \widetilde{\tau}, v \neq u\}$. Observe that $\varepsilon_1 > 0$ since, otherwise, there would be a sequence of points in $\widetilde{\tau}$ converging to u in X, which in turn would imply that $d_p(\tau, \sigma_A) = \infty$. Let $\varepsilon_2 > 0$ be such that $d(u,a) \geq \varepsilon_2$ for all $a \in A$, which exists since $u \in X \setminus A$ and $X \setminus A$ is open in X. By setting $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\} > 0$, it follows that for any $\widetilde{\alpha}, \widetilde{\beta} \in \widetilde{\mathcal{D}}(A)$, if $\phi\colon \widetilde{\sigma} + \widetilde{\alpha} \to \widetilde{\tau} + \widetilde{\beta}$ is a bijection, then ϕ must map some copy of $u \in \widetilde{\sigma}$ to a point $v \in \widetilde{\tau} + \widetilde{\beta}$ with $d(u,v) \geq \varepsilon$, implying that $\widetilde{d}_p(\widetilde{\sigma} + \widetilde{\alpha}, \widetilde{\tau} + \widetilde{\beta}) \geq \varepsilon$. By taking the infimum over all $\widetilde{\alpha}, \widetilde{\beta} \in \widetilde{\mathcal{D}}(A)$, it follows that $d_p(\sigma, \tau) \geq \varepsilon > 0$, as required. This shows that d_p is an extended metric on $\mathcal{D}(X,A)$ and a metric on $\mathcal{D}_p(X,A)$. □

Definition 3.2.4. The L^p*-space of persistence diagrams* on the metric pair (X, A) is the set $\mathcal{D}_p(X,A)$ endowed with the L^p*-Wasserstein metric* d_p.

From now on, and for the sake of simplicity, we will treat elements in $\mathcal{D}_p(X,A)$ as multisets, with the understanding that whenever we do so we are actually dealing with representatives of such elements in $\widetilde{\mathcal{D}}(X)$. Thus, for instance, we can talk about bijections $\phi\colon \sigma \to \tau$ for $\sigma, \tau \in \mathcal{D}_p(X,A)$, meaning there are representatives $\widetilde{\sigma}$ and $\widetilde{\tau}$ and a bijection $\widetilde{\phi}\colon \widetilde{\sigma} \to \widetilde{\tau}$.

Now we observe that, for any $p \in [1, \infty)$, $\mathcal{D}_p$ defines a functor from the category $\mathsf{Met}_{\mathsf{Pair}}$ to Met_* (see Definition 2.1.1).

Indeed, for any relative map $f\colon (X,A) \to (Y,B)$ between metric pairs (i.e. $f(A) \subset B$), we define a pointed map $f_*\colon \mathcal{D}(X,A) \to \mathcal{D}(Y,B)$ as follows: for any given $\sigma \in \mathcal{D}(X,A)$, let

$$f_*(\sigma) = \{\{f(x) : x \in \sigma\}\}\,. \tag{3.5}$$

Proposition 3.2.5. *For any* $p \in [1, \infty)$*, the maps*

$$\begin{cases} (X,A) \mapsto (\mathcal{D}_p(X,A), \sigma_A) \\ (f\colon (X,A) \to (Y,B)) \mapsto \left(f_*|_{\mathcal{D}_p(X,A)}\colon \mathcal{D}_p(X,A) \to \mathcal{D}_p(Y,B)\right) \end{cases} \tag{3.6}$$

define a functor from the category $\mathsf{Met}_{\mathsf{Pair}}$ *of metric pairs equipped with relative Lipschitz maps to the category* Met_* *of pointed metric spaces with pointed Lipschitz maps.*

Proof. First, we need to prove that for any morphism f in $\mathsf{Met}_{\mathsf{Pair}}$, the map f_* restricts to a morphism in Met_*. In other words, if $f\colon (X,A) \to (Y,B)$ is a relative map such that

$$d_Y(f(x), f(y)) \leq C d_X(x,y)$$

holds for all $x, y \in X$, and for some constant $C > 0$, we have to prove that f_* given by (3.5) restricts to a pointed Lipschitz map from $\mathcal{D}_p(X,A)$ to $\mathcal{D}_p(Y,B)$. Since the fact that f_* is a pointed map is clear, we will focus on proving that it maps $\mathcal{D}_p(X,A)$ into $\mathcal{D}_p(Y,B)$, and that it is Lipschitz.

Indeed, for any $\sigma \in \mathcal{D}_p(X,A)$, we have

$$d_p(f_*(\sigma), \sigma_B)^p = \sum_{x\in\sigma} d_Y(f(x), B)^p \leq \sum_{x\in\sigma} d_Y(f(x), f(a_x))^p \leq C^p \sum_{x\in\sigma} d_X(x, a_x)^p$$

for any choice $\{a_x\}_{x\in\sigma} \subset A$, due to f being a relative map. Since the choice $\{a_x\}_{x\in\sigma}$ is arbitrary,

$$d_p(f_*(\sigma), \sigma_B)^p \leq C^p \sum_{x\in\sigma} d_X(x, A)^p = C^p d_p(\sigma, \sigma_A) < \infty.$$

Therefore, the image of $f_*|_{\mathcal{D}_p(X,A)}$ is contained in $\mathcal{D}_p(Y,B)$.

Now consider two diagrams $\sigma, \tau \in \mathcal{D}_p(X,A)$. Observe that, if $\phi\colon \sigma \to \tau$ is a bijection, then $f_*\phi\colon f_*(\sigma) \to f_*(\tau)$ given by $f_*\phi(y) = f(\phi(x))$, whenever $y = f(x)$ for some $x \in \sigma$, is a bijection too. Therefore

$$\begin{aligned} d_p(f_*(\sigma), f_*(\tau))^p &\leq \sum_{y\in f_*(\sigma)} d(y, f_*\phi(y))^p \\ &= \sum_{x\in\sigma} d(f(x), f(\phi(x)))^p \\ &\leq C^p \sum_{x\in\sigma} d(x, \phi(x))^p. \end{aligned}$$

Since $\phi\colon \sigma \to \tau$ is an arbitrary bijection, we get

$$d_p(f_*(\sigma), f_*(\tau)) \leq C d_p(\sigma, \tau).$$

Thus, $f_*|_{\mathcal{D}_p(X,A)}$ is Lipschitz.

Finally, if $f\colon (X,A) \to (Y,B)$ and $g\colon (Y,B) \to (Z,C)$ are relative Lipschitz maps, then, for any $\sigma \in \mathcal{D}_p(X,A)$, we have

$$(g \circ f)_*(\sigma) = \{\{g \circ f(x) : x \in \sigma\}\} = g_*(\{\{f(x) : x \in \sigma\}\}) = g_* \circ f_*(\sigma),$$

which means that $(g \circ f)_* = g_* \circ f_*$. Moreover, it is clear that $\mathrm{id}_* |_{\mathcal{D}_p(X,A)}$ is the identity map, whenever $\mathrm{id}\colon (X,A) \to (X,A)$ is the identity. □

Remark 3.2.6. Note that $\mathcal{D}_p$ also defines a functor from the category of metric pairs equipped with relative isometries or relative bi-Lipschitz maps. However, Proposition 3.2.5 is more general.

Remark 3.2.7. Proposition 3.2.5 implies that, if (X,A) is a metric pair and $(g,x) \mapsto g \cdot x$ is an action of a group G on (X,A) via relative bi-Lipschitz maps, then we get an action of G on $\mathcal{D}_p(X,A)$ given by

$$g \cdot \sigma = \{\{ga : a \in \sigma\}\}.$$

Observe that the Lipschitz constants of the bi-Lipschitz maps in the group action is preserved by the functor $\mathcal{D}_p$. Hence, if G acts by relative isometries on (X,A), i.e. by isometries $f\colon X \to X$ such that $f(A) \subseteq A$, then the induced action on $(\mathcal{D}_p(X,A), \sigma_A)$ is by pointed isometries, i.e. isometries that fix σ_A.

Remark 3.2.8. We point out that $\mathcal{D}_p$, in fact, defines a functor from $\mathsf{Met}_{\mathsf{Pair}}$ to $\mathsf{CMon}(\mathsf{Met}_*)$, the category of commutative pointed metric monoids (see [8]). This means that, given a map $f\colon (X,A) \to (Y,B)$, the induced map $f_*\colon \mathcal{D}_p(X,A) \to \mathcal{D}_p(Y,B)$ is a monoid morphism. Composing such functor with a forgetful functor, one obtains the maps in (3.6). In this thesis we consider this last composition, since we are mainly interested in the metric geometry of the spaces $\mathcal{D}_p(X,A)$.

Remark 3.2.9. Consider now the quotient metric space X/A, namely, the quotient space induced by the partition $\{\{x\} : x \in X \setminus A\} \sqcup \{A\}$ endowed with the metric given by

$$d([x],[y]) = \min\{d(x,y), d(x,A) + d(y,A)\}$$

for any $x, y \in X$ (cf. [44, Ch. 2, §22] and [12, Definition 3.1.12]). It follows from [10, Remark 4.14 and Lemma 4.24] that $\mathcal{D}_p(X, A)$ and $\mathcal{D}_p(X/A, [A])$ are isometrically isomorphic. We have the following commutative diagram of functors:

$$\begin{array}{ccc} \mathsf{Met}_{\mathsf{Pair}} & \xrightarrow{\mathcal{D}_p} & \mathsf{Met}_* \\ {\scriptstyle \mathcal{Q}}\downarrow & & \downarrow{\scriptstyle \cong} \\ \mathsf{Met}_* & \xrightarrow{\mathcal{D}_p} & \mathsf{Met}_* \end{array} \tag{3.7}$$

given by

$$\begin{array}{ccc} (X, A) & \longmapsto & (\mathcal{D}_p(X, A), \sigma_A) \\ \big\downarrow & & \big\downarrow \\ (X/A, [A]) & \longmapsto & (\mathcal{D}_p(X/A, [A]), \sigma_{[A]}) \end{array} .$$

Observe that the map $\mathcal{D}_p(X, A) \mapsto \mathcal{D}_p(X/A, [A])$ is a natural isomorphism. Therefore, diagram (3.7) show that the functor $\mathcal{D}_p$ factors through the quotient functor $\mathcal{Q}\colon (X, A) \mapsto (X/A, [A])$ and the functor $(X/A, [A]) \mapsto \mathcal{D}_p(X/A, [A])$.

Note that we also have the following commutative diagram of functors:

$$\begin{array}{ccc} \mathsf{Met}_{\mathsf{Pair}} & \longrightarrow & \overline{\mathsf{Met}}_{\mathsf{Pair}} \\ {\scriptstyle \mathcal{Q}}\downarrow & & \downarrow{\scriptstyle \cong} \\ \mathsf{Met}_* & \longrightarrow & \overline{\mathsf{Met}}_* \end{array} \tag{3.8}$$

given by

$$\begin{array}{ccc} (X, A) & \longmapsto & (\widetilde{\mathcal{D}}(X), \widetilde{\mathcal{D}}(A), \widetilde{d}_p) \\ \big\downarrow & & \big\downarrow \\ (X/A, [A]) & \longmapsto & (\mathcal{D}(X/A, [A]), \sigma_{[A]}, d_p) \end{array} .$$

Here the categories $\overline{\mathsf{Met}}_{\mathsf{Pair}}$ and $\overline{\mathsf{Met}}_*$ consist of extended metric pairs and extended pointed metric spaces, respectively.

Remark 3.2.10. By replacing the L^p-norm of $\{d(x, \phi(x))\}_{x \in \widetilde{\sigma}}$ in equation (3.1) with the corresponding L^∞-norm, we get the function $\widetilde{d}_\infty\colon \widetilde{\mathcal{D}}(X) \times \widetilde{\mathcal{D}}(X) \to [0, \infty]$ given by

$$\widetilde{d}_\infty(\widetilde{\sigma}, \widetilde{\tau}) = \inf_{\phi\colon \widetilde{\sigma} \to \widetilde{\tau}} \sup\{d(x, \phi(x)) : x \in \widetilde{\sigma}\}.$$

This in turn induces a function $d_\infty\colon \mathcal{D}(X, A) \times \mathcal{D}(X, A) \to [0, \infty]$ which restricts to a pseudometric on the set

$$\mathcal{D}_\infty(X, A) = \{\sigma \in \mathcal{D}(X, A) : d_\infty(\sigma, \sigma_A) < \infty\}.$$

In this case, $\mathcal{D}_\infty$ defines a functor from $\mathsf{Met}_{\mathsf{Pair}}$ to PMet_*, and we get a commutative diagram of functors

$$\begin{array}{ccc} \mathsf{Met}_{\mathsf{Pair}} & \xrightarrow{\mathcal{D}_\infty} & \mathsf{PMet}_* \\ {\scriptstyle \mathcal{Q}}\downarrow & & \downarrow{\scriptstyle \cong} \\ \mathsf{Met}_* & \xrightarrow{\mathcal{D}_\infty} & \mathsf{PMet}_* \end{array} \tag{3.9}$$

completely analogous to diagram (3.7).

We do not discuss in this thesis neither the details of these constructions, nor the geometric properties of the resulting spaces $\mathcal{D}_\infty(X, A)$, all of which be found in [17, 18]. We do, however, prove that $\mathcal{D}_\infty$ is a continuous map with respect to the Gromov–Hausdorff convergence of metric pairs and pointed pseudometric spaces (see Section 3.7).

3.3 Existence of optimal bijections

In this section, we prove the existence of optimal bijections between persistence diagrams in $\mathcal{D}_p(X, A)$ under the assumption that X is a proper metric space (Theorem 3.3.3). In order to prove this, we need the following two lemmas, which are generalisations of [20, Lemmas 17 and 18], and their proofs are essentially the same with the only difference that, for a general metric pair (X, A) where X is assumed to be proper, points in X always have a closest point in A but such a point is not necessarily unique.

Lemma 3.3.1. *Let* $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$. *Let* $\sigma, \tau \in \mathcal{D}_p(X, A)$ *be diagrams,* $\phi_k \colon \sigma \to \tau$ *be a sequence of bijections such that* $\sum_{x\in\sigma} d(x, \phi_k(x))^p \to d_p(\sigma, \tau)^p$ *as* $k \to \infty$. *Then the following assertions hold:*

1. *If* $x \in \sigma$, $y \in \tau \setminus A$ *are such that* $\lim_{k\to\infty} \phi_k(x) = y$, *then there exists* $k_0 \in \mathbb{N}$ *such that* $\phi_k(x) = y$ *for all* $k \geq k_0$.

2. *If* $x \in \sigma \setminus A$, $y \in A$ *are such that* $\lim_{k\to\infty} \phi_k(x) = y$, *then* $d(x, y) = d(x, A)$.

Proof. 1. Since $p \in [1, \infty)$ and $\tau \in \mathcal{D}_p(X, A)$, there is some $\varepsilon > 0$ such that $B_\varepsilon(y) \cap \tau = \{y\}$. Since $\phi_k(x) \in B_\varepsilon(y) \cap \tau$ for sufficiently large k, the conclusion follows.

2. Assuming $d(x, y) > d(x, A)$, it follows that $d(x, \phi_k(x)) > d(x, A) + 2\varepsilon$ and $d(\phi_k(x), A) < \varepsilon$ for sufficiently large k, where $\varepsilon = (d(x, y) - d(x, A))/3$. This contradicts the fact that $\sum_{x \in \sigma} d(x, \phi_k(x))^p \to d_p(\sigma, \tau)^p$ as $k \to \infty$. ☐

Lemma 3.3.2. *Let $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$ and assume X is a proper metric space. Let $\sigma, \tau \in \mathcal{D}_p(X, A)$, and let $\phi_k \colon \sigma \to \tau$ be a sequence of bijections such that $\sum_{x \in \sigma} d(x, \phi_k(x))^p \to d_p(\sigma, \tau)^p$ as $k \to \infty$. Then there exists a subsequence $\{\phi_{k_l}\}$ and a limiting bijection ϕ_* such that $\phi_{k_l} \to \phi_*$ pointwise as $l \to \infty$ and*

$$\sum_{x \in \sigma} d(x, \phi_*(x))^p = d_p(\sigma, \tau)^p.$$

Proof. Since $d_p(\sigma, \tau) < \infty$, for each point $x \in \sigma \setminus A$ the sequence $\{\phi_k(x)\}_{k \in \mathbb{N}}$ consists of a bounded set of points in X and at most countably many copies of A. In particular, thanks to Lemma 3.3.1 and the fact that X is proper, and using a diagonal argument, we can assume that for each $x \in \sigma \setminus A$, the sequence $\{\phi_k(x)\}_{k \in \mathbb{N}}$ is eventually constant equal to some point $y \in \tau \setminus A$ or it is convergent to some point $y \in A$ such that $d(x, y) = d(x, A)$. In any case, we can define $\phi_* \colon \sigma \setminus A \to \tau$ as

$$\phi_*(x) = \lim_{k \to \infty} \phi_k(x).$$

By mapping enough points in A to all the points in τ that were not matched with points in $\sigma \setminus A$, we get the required bijection $\phi_* \colon \sigma \to \tau$. ☐

Theorem 3.3.3. *Let $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$ and assume X is a proper space. Then for any $\sigma, \sigma' \in \mathcal{D}_p(X, A)$ there exists an optimal bijection $\phi \colon \sigma \to \tau$, i.e. $d_p(\sigma, \tau)^p = \sum_{x \in \sigma} d(x, \phi(x))^p$.*

3.4 Completeness and separability

In this section we prove that $\mathcal{D}_p(X, A)$ preserves the completeness and separability of metric spaces, for any $p \in [1, \infty)$, in the sense of items 1 and 2 in Theorem A. More precisely, we prove the following stronger results.

Theorem 3.4.1. *For any $p \in [1, \infty)$, the space $\mathcal{D}_p(X, A)$ is complete if and only if X/A is complete.*

Theorem 3.4.2. *Let $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$ and $p \in [1, \infty)$. Then $\mathcal{D}_p(X, A)$ is separable if and only if X/A is separable.*

The proofs in this section follow closely the arguments in [41] about the classical spaces of persistence diagrams $\mathcal{D}_p(\mathbb{R}^2_{\geq}, \Delta)$.

Proof of Theorem 3.4.1. Let $p \in [1, \infty)$ and (X, A) be a metric pair. In view of the isometry $\mathcal{D}_p(X, A) \cong \mathcal{D}_p(X/A, [A])$ for any $p \in [1, \infty)$ (see Remark 3.2.9), Theorem 3.4.1 is equivalent to the statement that $\mathcal{D}_p(X, \{a_0\})$ is complete if and only if X is complete, where $a_0 \in X$ is any point.

The "only if" implication of Theorem 3.4.1 follows from Lemma 3.4.3 below.

Lemma 3.4.3. *If $\mathcal{D}_p(X, \{a_0\})$ is complete, then so is X.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in X. Then $\{d(x_n, a_0)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, therefore convergent. If $d(x_n, a_0) \to 0$ as $n \to \infty$ then $x_n \to a_0$. Thus, we may assume that $d(x_n, a_0) \to \delta$ as $n \to \infty$ for some $\delta > 0$.

Now, for each $n \in \mathbb{N}$, let $\sigma_n = \{\{x_n\}\} \in \mathcal{D}_p(X, \{a_0\})$. For each $n, m \in \mathbb{N}$, it is clear that $d_p(\sigma_n, \sigma_m) \leq d(x_n, x_m)$. Thus, $\{\sigma_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{D}_p(X, \{a_0\})$, therefore it converges to some $\sigma \in \mathcal{D}_p(X, \{a_0\})$. Now let $\varepsilon \in (0, \delta/2]$. Then there exists $N \in \mathbb{N}$ such that $d_p(\sigma_n, \sigma) < \varepsilon$ for $n \geq N$, which implies there exists a bijection $\phi_n \colon \sigma_n \to \sigma$ such that $d(x', \phi_n(x')) < \varepsilon \leq \delta/2$ for every $x' \in \sigma_n$. This implies that σ contains a unique point $x \in X \setminus B_{\delta/2}(a_0)$ and that $\phi_n(x_n) = x$, and hence $d(x_n, x) < \varepsilon$, for every $n \geq N$. Therefore, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to $x \in X$. $\square$

For the "if" implication, suppose now that X is complete and let $\{\sigma_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{D}_p(X, A)$; here we do not need the assumption $A = \{a_0\}$, the argument works for any non-empty closed set $A \subset X$.

For any $\alpha > 0$, let $u_\alpha \colon \mathcal{D}_p(X, A) \to \mathcal{D}_p(X, A)$ be the map defined by

$$u_\alpha(\sigma) = \{\{x \in \sigma : d(x, A) \geq \alpha\}\} .$$

We call $u_\alpha(\sigma)$ the *α-upper part* of σ. We define in a similar way the *α-lower part* of σ, by letting $l_\alpha \colon \mathcal{D}_p(X, A) \to \mathcal{D}_p(X, A)$ be given by

$$l_\alpha(\sigma) = \{\{x \in \sigma : d(x, A) < \alpha\}\} .$$

Lemma 3.4.4. *Let $\alpha > 0$. Then there exist $M_\alpha \in \mathbb{Z}_{\geq 0}$ and $\delta_\alpha \in (0, \alpha)$, such that, for all $\delta \in [\delta_\alpha, \alpha)$ the equation $|u_\delta(\sigma_n)| = M_\alpha$ holds for sufficiently large n.*

Proof. For $\delta \in (0, \alpha)$, let

$$M^\delta_{\sup} = \limsup_{n\to\infty} |u_\delta(\sigma_n)|$$

and

$$M^\delta_{\inf} = \liminf_{n\to\infty} |u_\delta(\sigma_n)| .$$

Since $\{\sigma_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, $\{d_p(\sigma_n, \sigma_A)\}_{n\in\mathbb{N}}$ is bounded, which implies that $M^\delta_{\sup} < \infty$. Also, if $\delta_1 > \delta_2$, then $|u_{\delta_1}(\sigma_n)| \leq |u_{\delta_2}(\sigma_n)|$ which means that

$$0 \leq M^{\delta_1}_{\sup} \leq M^{\delta_2}_{\sup} \ \text{ and } \ 0 \leq M^{\delta_1}_{\inf} \leq M^{\delta_2}_{\inf}.$$

Therefore, the limits $M_{\sup} = \lim_{\delta\to\alpha} M^\delta_{\sup}$ and $M_{\inf} = \lim_{\delta\to\alpha} M^\delta_{\inf}$ exist and, moreover, there exists a δ_α such that $M_{\sup} = M^\delta_{\sup}$ and $M_{\inf} = M^\delta_{\inf}$ whenever $\delta_\alpha \leq \delta < \alpha$, since $M^\delta_{\sup}$ and $M^\delta_{\inf}$ are integers for any δ. Now suppose that $M_{\inf} < M_{\sup}$. Fix $\delta \in (\delta_\alpha, \alpha)$ and let $\varepsilon = \delta - \delta_\alpha > 0$. Let $\{\sigma_{n_k}\}_{k\in\mathbb{N}}$ and $\{\sigma_{n_l}\}_{l\in\mathbb{N}}$ be two subsequences of $\{\sigma_n\}_{n\in\mathbb{N}}$ such that $|u_\delta(\sigma_{n_k})| = M_{\sup}$ and $|u_{\delta_\alpha}(\sigma_{n_l})| = M_{\inf}$. Since $\{\sigma_n\}_{n\in\mathbb{N}}$ is Cauchy, there exists $N \in \mathbb{N}$ such that $d_p(\sigma_{n_k}, \sigma_{n_l}) < \varepsilon$ for all

$k, l \geq N$. By assumption, $|u_\delta(\sigma_{n_k})| > |u_{\delta_\alpha}(\sigma_{n_l})|$, which implies that, for any bijection $\phi\colon \sigma_{n_k} \to \sigma_{n_l}$, there is a point $x \in \sigma_{n_k}$ such that $d(x, A) \geq \delta$ and $d(\phi(x), A) < \delta_\alpha$. This means that $d(x, \phi(x)) > \varepsilon$, leading to $d_p(\sigma_{n_k}, \sigma_{n_l}) \geq \varepsilon$, which is a contradiction. We then set $M_\alpha = M_{\sup} = M_{\inf}$. □

For any $\alpha > 0$, let $\sigma_n^\alpha = u_{\delta_\alpha}(\sigma_n)$ and $\sigma_{n,\alpha} = l_{\delta_\alpha}(\sigma_n)$.

Lemma 3.4.5. *The sequence $\{\sigma_n^\alpha\}_{n\in\mathbb{N}} \subset \mathcal{D}_p(X, A)$ is Cauchy, for any $\alpha > 0$.*

Proof. Let δ_α be as in Lemma 3.4.4 and let $\delta \in (\delta_\alpha, \alpha)$. Let $\varepsilon > 0$ and let $\varepsilon_0 = \min\{\varepsilon, (\delta - \delta_\alpha)/2\}$. By Lemma 3.4.4, there is $N \in \mathbb{N}$ such that, for all $n \geq N$, there is no point $x \in \sigma_n$ with $d(x, A) \in [\delta_\alpha, \delta)$, and since $\{\sigma_n\}_{n\in\mathbb{N}}$ is Cauchy, we can further assume that $d_p(\sigma_m, \sigma_n) < \varepsilon_0$ for all $m, n \geq N$. Then there is a bijection $\phi\colon \sigma_m \to \sigma_n$ such that

$$\left(\sum_{x\in\sigma_m} d(x, \phi(x))^p\right)^{\frac{1}{p}} < \varepsilon_0 \leq \frac{\delta - \delta_\alpha}{2},$$

which implies that $\phi(\sigma_m^\alpha) = \sigma_n^\alpha$. Therefore,

$$d_p(\sigma_m^\alpha, \sigma_n^\alpha) \leq \left(\sum_{x\in\sigma_m^\alpha} d(x, \phi(x))^p\right)^{\frac{1}{p}} < \varepsilon_0 \leq \varepsilon.$$ □

Lemma 3.4.6. *For any $\alpha > 0$ there exists $\sigma^\alpha \in \mathcal{D}_p(X, A)$, with $|\sigma^\alpha| = M_\alpha$ and $u_\alpha(\sigma^\alpha) = \sigma^\alpha$, such that $d_p(\sigma_n^\alpha, \sigma^\alpha) \to 0$ as $n \to \infty$. Moreover, if $\alpha_1 > \alpha_2$, then $\sigma^{\alpha_1} \subset \sigma^{\alpha_2}$.*

Proof. We know that the sequence $\{\sigma_n^\alpha\}_{n\in\mathbb{N}}$ is Cauchy, thanks to Lemma 3.4.5, and $|\sigma_n^\alpha| = M_\alpha$ for sufficiently large n, thanks to Lemma 3.4.4. In particular, up to passing to a subsequence, we can assume that $x^1, \ldots, x^{M_\alpha}$ are the points in σ_1^α outside A and that there exist bijections $\phi_k\colon \sigma_k^\alpha \to \sigma_{k+1}^\alpha$ such that the sequences $\{x_n^1\}_{n\in\mathbb{N}}, \ldots, \{x_n^{M_\alpha}\}_{n\in\mathbb{N}}$, given by $x_1^i = x^i$, for $i = 1, \ldots, M_\alpha$, and $x_{k+1}^i = \phi_k(x_k^i)$, are Cauchy. Since X is complete, these sequences converge to points $\hat{x}^1, \ldots, \hat{x}^{M_\alpha}$, respectively, which actually are away from A since $d(x_n^i, A) \geq \alpha > 0$ for $i = 1, \ldots, M_\alpha$ and $n \in \mathbb{N}$. Let σ^α be the diagram given by the multiset $\left\{\!\!\left\{\hat{x}^1, \ldots, \hat{x}^{M_\alpha}\right\}\!\!\right\}$.

Hence, up to passing to a subsequence, $d_p(\sigma^\alpha, \sigma_n^\alpha) \to 0$ as $n \to \infty$. Nevertheless, since $\{\sigma_n^\alpha\}_{n\in\mathbb{N}}$ is Cauchy, than the previous limit holds for the whole sequence.

Finally, if $\alpha_1 > \alpha_2$ then $\sigma_n^{\alpha_1} \subset \sigma_n^{\alpha_2}$ for every $n \in \mathbb{N}$. The last part of the Lemma follows from this observation and the argument above. □

Lemma 3.4.7. *Let* $\sigma^* = \bigcup_{\alpha>0} \sigma^\alpha$. *Then* $\sigma^* \in \mathcal{D}_p(X, A)$ *and* $d_p(\sigma^\alpha, \sigma^*) \to 0$ *as* $\alpha \to 0$.

Proof. Let $\alpha > 0$ and $n \in \mathbb{N}$ be sufficiently large such that $d_p(\sigma^\alpha, \sigma_n^\alpha) < 1$. Then

$$d_p(\sigma^\alpha, \sigma_A) \leq d_p(\sigma^\alpha, \sigma_n^\alpha) + d_p(\sigma_n^\alpha, \sigma_A) \leq 1 + C$$

for some constant $C > 0$, independent of α, since $\{\sigma_n\}_{n\in\mathbb{N}}$ is Cauchy, and therefore bounded. This implies that $d_p(\sigma^*, \sigma_A) \leq 1 + C$, thus $\sigma^* \in \mathcal{D}_p(X, A)$.

Finally, note that

$$d_p(\sigma^\alpha, \sigma^*)^p \leq d_p(l_\alpha(\sigma^*), \sigma_A)^p = \sum_{\substack{x\in\sigma^* \\ d(x,A)<\alpha}} d(x, A)^p$$

and the right hand side of this inequality vanishes as $\alpha \to 0$, since it is the tail of an absolutely convergent series. □

Lemma 3.4.8. *For each* $\varepsilon > 0$, *there exists an* $\alpha_0 > 0$ *such that, for all* $n \in \mathbb{N}$ *and* $\alpha \in (0, \alpha_0]$, *we have* $d_p(\sigma_{n,\alpha}, \sigma_A) < \varepsilon$ *and, therefore,* $d_p(\sigma_n^\alpha, \sigma_n) < \varepsilon$.

Proof. Suppose there exists $\varepsilon > 0$ such that, for all $\alpha > 0$, there exists $n \in \mathbb{N}$ with $d_p(\sigma_{n,\alpha}, \sigma_A) \geq \varepsilon$. In particular, we obtain a subsequence $\{\sigma_{n_i}\}_{i\in\mathbb{N}}$ such that $d_p(\sigma_{n_i,1/i}, \sigma_A) \geq \varepsilon$.

Let $\delta \in (0, \varepsilon/4)$ and choose $k \in \mathbb{N}$ such that $d_p(\sigma_{n_k}, \sigma_{n_i}) < \delta$, for all $i \geq k$. Now, pick $j \geq k$ such that $d_p(\sigma_{n_k,1/i}, \sigma_A) < \delta$ for all $i \geq j$. This implies that $d_p(\sigma_{n_i,1/i}, \sigma_{n_k,1/j}) > 3\delta$ for $i \geq j$ by an application of the triangle inequality. For $i \geq j$ let $\phi_i \colon \sigma_{n_i} \to \sigma_{n_k}$ be a bijection such that $\sum_{x\in\sigma_{n_i}} d(x, \phi_i(x))^p < \delta^p$. Then also $\sum_{x\in\sigma_{n_i,1/i}} d(x, \phi_i(x))^p < \delta^p$.

Since $\delta_{1/j} > 0$, we can pick $l \geq j$ such that $\delta_{1/j} > 2/i$ for all $i \geq l$. If we now take $i \geq l$ and $x \in \sigma_{n_i,1/i}$ such that $\phi_i(x) \in \sigma_{n_k}^{1/j}$, we see that $d(x, \phi_i(x)) \geq d(x, A)$ by another application of the triangle inequality. Let $\hat{\phi}_i \colon \sigma_{n_i,1/i} \to \sigma_{n_k,1/j}$ be a bijection such that $\hat{\phi}_i(x) = \phi_i(x)$ if $\phi_i(x) \in \sigma_{n_k,1/j}$ and $\hat{\phi}_i(x) \in A$ if $\phi_i(x) \in \sigma_{n_k}^{1/j}$. Then, for $i \geq l$, we have

$$\sum_{x\in\sigma_{n_i,1/i}} d(x, \hat{\phi}_i(x))^p \leq \sum_{\substack{x\in\sigma_{n_i,1/i} \\ \phi_i(x)\in\sigma_{n_k,1/j}}} d(x, \phi_i(x))^p + \sum_{\substack{x\in\sigma_{n_i,1/i} \\ \phi_i(x)\in\sigma_{n_k}^{1/j}}} d(x, A)^p + \delta^p < 2\delta^p.$$

Therefore, $d_p(\sigma_{n_i,1/i}, \sigma_{n_k,1/j}) < 2^{1/p}\delta < 3\delta$ for $i \geq l$, which is a contradiction. □

By the triangle inequality,

$$d_p(\sigma^*, \sigma_n) \leq d_p(\sigma^*, \sigma^\alpha) + d_p(\sigma^\alpha, \sigma_n^\alpha) + d_p(\sigma_n^\alpha, \sigma_n)$$

and the Theorem follows as a consequence of Lemmas 3.4.6, 3.4.7 and 3.4.8. □

Remark 3.4.9. Observe that if X is complete then X/A, endowed with the quotient metric (see Remark 3.2.9), is complete as well. Indeed, if $\{[x_n]\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X/A, then $\{d([x_n], [A])\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, therefore it is convergent. If $d([x_n], [A]) \to 0$ as $n \to \infty$ then $[x_n] \to [A]$. Otherwise, $d([x_n], [A]) \to \delta > 0$ as $n \to \infty$, and for any $\varepsilon \in (0, \delta/2)$ there exists N such that $d([x_n], [x_m]) \leq \varepsilon < \delta/2$ for all $n, m \geq N$, implying that $[x_n], [x_m] \notin B_{\delta/2}([A])$, therefore and $d(x_n, x_m) = d([x_n], [x_m])$. In particular, $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X, thus it converges, which implies $\{[x_n]\}_{n\in\mathbb{N}}$ is convergent as well.

This observation, combined with Theorem 3.4.1, implies item 1 in Theorem A.

Proof of Theorem 3.4.2. Observe that the map $X/A \to \mathcal{D}_p(X/A, [A])$ given by

$[x] \mapsto \{\{[x]\}\}$ is a bi-Lipschitz embedding. Indeed,

$$\begin{aligned} d_p(\{\{[x]\}\}, \{\{[y]\}\}) &= \min\left\{d([x],[y]), (d([x],[A])^p + d([y],[A])^p)^{1/p}\right\} \\ &\leq d([x],[y]) \\ &= \min\{d(x,y), d(x,A) + d(y,A)\} \\ &\leq d(x,A) + d(y,A) \\ &\leq C(d(x,A)^p + d(y,A)^p)^{1/p} \end{aligned}$$

for some constant $C > 1$, due to the equivalence of norms in $\mathbb{R}^2$. This clearly implies

$$d_p(\{\{[x]\}\}, \{\{[y]\}\}) \leq d([x],[y]) \leq C d_p(\{\{[x]\}\}, \{\{[y]\}\}).$$

Since any subset of a separable metric space is separable, and separability is preserved by homeomorphisms, we conclude that the separability of $\mathcal{D}_p(X, A) \cong \mathcal{D}_p(X/A, [A])$ implies the separability of X/A.

Conversely, let S be a countable dense subset of X/A and define

$$\hat{S} = \{\sigma \in \mathcal{D}_p(X/A, [A]) : |\sigma| < \infty \text{ and } \sigma \subset S\}.$$

Let $\sigma \in \mathcal{D}_p(X/A, [A])$. Then, for each $\varepsilon > 0$, we can find $\alpha > 0$ such that $d_p(l_\alpha(\sigma), \sigma_{[A]}) < \varepsilon/2$, which implies $d_p(\sigma, u_\alpha(\sigma)) < \varepsilon/2$. Since $S^{|u_\alpha(\sigma)|}$ is dense in $(X/A)^{|u_\alpha(\sigma)|}$, we can find $\sigma' \in \hat{S}$ such that $d_p(\sigma', u_\alpha(\sigma)) < \varepsilon/2$. Then,

$$d_p(\sigma, \sigma') \leq d_p(\sigma, u_\alpha(\sigma)) + d_p(\sigma', u_\alpha(\sigma)) < \varepsilon,$$

which implies that $\hat{S}$ is dense.

Finally, note that $\hat{S} = \bigcup_{m=0}^{\infty} \hat{S}_m$, where $\hat{S}_m = \{\sigma \in \hat{S} : |\sigma| = m\}$. Each $\hat{S}_m$ can be embedded into S^m, thus it is countable. Hence, $\hat{S}$ is countable. □

Remark 3.4.10. Observe that, whenever X is separable, the quotient space X/A is separable too. Indeed, it is easy to see that if $\pi\colon X \to X/A$ is the quotient map, and $S \subset X$ is countable and dense, then $\pi(S) \subset X/A$ is countable and dense as well, since π is surjective and Lipschitz. In particular, Theorem 3.4.2 implies item 2 in Theorem A.

3.5 Geodesics

In this section, we show that the functor $\mathcal{D}_p$, with $p \in [1, \infty)$, preserves the geodesicity of metric spaces (Theorem 3.5.2), which proves item 3 in Theorem A. We also characterise geodesics in the space $\mathcal{D}_p(X, A)$ as given by geodesic interpolations (Theorem 3.5.4), adapting ideas from [20] to the context of general metric pairs.

Definition 3.5.1. A *geodesic interpolation* in $\mathcal{D}_p(X, A)$ is a curve $(\sigma_t)_{t\in[0,1]}$ in $\mathcal{D}_p(X, A)$ such that there exist a bijection $\phi\colon \sigma_0 \to \sigma_1$ and $\{\xi^x\}_{x\in\sigma_0} \subset \mathrm{Geo}(X)$ such that ξ^x joins x with $\phi(x)$, for each $x \in \sigma_0$, and $\sigma_t = \{\{\xi_t^x : x \in \sigma_0\}\}$ for each $t \in [0, 1]$.

Theorem 3.5.2. *Let $(X, A) \in \mathsf{Met_{Pair}}$. If X is a proper geodesic space, then $\mathcal{D}_p(X, A)$ is a geodesic space.*

Proof. Let $\sigma_0, \sigma_1 \in \mathcal{D}_p(X, A)$ be diagrams, $\phi\colon \sigma_0 \to \sigma_1$ be an optimal bijection as in Theorem 3.3.3 and let $(\sigma_t)_{t\in[0,1]}$ be a geodesic interpolation induced by ϕ. Then $(\sigma_t)_{t\in[0,1]}$ is a geodesic joining σ_0 and σ_1. Indeed, if for any $s, t \in [0, 1]$ we consider the bijection $\phi_s^t\colon \sigma_s \to \sigma_t$ given by $\phi_s^t(\xi_s^x) = \xi_t^x$, then

$$\begin{aligned} d_p(\sigma_s, \sigma_t)^p &\le \sum_{x'\in\sigma_s} d(x', \phi_s^t(x'))^p \\ &= \sum_{x\in\sigma_0} d(\xi_s^x, \xi_t^x)^p \\ &= |s - t|^p \sum_{x\in\sigma_0} d(x, \phi(x))^p \\ &= |s - t|^p d_p(\sigma_0, \sigma_1)^p. \end{aligned}$$

Therefore $(\sigma_t)_{t\in[0,1]}$ is a geodesic from σ_0 to σ_1. □

Lemma 3.5.3. *Let $(\sigma_t)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{D}_p(X, A))$ be a geodesic, $t_0 \in [0, 1]$, and $\phi_i\colon \sigma_{t_0} \to \sigma_i$, $i = 0, 1$, be optimal bijections. Then $\phi = \phi_1 \circ \phi_0^{-1}\colon \sigma_0 \to \sigma_1$ is an optimal bijection, and, for any $x \in \sigma_t$, x is a t-intermediate between $\phi_0(x)$ and $\phi_1(x)$, i.e. $d(x, \phi_0(x)) = td(\phi_0(x), \phi_1(x))$ and $d(x, \phi_1(x)) = (1 - t)d(\phi_0(x), \phi_1(x))$.*

Proof. By the triangle inequality both in X and in space of ℓ^p-summable sequences, we have the following:

$$\begin{aligned}
d_p(\sigma_0,\sigma_1) &\le \left(\sum_{x\in\sigma_t} d(\phi_0(x),\phi_1(x))^p\right)^{1/p} \\
&\le \left(\sum_{x\in\sigma_t} (d(\phi_0(x),x)+d(x,\phi_1(x)))^p\right)^{1/p} \\
&\le \left(\sum_{x\in\sigma_t} d(\phi_0(x),x)^p\right)^{1/p} + \left(\sum_{x\in\sigma_t} d(x,\phi_1(x))^p\right)^{1/p} \\
&= d_p(\sigma_0,\sigma_t)+d_p(\sigma_t,\sigma_1) \\
&= d_p(\sigma_0,\sigma_1).
\end{aligned}$$

It follows that

$$d_p(\sigma_0,\sigma_1) = \left(\sum_{z\in\sigma_0} d(z,\phi(z))^p\right)^{1/p}$$

and

$$d(\phi_0(x),x) = td(\phi_0(x),\phi_1(x)), \quad d(x,\phi_1(x)) = (1-t)d(\phi_0(x),\phi_1(x))$$

for all $x\in\sigma_t$, which proves the claim. □

Theorem 3.5.4. *Let $(X,A)\in \mathsf{Met}_{\mathsf{Pair}}$ and assume X is a proper geodesic space. Then every geodesic in $\mathcal{D}_p(X,A)$ is a geodesic interpolation.*

Proof. This argument closely follows the proofs of [20, Theorems 10 and 11]. We repeat some of the constructions for the convenience of the reader.

Fix $(\sigma_t)_{t\in[0,1]}\in \mathrm{Geo}(\mathcal{D}_p(X,A))$. We first claim there exists a sequence of geodesic interpolations $\{(\sigma_t^n)_{t\in[0,1]}\}_{n\in\mathbb{N}}$ such that $\sigma_{i/2^n}=\sigma^n_{i/2^n}$, for each $n\in\mathbb{N}$ and $i\in\{0,\dots,2^n\}$.

Indeed, for every $n\in\mathbb{N}$ and every $i\in\{1,\dots,2^{n-1}\}$, consider optimal bijections $\phi^{\pm}_{n,i}\colon \sigma_{(2i-1)/2^n}\to\sigma_{(2i-1\pm1)/2^n}$. By Lemma 3.5.3,

$$\phi_n=\phi^+_{n,2^{n-1}}\circ(\phi^-_{n,2^{n-1}})^{-1}\circ\cdots\circ\phi^+_{n,1}\circ(\phi^-_{n,1})^{-1}$$

is an optimal bijection between σ_0 and σ_1. Moreover, Lemma 3.5.3 implies that, for each $x\in\sigma_{(2i-1)/2^n}$, there is some geodesic joining $\phi^-_{n,i}(x)$ with $\phi^+_{n,i}(x)$ which has x as its midpoint. This way, starting from any point $x \in \sigma_0$ and following the bijections $\phi^{\pm}_{n,i}$, we construct a geodesic $\xi^{x,n}$ joining x with $\phi_n(x)$.

Now, thanks to Lemma 3.3.2, up to passing to a subsequence, we can assume that $\{\phi_n\}_{n\in\mathbb{N}}$ is pointwise convergent to some optimal bijection $\phi\colon \sigma_0 \to \sigma_1$. Moreover, we can extract a further subsequence $\{\phi_{n_k}\}_{k\in\mathbb{N}}$ such that, for fixed dyadic rationals $l/2^j$ and $l'/2^j$, the sequence of bijections $\sigma_{l/2^j} \to \sigma_{l'/2^j}$ induced by $\{\phi_{n_k,i}\}_{k\in\mathbb{N}}$ is pointwise convergent as well. By Arzelà–Ascoli theorem and a diagonal argument, we may assume that for each $x \in \sigma_0$ the sequence $\{\xi^{x,n_k}\}_{k\in\mathbb{N}}$ is uniformly convergent to some geodesic ξ^x joining x with $\phi(x)$. By continuity, it follows that $(\sigma_t)_{t\in[0,1]}$ is the geodesic interpolation induced by ϕ and the set of geodesics $\{\xi^x\}$. □

3.6 Non-negative curvature

In this section, we prove that the functor $\mathcal{D}_2$ preserves non-negative curvature in the sense of Definition 2.2.2 (cf. [53, Theorem 2.5] and [20, Theorems 10 and 11]). On the other hand, it is known that the functor $\mathcal{D}_p$ does not preserve non-negative curvature for $p \neq 2$ (see [52]). Also, $\mathcal{D}_p$ does not preserve upper curvature bounds in the sense of CAT spaces for any p (cf. [53, Proposition 2.4] and [52, Proposition 2.4]). Whether the functor $\mathcal{D}_2$ preserves strictly negative lower curvature bounds remains an open question.

Theorem 3.6.1. *Let $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$. If X is a proper Alexandrov space with non-negative curvature, then, $\mathcal{D}_2(X, A)$ is also an Alexandrov space with non-negative curvature.*

Proof. Since X is an Alexandrov space, it is complete and geodesic. Thus, by Theorem 3.4.1, the space $\mathcal{D}_2(X, A)$ is complete, and, since X is assumed to be proper, Theorems 3.5.2 imply that $\mathcal{D}_2(X, A)$ is geodesic. Now we must show that $\mathcal{D}_2(X, A)$ has non-negative curvature.

Let $\sigma, \sigma_0, \sigma_1 \in \mathcal{D}_2(X, A)$ be diagrams and $(\sigma_t)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{D}_2(X, A))$ be a geodesic from σ_0 to σ_1. We want to show the inequality

$$d_2(\sigma, \sigma_t)^2 \geq (1-t)d_2(\sigma, \sigma_0)^2 + td_2(\sigma, \sigma_1)^2 - (1-t)td_2(\sigma_0, \sigma_1)^2$$

for any $t \in [0, 1]$ (see Remark 2.2.3).

Fix $t \in [0, 1]$ and let $\phi_i \colon \sigma_t \to \sigma_i$, $i = 0, 1$, and $\phi \colon \sigma_t \to \sigma$, be optimal bijections, and define $\Phi = \phi_1 \circ \phi_0^{-1} \colon \sigma_0 \to \sigma_1$. From the formula for the distance in $\mathcal{D}_2(X, A)$ we observe that the following inequalities hold:

$$\begin{aligned} d_2(\sigma, \sigma_t)^2 &= \sum_{x\in\sigma_t} d(x, \phi(x))^2; \\ d_2(\sigma, \sigma_0)^2 &\leq \sum_{x\in\sigma_t} d(\phi(x), \phi_0(x))^2; \\ d_2(\sigma, \sigma_1)^2 &\leq \sum_{x\in\sigma_t} d(\phi(x), \phi_1(x))^2. \end{aligned}$$

Now, since $\mathrm{curv}(X) \geq 0$, we have that

$$d(x, \phi(x))^2 \geq (1-t)d(\phi(x), \phi_0(x))^2 + td(\phi(x), \phi_1(x))^2 - (1-t)td(\phi_0(x), \phi_1(x))^2$$

for all $x \in \sigma_t$. Therefore, thanks to Lemma 3.5.3,

$$\begin{aligned} d_2(\sigma, \sigma_t)^2 &= \sum_{x\in\sigma_t} d(x, \phi(x))^2 \\ &\geq \sum_{x\in\sigma_t} (1-t)d(\phi(x), \phi_0(x))^2 \\ &\qquad\qquad + td(\phi(x), \phi_1(x))^2 - (1-t)td(\phi_0(x), \phi_1(x))^2 \\ &\geq (1-t)d_2(\sigma, \sigma_0)^2 + td_2(\sigma, \sigma_1)^2 - (1-t)td_2(\sigma_0, \sigma_1)^2. \end{aligned}$$

□

Remark 3.6.2. We note that $\mathcal{D}_2(X, A)$ does not satisfy $\mathrm{curv}(X) \geq k$ for any $\kappa > 0$ in general. To see this, let (X, A) be a metric pair, where X is proper and geodesic. For $i \in \{1, 2, 3\}$, let $x_i \in X \setminus A$ and let $\xi_i \colon [0, 1] \to X$ be a constant speed geodesic with $\xi_i(0) \in A$ and $\xi_i(1) = x_i$ of minimal length, i.e. of length $d(x_i, A) = \min_{a\in A} d(x_i, a)$; such ξ_i exists since X is proper and A is closed. Suppose that

$$d(\xi_i(s), \xi_j(t))^2 \geq d(\xi_i(0), \xi_i(s))^2 + d(\xi_j(0), \xi_j(t))^2 \text{ whenever } i \neq j. \tag{3.10}$$

For $i = 1, 2, 3$, let $\sigma_i = \{\{x_i\}\} \in \mathcal{D}_2(X, A)$. It follows from (3.10) that $d(x_i, x_j)^2 \geq d(x_i, A)^2 + d(x_j, A)^2$ for $i \neq j$, and therefore $d_2(\sigma_i, \sigma_j) = \sqrt{d(x_i, A)^2 + d(x_j, A)^2}$. It is then easy to see that the path $\eta_{i,j} \colon [0, 1] \to \mathcal{D}_2(X, A)$, where

$$\eta_{i,j}(t) = \{\{\xi_i(1 - t), \xi_j(t)\}\},$$

is a constant speed geodesic in $\mathcal{D}_2(X, A)$ from σ_i to σ_j. But it is then easy to verify, again using (3.10), that

$$d_2(\sigma_k, \eta_{i,j}(t)) = \sqrt{d(x_k, A)^2 + d(\xi_i(1 - t), A)^2 + d(\xi_j(t), A)^2},$$

where $k \notin \{i, j\}$. In particular, it follows that the geodesic triangle in $\mathcal{D}_2(X, A)$ formed by geodesics $\eta_{1,2}$, $\eta_{2,3}$ and $\eta_{3,1}$ is isometric to the geodesic triangle in $\mathbb{R}^3$ with vertices $(d(x_1, A), 0, 0)$, $(0, d(x_2, A), 0)$ and $(0, 0, d(x_3, A))$.

The condition (3.10) is not hard to achieve: it can be achieved whenever X is a connected Riemannian manifold of dimension ≥ 2 and $A \neq X$, for instance. Indeed, in that case, if $|\partial A| \geq 3$ then (3.10) is satisfied for any $x_1, x_2, x_3 \in X \setminus A$ with $d(x_i, a_i) \leq \varepsilon/6$, where $a_1, a_2, a_3 \in \partial A$ are distinct elements and $\varepsilon = \min\{d(a_i, a_j) : i \neq j\}$. On the other hand, if $|\partial A| \geq 2$ then $|A| \leq 2$ since X is connected of dimension ≥ 2, and so we may pick $x_1, x_2, x_3 \in X \setminus A$ in such a way that $d(x_1, a) = d(x_2, a) = d(x_3, a) = \varepsilon < d(x_i, b)$ for any i and any $b \in A \setminus \{a\}$, where $a \in A$ is a fixed element. It then follows that $\xi_i(0) = a$ for each i. Since $\dim X \geq 2$, we may do this in such a way that the angle between ξ_i and ξ_j at a is $> \pi/2$ when $i \neq j$; but then, as a consequence of the Rauch comparison theorem, (3.10) will be satisfied whenever $\varepsilon > 0$ is chosen small enough.

Remark 3.6.3. Let X be an Alexandrov space and let $K \subset X$ be a convex subset, i.e. such that any geodesic joining any two points in K remains inside K (cf. [12, p. 90]). It is a direct consequence of the definition that K is also an Alexandrov space with the same lower curvature bound as X. In particular, if $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$ with $\mathrm{curv}(X) \geq 0$, and $K \subset X$ is a convex subset with $A \subset K$, then $\mathrm{curv}(\mathcal{D}_2(K, A)) \geq 0$.

Remark 3.6.4. Observe that, finally, as a consequence of Theorems 3.4.1, 3.4.2, 3.5.2, 3.6.1, and Remarks 3.4.9 and 3.4.10, Theorem A follows.

3.6.1 Spaces of directions

In this section we prove some metric properties of the space of directions Σ_{σ_A} at the empty diagram $\sigma_A \in \mathcal{D}_2(X, A)$ for $(X, A) \in \mathsf{Met}_{\mathsf{Pair}}$ with X a proper Alexandrov space with non-negative curvature.

The following proposition shows that $\mathcal{D}_2(X, A)$ always has at least one *extremal point*, i.e. a point such that the corresponding space of directions has diameter at most $\pi/2$ (see [46] for more details about extremal points and extremal sets in Alexandrov spaces).

Proposition 3.6.5. *The empty diagram σ_A is an extremal point in $\mathcal{D}_2(X, A)$.*

Proof. Consider $\sigma, \sigma' \in \mathcal{D}_2(X, A)$. We can always consider a bijection $\phi\colon \sigma \to \sigma'$ such that $\phi(a) = A$ for every $a \in \sigma$ different from A and $\phi^{-1}(a') = A$ for every $a' \in \sigma'$ different from A, since both σ and σ' are countable and contain countably infinite copies of A. Thus, by definition of the distance function d_2, we have

$$d_2(\sigma, \sigma')^2 \leq \sum_{a \in \sigma} d(a, A)^2 + \sum_{a' \in \sigma'} d(a', A)^2 = d_2(\sigma, \sigma_A)^2 + d_2(\sigma', \sigma_A)^2.$$

Therefore,

$$\cos \widetilde{\measuredangle}_0 \sigma \sigma_A \sigma' = \frac{d_2(\sigma, \sigma_A)^2 + d_2(\sigma', \sigma_A)^2 - d_2(\sigma, \sigma')^2}{2 d_2(\sigma, \sigma_A) d_2(\sigma', \sigma_A)} \geq 0,$$

i.e. $\widetilde{\measuredangle}_0 \sigma \sigma_A \sigma' \leq \pi/2$. This immediately implies the result. □

Proposition 3.6.6. *For any $\sigma \in \mathcal{D}_2(X, A)$, geodesics directions in Σ_σ corresponding to diagrams with finitely many points are dense in Σ_σ.*

Proof. Let $\sigma, \tau \in \mathcal{D}_2(X, A)$ and $(\tau_t)_{t \in [0,1]} \in \mathrm{Geo}(\mathcal{D}_2(X, A))$ be such that $\tau_0 = \sigma$ and $\tau_1 = \tau$. By Theorem 3.5.4, we know that there exist an optimal bijection $\phi\colon \sigma \to \tau$, and a set of geodesics $\{\xi^x\}_{x\in\sigma} \subset \mathrm{Geo}(X)$ such that ξ^x joins x with $\phi(x)$ for any $x \in \sigma$, and such that $\tau_t = \{\{\xi^x_t : x \in \sigma\}\}$ for any $t \in [0,1]$.

Now, let $\{(a_i, b_i)\}_{i\in\mathbb{N}} \subset X \times X$ be such that $\sigma = \{\{a_i : i \in \mathbb{N}\}\}$ and $\tau = \{\{b_i : i \in \mathbb{N}\}\}$, and such that $\phi(a_i) = b_i$. We can then define sequences of finite diagrams $\{\sigma_n\}_{n\in\mathbb{N}}$ and $\{\tau_n\}_{n\in\mathbb{N}}$ given by

$$\sigma_n = \{\{a_1, \ldots, a_n\}\}\,, \quad \tau_n = \{\{b_1, \ldots, b_n\}\}\,.$$

Let $(\tau^n_t)_{t\in[0,1]}$ be the geodesic interpolation given by $\phi|_{\sigma_n} : \sigma_n \to \tau_n$ and $\{\xi^x\}_{x\in\sigma_n}$. Moreover, observe that $\phi|_{\sigma_n}$ is an optimal bijection, since being otherwise would contradict the optimality of ϕ. It is then clear that

$$\begin{aligned} d_2(\tau^n_t, \tau_t)^2 &\le \sum_{x\in\sigma\setminus\sigma_n} d(x, \xi^x_t)^2 \\ &= t^2 \sum_{x\in\sigma\setminus\sigma_n} d(x, \phi(x))^2 \\ &= t^2 \left(d_2(\sigma,\tau)^2 - d_2(\sigma_n, \tau_n)^2 \right). \end{aligned}$$

Thus, using the definition of angle between geodesics in an Alexandrov space (see Remark 2.2.4) and the law of cosines, we get that

$$\begin{aligned} 1 &\ge \cos \measuredangle \tau_n \sigma \tau \\ &= \lim_{t\to 0} \frac{t^2(d_2(\tau_n,\sigma)^2 + d_2(\tau,\sigma)^2) - d_2(\tau^n_t, \tau_t)^2}{2t^2 d_2(\tau_n,\sigma) d_2(\tau,\sigma)} \\ &\ge \lim_{t\to 0} \frac{t^2(d_2(\tau_n,\sigma)^2 + d_2(\tau,\sigma)^2 - d_2(\sigma,\tau)^2 + d_2(\sigma_n,\tau_n)^2)}{2t^2 d_2(\tau_n,\sigma) d_2(\tau,\sigma)} \\ &= \frac{d_2(\tau_n,\sigma)^2 + d_2(\sigma_n,\tau_n)^2}{2 d_2(\tau_n,\sigma) d_2(\tau,\sigma)}, \end{aligned}$$

and the last quotient converges to 1 as $n \to \infty$ due to the continuity of d_2 and the fact that $\tau_n \to \tau$ and $\sigma_n \to \sigma$ as $n \to \infty$. Therefore, $\measuredangle \tau_n \sigma \tau$ converges to 0. This proves that geodesic directions joining σ with finite diagrams are dense in the set of all geodesic directions at σ, and since Σ_σ is the metric completion of the set of geodesic directions at σ, the result follows. □

Moreover, we can express the angle between any two geodesic directions in Σ_{σ_A} determined by finite diagrams, as the following result show.

Lemma 3.6.7. *Let σ and σ' be diagrams with finitely many points, and let $\xi, \xi' \in \mathrm{Geo}(\mathcal{D}_2(X, A))$ be such that $\xi_0 = \xi'_0 = \sigma_A$, $\xi_1 = \sigma$, and $\xi'_1 = \sigma'$, and such that $\xi_t = \{\{\xi^x_t : x \in \sigma\}\}$ and $\xi'_t = \left\{\left\{\xi^{x'}_t : x' \in \sigma'\right\}\right\}$ for some $\{\xi^x\}_{x\in\sigma}$, $\{\xi^{x'}\}_{x'\in\sigma'} \subset \mathrm{Geo}(X)$. Then*

$$d_2(\sigma, \sigma_A) d_2(\sigma', \sigma_A) \cos \measuredangle(\xi, \xi') = \sum_{x\in\tau} d(x, A) d(\phi(x), A) \cos \measuredangle(\xi^x, \xi^{\phi(x)}),$$

for some bijection $\phi \colon \tau \to \tau'$ between $\tau \subset \sigma$ and $\tau' \subset \sigma'$, and such that $\xi^x_0 = \xi^{\phi(x)}_0$ for all $x \in \tau$.

Proof. For each $t \in (0, 1]$, let $\phi'_t \colon \xi_t \to \xi'_t$ be an optimal bijection. Then there exists a bijection $\phi_t \colon \tau_t \to \tau'_t$ for some $\tau_t \subset \sigma$ and $\tau'_t \subset \sigma'$ such that $\phi'_t(\xi^x_t) = \xi^{\phi(x)}_t$ for any $x \in \tau_t$, and such that ϕ'_t matches all the other points in $\xi_t \cup \xi'_t$ to points in A in their corresponding geodesic ξ^x or $\xi^{x'}$. Moreover, by the optimality of ϕ_t, we have

$$d(\xi^x_t, \xi^{\phi_t(x)}_t)^2 \leq t^2 d(x, A)^2 + t^2 d(\phi_t(x), A)^2$$

for all $x \in \tau_t$. Therefore, by the triangle inequality, for t is sufficiently small, $\xi^x_0 = \xi^{\phi_t(x)}_0$ for all $x \in \tau_t$.

Furthermore, since σ and σ' are finite, there are only finitely many choices of τ_t, τ'_t and ϕ_t, which implies there exist some fixed $\tau \subset \sigma$, $\tau' \subset \sigma'$, and a bijection $\phi \colon \tau \to \tau'$ such that, for some $\{t_i\}_{n\in\mathbb{N}} \subset (0, 1]$ with $t_i \to 0$ as $i \to \infty$, all the conditions above hold with $\tau_{t_i} = \tau$, $\tau'_{t_i} = \tau'$ and $\phi_{t_i} = \phi$.

Since

$$d_2(\xi_{t_i}, \xi'_{t_i})^2 = t_i^2 \left(\sum_{x\in\sigma\setminus\tau} d(x, A)^2 + \sum_{x'\in\sigma'\setminus\tau'} d(x', A)^2 \right) + \sum_{x\in\tau} d\left(\xi^x_{t_i}, \xi^{\phi(x)}_{t_i}\right)^2,$$

it follows that

$$t_i^2 d_2(\sigma, \sigma_A)^2 + t_i^2 d_2(\sigma', \sigma_A)^2 - d_2(\xi_{t_i}, \xi'_{t_i})^2 = \sum_{x\in\tau} t_i^2 d(x, A)^2 + t_i^2 d(\phi(x), A)^2 - d\left(\xi^x_{t_i}, \xi^{\phi(x)}_{t_i}\right)^2, \quad (3.11)$$

which implies that

$$
\begin{aligned}
d_2(\sigma,\sigma_A)d_2(\sigma',\sigma_A)\cos\measuredangle(\xi,\xi') &= \lim_{i\to\infty}\frac{t_i^2 d_2(\sigma,\sigma_A)^2 + t_i^2 d_2(\sigma',\sigma_A)^2 - d_2(\xi_{t_i},\xi'_{t_i})^2}{2t_i^2} \\
&= \lim_{i\to\infty}\sum_{x\in\tau}\frac{t_i^2 d(x,A)^2 + t_i^2 d(\phi(x),A)^2 - d\left(\xi^x_{t_i},\xi^{\phi(x)}_{t_i}\right)^2}{2t_i^2} \\
&= \sum_{x\in\tau} d(x,A)d(\phi(x),A)\cos\measuredangle(\xi^x,\xi^{\phi(x)})
\end{aligned} \tag{3.12}
$$

due to the finiteness of τ. The result follows. □

Using Lemma 3.6.7 and the density of geodesic directions in Σ_{σ_A} corresponding to diagrams with finitely many points, we get Theorem B.

3.7 Gromov–Hausdorff continuity

In this section, we investigate the continuity of the functor $\mathcal{D}_p$ with respect to the Gromov–Hausdorff convergence of metric pairs and pointed metric spaces. We first recall the notion of Gromov–Hausdorff convergence of metric pairs (X,A), as introduced in [17] and studied in [1]. This is a natural extension of the definition of Gromov–Hausdorff convergence for pointed metric spaces (Definition 2.3.3).

Definition 3.7.1. A sequence $\{(X_i,A_i)\}_{i\in\mathbb{N}}$ *converges in the Gromov–Hausdorff topology to a metric pair* (X,A) if there exist sequences $\{\varepsilon_i\}_{i\in\mathbb{N}}$ and $\{R_i\}_{i\in\mathbb{N}}$ of positive numbers with $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$, and maps $\phi_i\colon \overline{B}_{R_i}(A_i)\to X$ satisfying the following three conditions:

1. $|d_{X_i}(x,y) - d_X(\phi_i(x),\phi_i(y))| \leq \varepsilon_i$ for any $x,y\in\overline{B}_{R_i}(A_i)$;
2. $d_{\mathsf{H}}^{d_X}(\phi_i(A_i),A)\leq\varepsilon_i$;
3. $\overline{B}_{R_i}(A)\subset\overline{B}_{\varepsilon_i}(\phi_i(\overline{B}_{R_i}(A_i)))$.

We will denote the Gromov–Hausdorff convergence of metric pairs by

$$(X_i,A_i)\xrightarrow{\mathsf{GH}}(X,A).$$

We first observe that $\mathcal{D}_p\colon \mathsf{Met}_{\mathsf{Pair}} \to \mathsf{Met}_*$ is not sequentially continuous for any $p \in [1, \infty)$, as the following example shows.

Example 3.7.2. Let $X_i = [-\frac{1}{i}, \frac{1}{i}] \subset \mathbb{R}$ and set $A_i = X = A = \{0\}$. Then $\mathcal{D}_p(X, A) = \{\sigma_A\}$. Observe that for $p \neq \infty$, the space $\mathcal{D}_p(X_i, A_i)$ is unbounded. Indeed, if σ_n is the diagram that contains a single point, $1/i$, with multiplicity n, then $d_p(\sigma_n, \sigma_\varnothing) = \sqrt[p]{n}/i \to \infty$ as $n \to \infty$.

Now, let $\sigma_{A_i} \in \mathcal{D}_p(X_i, A_i)$ be the empty diagram and suppose, for the sake of contradiction, that there exist ε_i-approximations $f_i\colon \overline{B}_{R_i}(\sigma_{A_i}) \to \mathcal{D}_p(X, A)$ for some $\varepsilon_i \searrow 0$ and $R_i \nearrow \infty$. Then

$$|d_p(\sigma, \sigma_{A_i}) - d_p(f_i(\sigma), f_i(\sigma_{A_i}))| \leq \varepsilon_i$$

for all $\sigma \in \overline{B}_{R_i}(\sigma_{A_i})$. However, we have $d_p(f_i(\sigma), f(\sigma_{A_i})) = d_p(\sigma_A, \sigma_A) = 0$, implying that

$$d_p(\sigma, \sigma_{A_i}) \leq \varepsilon_i \tag{3.13}$$

for all $\sigma \in \overline{B}_{R_i}(\sigma_{A_i})$. As $\varepsilon_i \to 0$ and $R_i \to \infty$ as $i \to \infty$, inequality (3.13) contradicts the fact that $\mathcal{D}_p(X_i, A_i)$ is unbounded for each i.

In order to prove the continuity of $\mathcal{D}_\infty$, we first prove the sequential continuity of the quotient functor $\mathcal{Q}\colon \mathsf{Met}_{\mathsf{Pair}} \to \mathsf{Met}_*$.

Proposition 3.7.3. *The quotient functor $\mathcal{Q}\colon \mathsf{Met}_{\mathsf{Pair}} \to \mathsf{Met}_*$, given by $(X, A) \mapsto (X/A, [A])$, is sequentially continuous with respect to the Gromov–Hausdorff convergence of metric pairs.*

Proof. We will prove that, if there exist $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$ and ε_i-approximations from $\overline{B}_{R_i}(A_i)$ to $\overline{B}_{R_i}(A)$, then, they there exist $(5\varepsilon_i)$-approximations from $\overline{B}_{R_i}([A_i])$ to $\overline{B}_{R_i}([A])$. For ease of notation, we will omit the subindices in the metric which indicate the corresponding metric space.

Let f_i be an ε_i-approximation from $\overline{B}_{R_i}(A_i)$ to $\overline{B}_{R_i}(A)$ in the sense of Definition 3.7.1. Then, for any $x \in \overline{B}_{R_i}(A_i)$, $a_i \in A_i$, we have

$$|d(x, a_i) - d(f_i(x), f_i(a_i))| \leq \varepsilon_i$$

which implies

$$|d(x, A_i) - d(f_i(x), f_i(A_i))| \leq \varepsilon_i. \tag{3.14}$$

Moreover, for any $a_i \in A_i$ and $a \in A$, we have

$$|d(f_i(x), f_i(a_i)) - d(f_i(x), a)| \leq d(f_i(a_i), a)$$

and, since $d_H(f_i(A_i), A) \leq \varepsilon_i$, this yields

$$|d(f_i(x), f_i(A_i)) - d(f_i(x), A)| \leq \varepsilon_i. \tag{3.15}$$

Combining inequalities (3.14) and (3.15), we get

$$|d(x, A_i) - d(f_i(x), A)| \leq 2\varepsilon_i.$$

Now, for each i, define $\underline{f}_i \colon \overline{B}_{R_i}([A_i]) \to X/A$ by

$$\underline{f}_i([x]) = \begin{cases} [f_i(x)] & \text{if } [x] \neq [A_i], \\ [A] & \text{if } [x] = [A_i]. \end{cases}$$

We will prove that $\underline{f}$ is a $(5\varepsilon_i)$-approximation from $\overline{B}_{R_i}([A_i])$ to $\overline{B}_{R_i}([A])$. Indeed, consider $[x], [y] \in \overline{B}_{R_i}([A_i]) \setminus \{[A_i]\}$. Then $x, y \in B_{R_i}(A_i)$ and therefore

$$\begin{aligned}
&|d([x],[y]) - d(\underline{f}_i([x]), \underline{f}_i([y]))| \\
&\qquad = |\min\{d(x,y), d(x,A_i) + d(y,A_i)\} \\
&\qquad\qquad - \min\{d(f_i(x), f_i(y)), d(f_i(x), A) + d(f_i(y), A)\}| \\
&\qquad \leq |d(x,y) - d(f_i(x), f_i(y))| + |d(x, A_i) - d(f_i(x), A)| \\
&\qquad\qquad + |d(y, A_i) - d(f_i(y), A)| \\
&\qquad \leq \varepsilon_i + 2\varepsilon_i + 2\varepsilon_i = 5\varepsilon_i.
\end{aligned}$$

If $[x] \neq [A_i]$ and $[y] = [A_i]$, then

$$|d([x],[y]) - d(\underline{f}_i([x]), \underline{f}_i([y]))| = |d(x, A_i) - d(f_i(x), A)| \leq 2\varepsilon_i.$$

A similar inequality is obtained when $[y] \neq [A_i]$ and $[x] = [A_i]$. When both $[x] = [A_i]$ and $[y] = [A_i]$, we get

$$|d([x],[y]) - d(\underline{f}_i([x]), \underline{f}_i([y]))| = 0.$$

In any case, we see that the distortion of $\underline{f}_i$ is $\leq 5\varepsilon_i$, which is item (1) in Definition 3.7.1.

For item (2) in Definition 3.7.1, we simply observe that by definition of $\underline{f}_i$ they are pointed maps, therefore

$$d_H(\underline{f}_i(\{[A_i]\}), \{[A]\}) = d(\underline{f}_i([A_i]), [A]) = 0.$$

Finally, we see that for $[y] \in B_{R_i}([A])$ we have $d(y, A) \leq R_i$, so given that f_i is an ε_i-approximation from $B_{R_i}(A_i)$ to $B_{R_i}(A)$ there exists $x \in B_{R_i}(A_i)$ such that $d(y, f_i(x)) \leq \varepsilon_i$. Therefore,

$$d([y], \underline{f}_i[x]) \leq d(y, f_i(x)) \leq \varepsilon_i.$$

Thus $[y] \in B_{\varepsilon_i}(\underline{f}_i(B_{R_i}(A_i)))$. This gives item (3) in Definition 3.7.1. □

Recalling that $\mathcal{D}_\infty$ maps $\mathsf{Met}_{\mathsf{Pair}}$ into PMet_*, we need the following notation. Namely, given a pseudometric space X, we denote by $\underline{X}$ the metric space canonically obtained by identifying points at zero distance (see, for example, [12, Chapter 1]). We also denote sometimes by $\underline{x}$ the image of $x \in X$ under this identification.

The following proposition shows that pointed Gromov–Hausdorff convergence of pseudometric spaces induces pointed Gromov–Hausdorff convergence of the corresponding metric quotients.

Proposition 3.7.4. *Let $\{(X_i, x_i)\}_{i\in\mathbb{N}}$, (X, x) be pointed pseudometric spaces and let $\pi_i \colon X_i \to \underline{X}_i$, $\pi \colon X \to \underline{X}$ be the canonical identifications. Then the following assertions hold:*

1. *If $(X_i, x_i) \xrightarrow{\mathsf{GH}_*} (X, x)$, then $(\underline{X}_i, \underline{x}_i) \xrightarrow{\mathsf{GH}_*} (\underline{X}, \underline{x})$.*

2. *If* $(\underline{X}_i, \underline{x}_i) \xrightarrow{\mathsf{GH}_*} (\underline{X}, \underline{x})$, *then* $(X_i, x_i) \xrightarrow{\mathsf{GH}_*} (X, x)$.

Proof. For each i, consider $s_i\colon \underline{X}_i \to X_i$ such that $\pi_i(s_i(x)) = x$ for all $x \in \underline{X}_i$ and $s\colon \underline{X} \to X$ similarly. These maps exist due to the axiom of choice. Let f_i be ε_i-approximations from $\overline{B}_{R_i}(x_i)$ to $\overline{B}_{R_i}(x)$. Define $\underline{f}_i\colon \overline{B}_{R_i}(\underline{x}_i) \to \underline{X}$ as

$$\underline{f}_i(x) = \pi(f_i(s_i(x)))$$

for any $x \in \underline{X}_i$. Then $\underline{f}_i$ is a $(2\varepsilon_i)$-approximation from $\overline{B}_{R_i}(\underline{x}_i))$ to $\overline{B}_{R_i}(\underline{x})$. Indeed,

$$|d(x, y) - d(\underline{f}_i(x), \underline{f}_i(y))| = |d(s_i(x), s_i(y)) - d(f_i(s_i(x)), f_i(s_i(y)))| \le \varepsilon_i.$$

Also

$$\begin{aligned} d(\underline{f}_i(\underline{x}_i), \underline{x}) &= d(f_i(s_i(\underline{x}_i)), x) \\ &\le d(f_i(s_i(\underline{x}_i)), f_i(x_i)) + d(f_i(x_i), x) \\ &\le d(s_i(\underline{x}_i), x_i) + \varepsilon_i + d(f_i(x_i), x) \\ &\le 2\varepsilon_i. \end{aligned}$$

Moreover, if $d(x, \pi(x)) \le R_i$ then $d(s(x), x) \le R_i$. Then there is some $y \in X_i$ with $d(y, x_i) \le R_i$ such that $d(s(x), f_i(y)) \le \varepsilon_i$. Therefore,

$$\begin{aligned} d(x, \underline{f}_i(\underline{y})) &= d(s(x), f_i(s_i(\underline{y}))) \\ &\le d(s(x), f_i(y)) + d(f_i(y), f_i(s_i(\underline{y}))) \\ &\le \varepsilon_i + d(y, s_i(\underline{y})) + \varepsilon_i \\ &= 2\varepsilon_i. \end{aligned}$$

This proves item (1).

Conversely, given $\underline{f}_i$ an ε_i-approximation from $\overline{B}_{R_i}(\underline{x}_i)$ to $\overline{B}_{R_i}(\underline{x})$, we can define $f_i\colon \overline{B}_{R_i}(x_i) \to X$ as

$$f_i(x) = s(\underline{f}_i(\underline{x}))$$

for any $x \in X_i$. Then f_i is an ε_i-approximation from $\overline{B}_{R_i}(x_i)$ to $\overline{B}_{R_i}(x)$. Indeed,

$$|d(x, y) - d(f_i(x), f_i(y))| = |d(\underline{x}, \underline{y}) - d(\underline{f}_i(\underline{x}), \underline{f}_i(\underline{y}))| \le \varepsilon_i.$$

Moreover

$$d(f_i(x_i), x) = d(\underline{f}_i(\underline{x}_i), \underline{x}) \leq \varepsilon_i.$$

Finally, if $d(\underline{x}, x) \leq R_i$ then there exists $y \in X_i$ such that $d(\underline{y}, \underline{x}_i) \leq R_i$ and $d(\underline{x}, \underline{f}_i(\underline{y})) \leq \varepsilon_i$, or equivalently, $d(x, f_i(y)) \leq \varepsilon_i$. This proves item (2). □

Proposition 3.7.4 implies that, if we consider the following commutative diagram

$$\begin{array}{ccc} \mathsf{Met}_{\mathsf{Pair}} & \xrightarrow{\mathcal{D}_\infty} & \mathsf{PMet}_* \\ & \searrow^{\pi\circ\mathcal{D}_\infty} & \downarrow \pi \\ & & \mathsf{Met}_* \end{array}$$

where $\pi\colon \mathsf{PMet}_* \to \mathsf{Met}_*$ is the canonical metric identification functor, then $\mathcal{D}_\infty$ is continuous if and only if $\pi \circ \mathcal{D}_\infty$ is continuous.

Proposition 3.7.5. *The functor* $(X, A) \mapsto (\mathcal{D}_\infty(X, A), \sigma_A)$ *is sequentially continuous with respect to the Gromov–Hausdorff convergence.*

Proof. Let $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$, $R_i \nearrow \infty$, $\varepsilon_i \searrow 0$, and f_i be ε_i-approximations from $\overline{B}_{R_i}(A_i)$ to $\overline{B}_{R_i}(A)$. We can define a map $(f_i)_*\colon \overline{B}_{R_i}(\sigma_{A_i}) \to \mathcal{D}_\infty(X, A)$ as

$$(f_i)_*(\sigma) = \{\{f_i(x) : x \in \sigma \setminus A_i\}\}.$$

We will prove that $(f_i)_*$ is a $(3\varepsilon_i)$-approximation from $\overline{B}_{R_i}(\sigma_{A_i})$ to $\overline{B}_{R_i}(\sigma_A)$.

Let $\sigma, \sigma' \in \mathcal{D}_\infty(X_i, A_i)$. We now show that, for any bijection $\phi\colon \sigma \to \sigma'$, there exists a bijection $\phi_*\colon (f_i)_*(\sigma) \to (f_i)_*(\sigma')$ such that

$$\left| \sup_{x\in\sigma} d_{X_i}(x, \phi(x)) - \sup_{y\in(f_i)_*(\sigma)} d_X(y, \phi_*(y)) \right| \leq 3\varepsilon_i, \tag{3.16}$$

and, conversely, that for any bijection $\phi_*\colon (f_i)_*(\sigma) \to (f_i)_*(\sigma')$, there exists a bijection $\phi\colon \sigma \to \sigma'$ such that inequality (3.16) holds.

Indeed, let $\phi\colon \sigma \to \sigma'$ be a bijection, and let $x \in \sigma$ and $x' \in \sigma'$ be such that $\phi(x) = x'$. We set $\phi_*(\widehat{x}) = \widehat{x'}$, where, given any $z \in X_i$, we set $\widehat{z} = f_i(z)$ if $z \notin A_i$, and we set $\widehat{z} \in A$ to be a point such that $d_X(f_i(z), \widehat{z}) \leq \varepsilon_i$ if $z \in A_i$. In the latter case, such a choice is possible by item (2) in Definition 3.7.1. In particular, in either case we have $d_X(f_i(z), \widehat{z}) \leq \varepsilon_i$. Up to changing representatives of $(f_i)_*(\sigma)$ and $(f_i)_*(\sigma')$ in $\mathcal{D}_\infty(X, A)$, this completely defines a bijection $\phi_*\colon (f_i)_*(\sigma) \to (f_i)_*(\sigma')$, and we have

$$\begin{aligned}\left|d_{X_i}(x, x') - d_X(\widehat{x}, \widehat{x'})\right| &\leq |d_{X_i}(x, x') - d_X(f_i(x), f_i(x'))| \\ &\quad + |d_X(f_i(x), f_i(x')) - d_X(\widehat{x}, f_i(x'))| \\ &\quad + \left|d_X(\widehat{x}, f_i(x')) - d_X(\widehat{x}, \widehat{x'})\right| \\ &\leq \varepsilon_i + d_X(f_i(x), \widehat{x}) + d_X(f_i(x'), \widehat{x'}) \\ &\leq 3\varepsilon_i\end{aligned}$$

by item (1) in Definition 3.7.1 and the triangle inequality. Taking the supremum over all $x \in \sigma$ yields inequality (3.16).

Conversely, let $\theta\colon (f_i)_*(\sigma) \to (f_i)_*(\sigma')$ be a bijection, and let $y \in (f_i)_*(\sigma)$ and $y' \in (f_i)_*(\sigma')$ be such that $\theta(y) = y'$. We define a bijection $\breve{\theta}\colon \sigma \to \sigma'$ by setting $\breve{\theta}(\breve{y}) = \breve{y}'$, where, given any $z \in X$ (viewed as an element in the multiset $(f_i)_*(\sigma)$ or $(f_i)_*(\sigma')$), we set $\breve{z} \in X_i$ to be such that $f_i(\breve{z}) = z$ if z is defined as $f_i(x)$ for some $x \in X_i$, and such that $\breve{z} \in A_i$ and $d_X(f_i(\breve{z}), z) \leq \varepsilon_i$ otherwise. In the latter case, we must have $z \in A$ and hence such a choice is possible by item (2) in Definition 3.7.1. Similarly as above, we can then show that $\left|d_{X_i}(\breve{y}, \breve{y}') - d_X(y, y')\right| \leq 3\varepsilon_i$, and hence (3.16) holds with $\phi = \breve{\theta}$ and $\phi_* = \theta$.

Therefore, for any $\sigma, \sigma' \in \overline{B}_{R_i}(\sigma_{A_i})$, we have

$$\begin{aligned}|d_\infty(\sigma, \sigma') - d_\infty(f_i(\sigma), f_i(\sigma'))| &= \left|\inf_\phi \sup_{x \in \sigma}\{d(x, \phi(x))\} - \inf_\theta \sup_{y \in (f_i)_*(\sigma)}\{d(y, \theta(y))\}\right| \\ &\leq 3\varepsilon_i.\end{aligned}$$

On the other hand, by definition, we have that

$$d_\infty(f_i(\sigma_{A_i}), \sigma_A) = d_\infty(\sigma_A, \sigma_A) = 0 \leq 3\varepsilon_i.$$

Finally, if $d_\infty(\sigma, \sigma_A) \leq R_i$, then $d(y, A) \leq R_i$ for any $y \in \sigma$, and since f_i is an ε_i-approximation from $\overline{B}_{R_i}(A_i)$ to $\overline{B}_{R_i}(A)$, we know that there is some $x_y \in \overline{B}_{R_i}(A_i)$

such that $d(y, f_i(x_y)) \le \varepsilon_i$. Hence, the diagram $\hat{\sigma} \in \mathcal{D}_\infty(X_i, A_i)$ given by

$$\hat{\sigma} = \{\{x_y : x \in \sigma\}\}$$

satisfies $d_\infty(\sigma, (f_i)_*(\hat{\sigma})) \le \varepsilon_i \le 3\varepsilon_i$ and $d_\infty(\hat{\sigma}, \sigma_{A_i}) \le R_i$, so we conclude that $\overline{B}_{R_i}(\sigma_A) \subset \overline{B}_{3\varepsilon_i}(\overline{B}_{R_i}(\sigma_{A_i}))$.

Thus, $(f_i)_*$ is a $3\varepsilon_i$-approximation from $\overline{B}_{R_i}(\sigma_{A_i})$ to $\overline{B}_{R_i}(\sigma_A)$. □

Remark 3.7.6. Note that we have only shown that $\mathcal{D}_\infty$ is sequentially continuous. To show continuity, we must prove that the Gromov–Hausdorff convergence in $\mathsf{Met}_{\mathsf{Pair}}$ is metrisable. This is done for proper metric spaces in Proposition A.1.29 of the Appendix, which yields the proof of Theorem C.

Chapter **4**

Optimal partial transport for metric pairs

4.1 Introduction

Optimal transport provides a geometric way to compare probability measures, or more generally, up to re-scaling, measures with the same finite total mass. However, in different contexts it is natural to consider measures with different total masses and try to compare them (see, for example, [14, 19, 26, 27, 31, 38, 49] and references therein). In [27], Figalli and Gigli introduced a version of the optimal transport problem for non-negative Radon measures on bounded domains in $\mathbb{R}^n$, which we refer to as *optimal partial transport*, motivated by finding solutions to evolution equations with constant Dirichlet boundary conditions, in analogy to the Jordan-Kinderlehrer-Otto scheme [35]. Recently, Divol and Lacombe [24] established a connection between optimal partial transport and spaces of persistence diagrams. In this chapter we study optimal partial transport in the setting of proper metric pairs, and we carry over basic results from classical optimal transport to this setting.

In section 4.2 we define the spaces $(\mathcal{M}_p(X, A), \mathrm{Wb}_p)$ of Radon measures on proper metric pairs endowed with the L^p-optimal partial transport metric, and prove the existence of optimal partial transport plans, yielding Theorem E. In section

4.3 we prove criteria for optimality of partial transport plans. In sections 4.4, 4.5, and 4.6 we prove that the map $(X, A) \mapsto \mathcal{M}_p(X, A)$ preserves completeness, separability, and geodesicity. Moreover, we prove that for $p > 1$ it also preserves the non-branching property. For $p = 2$, we prove it preserves non-negative curvature in the Alexandrov sense, yielding Theorem F. Finally, in section 4.7, we prove that there is an isometric embedding of generalised spaces of persistence diagrams into spaces of optimal partial transport, proving Theorem G.

4.2 Optimal partial transport for metric pairs

We consider metric pairs (X, A) as in Definition 2.1.2. Moreover, throughout this chapter, we assume that X is proper, which in particular implies that it is complete, separable, and locally compact.

Given a metric pair (X, A) and $p \in [1, \infty)$, we define

$$\mathcal{M}_p(X, A) = \left\{ \mu \in \mathcal{M}(\Omega) : \int_\Omega d(x, A)^p \, d\mu(x) < \infty \right\},$$

where $\Omega = X \setminus A$, and $\mathcal{M}(\Omega)$ is the set of Radon measures on Ω, as in Definition 2.4.12. For any $\mu, \nu \in \mathcal{M}_p(X, A)$, we set the *$L^p$-optimal partial transport metric*

$$\mathrm{Wb}_p(\mu, \nu) = \inf_{\gamma \in \mathrm{Adm}(\mu,\nu)} \left(\int_{E_\Omega} d(x, y)^p \, d\gamma(x, y) \right)^{1/p} \tag{4.1}$$

where $\mathrm{Adm}(\mu, \nu)$ is the set of *partial transport plans* between μ and ν, i.e. the set of $\gamma \in \mathcal{B}(E_\Omega)$, where $E_\Omega = X \times X \setminus A \times A$, such that

$$\pi^1_\# \gamma|_\Omega = \mu, \quad \pi^2_\# \gamma|_\Omega = \nu, \tag{4.2}$$

where $\pi^1, \pi^2 \colon X \times X \to X$ are the projections onto the first and second factor, respectively. We prove in Theorem 4.2.10 that Wb_p is a metric on $\mathcal{M}_p(X, A)$. Observe that this is an adaptation of [27, Problem 1.1] for metric spaces.

Remark 4.2.1. Observe that condition (4.2) implies that $\gamma \in \mathcal{M}(E_\Omega)$. Indeed, for any compact $K \subset E_\Omega$ it is easy to see that there are compact sets $K', K'' \subset \Omega$ such that $K \subset K' \times X \cup X \times K''$, therefore

$$\gamma(K) \leq \gamma(K' \times X) + \gamma(X \times K'') = \mu(K') + \nu(K'') < \infty, \tag{4.3}$$

since $\mu, \nu \in \mathcal{M}(\Omega)$. Moreover, since E_Ω is an open subset of the separable and locally compact metric space $X \times X$, it is separable and locally compact itself, and the claim follows from Remark 2.4.13. Furthermore, (4.3) proves that $\mathrm{Adm}(\mu, \nu)$ is a vaguely relatively compact subset of $\mathcal{M}(E_\Omega)$, due to Lemma 2.4.15.

Remark 4.2.2. Given $\gamma \in \mathrm{Adm}(\mu, \nu)$, we define

$$C(\gamma) = \int_{E_\Omega} d(x,y)^p \; d\gamma(x,y)$$

and denote $\gamma_R^S = \gamma|_{R \times S}$, for any $R, S \subset X$ such that $R \times S \subset E_\Omega$. In particular,

$$\gamma = \gamma_\Omega^\Omega + \gamma_\Omega^A + \gamma_A^\Omega.$$

Remark 4.2.3. Regarding the terminology of *partial* transport plans, this comes from the fact that, whenever γ satisfies (4.2), the measure γ_Ω^Ω can be regarded as a classical transport plan (in the sense of Definition 2.4.4) between its marginals $\widetilde{\mu} = \pi^1_\#(\gamma_\Omega^\Omega)$ and $\widetilde{\nu} = \pi^2_\#(\gamma_\Omega^\Omega)$, which satisfy $\widetilde{\mu} \leq \mu$ and $\widetilde{\nu} \leq \nu$ and $\widetilde{\mu}(\Omega) = \widetilde{\nu}(\Omega) = \gamma(\Omega \times \Omega)$. Therefore γ_Ω^Ω effectively transports part of μ into part of ν.

As a consequence of the Borel measurable selection principle (Theorem 2.4.3), we obtain the following lemma, which will be useful in the sequel.

Lemma 4.2.4. *Let (X, A) be a metric pair. Then there exists a Borel measurable map $\mathrm{proj}_A \colon X \to A$ such that $d(x, \mathrm{proj}_A(x)) = d(x, A)$ for all $x \in X$.*

Proof. Since A is closed and X is complete and separable, it follows that A is complete and separable endowed with the restricted metric. Moreover, since X is proper and d is continuous, the set

$$E = \{(x, y) \in X \times A : d(x, y) = d(x, A)\}$$

is σ-compact, closed, and $\pi^1(E) = X$. Therefore, by Theorem 2.4.3, the claim follows. □

Remark 4.2.5. Observe that $\mathrm{Wb}_p(\mu,\nu)$ is well-defined, non-negative, and finite. Indeed, let $\mu,\nu \in \mathcal{M}_p(X,A)$. Then a straightforward computation shows that the measure

$$\gamma = (\mathrm{id}, \mathrm{proj}_A)_\#\mu + (\mathrm{proj}_A, \mathrm{id})_\#\nu$$

is in $\mathrm{Adm}(\mu,\nu)$, and it satisfies

$$0 \leq C(\gamma) = \int_\Omega d(x,A)^p \; d\mu(x) + \int_\Omega d(y,A)^p \; d\nu(y) < \infty.$$

Furthermore, the zero measure belongs to $\mathcal{M}_p(X,A)$, and due to Lemma 4.2.4, for any $\mu \in \mathcal{M}_p(X,A)$ we have

$$\mathrm{Wb}_p^p(\mu,0) = \int_X d(x,A)^p \; d\mu(x). \tag{4.4}$$

Indeed, the partial transport plan $\gamma = (\mathrm{id}, \mathrm{proj}_A)_\#\mu \in \mathrm{Adm}(\mu,0)$ satisfies

$$C(\gamma) = \int_\Omega d(x,\mathrm{proj}_A(x))^p \; d\mu(x) = \int_\Omega d(x,A)^p \; d\mu(x).$$

On the other hand, for any $\widetilde{\gamma} \in \mathrm{Adm}(\mu,0)$ we have that $\pi^2_\#\widetilde{\gamma}|_\Omega = 0$, which implies that $\widetilde{\gamma}(X\times\Omega) = 0$. Therefore we have

$$\begin{aligned} C(\widetilde{\gamma}) &= \int_{E_\Omega} d(x,y)^p \; d\widetilde{\gamma}(x,y) \\ &= \int_{\Omega\times A} d(x,y)^p \; d\widetilde{\gamma}(x,y) \\ &\geq \int_{\Omega\times A} d(x,A)^p \; d\widetilde{\gamma}(x,y) \\ &= \int_\Omega d(x,A)^p \; d\mu(x). \end{aligned}$$

This proves equation (4.4).

The following lemma is a natural generalisation of the gluing lemma (Theorem 2.4.7), and the proof is an adaptation of that of [27, Lemma 2.1]. We include it for the sake of completeness.

Lemma 4.2.6. *Let $\mu^1,\mu^2,\mu^3 \in \mathcal{M}_p(X,A)$, and consider $\gamma^{12} \in \mathrm{Adm}(\mu^1,\mu^2)$, and $\gamma^{23} \in \mathrm{Adm}(\mu^2,\mu^3)$. Then there exists $\gamma^{123} \in \mathcal{B}(X\times X\times X)$ such that*

$$\pi^{12}_\#\gamma^{123} = \gamma^{12} + \sigma^{12},$$

$$\pi^{23}_\#\gamma^{123} = \gamma^{23} + \sigma^{23},$$

where $\sigma^{12}, \sigma^{23} \in \mathcal{B}(X \times X)$ *are supported on* $\Delta(A \times A)$*.*

Proof. From the hypothesis, we can see that

$$\mu^2 = \pi^2_\# \gamma^{12}|_\Omega = \pi^2_\#((\gamma^{12})^\Omega_X + (\gamma^{12})^A_\Omega)|_\Omega = \pi^2_\#(\gamma^{12})^\Omega_X.$$

Analogously, $\mu^2 = \pi^1_\#(\gamma^{23})^X_\Omega$. Therefore, by applying Theorem 2.4.7, since μ^2 is a Radon measure, we can find a measure $\widetilde{\gamma}^{123}$ on $X \times \Omega \times X$ such that

$$\begin{aligned} \pi^{12}_\# \widetilde{\gamma}^{123} &= (\gamma^{12})^\Omega_X, \\ \pi^{23}_\# \widetilde{\gamma}^{123} &= (\gamma^{23})^X_\Omega. \end{aligned}$$

We define

$$\begin{aligned} \widetilde{\sigma}^{12} &= (\pi^1, \pi^1, \pi^2)_\#(\gamma^{23})^\Omega_A, \\ \widetilde{\sigma}^{23} &= (\pi^1, \pi^2, \pi^2)_\#(\gamma^{12})^A_\Omega, \\ \gamma^{123} &= \widetilde{\gamma}^{123} + \widetilde{\sigma}^{12} + \widetilde{\sigma}^{23}, \\ \sigma^{12} &= (\pi^1, \pi^1)_\#(\gamma^{23})^\Omega_A, \\ \sigma^{23} &= (\pi^2, \pi^2)_\#(\gamma^{12})^A_\Omega. \end{aligned}$$

We can check that these measures work. Indeed,

$$\begin{aligned} \pi^{12}_\# \gamma^{123} &= \pi^{12}_\# \widetilde{\gamma}^{123} + \pi^{12}_\# \widetilde{\sigma}^{12} + \pi^{12}_\# \widetilde{\sigma}^{23} \\ &= (\gamma^{12})^\Omega_X + \pi^{12}_\#(\pi^1, \pi^1, \pi^2)_\#(\gamma^{23})^\Omega_A + \pi^{12}_\#(\pi^1, \pi^2, \pi^2)_\#(\gamma^{12})^A_\Omega \\ &= (\gamma^{12})^\Omega_X + (\pi^1, \pi^1)_\#(\gamma^{23})^\Omega_A + (\pi^1, \pi^2)_\#(\gamma^{12})^A_\Omega \\ &= (\gamma^{12})^\Omega_X + \sigma^{12} + (\gamma^{12})^A_\Omega \\ &= \gamma^{12} + \sigma^{12}. \end{aligned}$$

Similarly with $\pi^{23}_\# \gamma^{123}$. □

The following result guarantees the existence of optimal partial transport plans in the setting of proper metric spaces, generalising [27, p. 4] and [24, Proposition 3.2].

Theorem 4.2.7. *Let (X, A) be a metric pair and $p \in [1, \infty)$. Then, for any $\mu, \nu \in \mathcal{M}_p(X, A)$, the set*

$$\mathrm{Opt}(\mu, \nu) = \{\gamma \in \mathrm{Adm}(\mu, \nu) : C(\gamma) = \mathrm{Wb}_p^p(\mu, \nu)\}$$

is non-empty. Moreover, the set $\mathrm{Opt}(\mu, \nu)$ *is vaguely compact.*

We will need the following technical lemma for the proof of Theorem 4.2.7, and later on, for the proof of Theorem 4.2.10. The proof of this lemma follows ideas from [2, Footnotes in Sections 2.1 and 2.2].

Lemma 4.2.8. *Let $\mu, \nu \in \mathcal{M}_p(X, A)$, $\varepsilon > 0$, and let $\mathrm{Opt}_\varepsilon(\mu, \nu)$ be the set of $\gamma \in \mathrm{Adm}(\mu, \nu)$ such that*

$$\mathrm{Wb}_p^p(\mu, \nu) \le C(\gamma) \le \mathrm{Wb}_p^p(\mu, \nu) + \varepsilon. \tag{4.5}$$

Then, for any compact $C \subset \Omega$, the set $\{\gamma_C^X : \gamma \in \mathrm{Opt}_\varepsilon(\mu, \nu)\}$ is weakly relatively compact.

Proof. Take $p_0 \in X$ and $0 < r < R$ such that $C \subset \overline{B}_r(p_0) \subset \overline{B}_R(p_0)$. By Hölder's inequality, for any $\gamma \in \mathrm{Adm}(\mu, \nu)$, we have

$$\left| \int_{C \times (X \setminus \overline{B}_R(p_0))} d(y, p_0) - d(x, p_0) \; d\gamma(x, y) \right|^p \le \gamma(C \times (X \setminus \overline{B}_R(p_0)))^{p-1} \int_{C \times (X \setminus \overline{B}_R(p_0))} d(x, y)^p \; d\gamma(x, y), \tag{4.6}$$

whereas

$$\left| \int_{C \times (X \setminus \overline{B}_R(p_0))} d(y, p_0) - d(x, p_0) \; d\gamma(x, y) \right|^p \ge (R - r)^p \gamma(C \times (X \setminus \overline{B}_R(p_0)))^p \tag{4.7}$$

since $\mathrm{dist}(\overline{B}_r(p_0), X \setminus \overline{B}_R(p_0)) \ge R - r$. Combining (4.5), (4.6), and (4.7), we get

$$\gamma(C \times (X \setminus \overline{B}_R(p_0))) \le \frac{\mathrm{Wb}_p^p(\mu, \nu) + \varepsilon}{(R - r)^p},$$

for any $\gamma \in \mathrm{Opt}_\varepsilon(\mu, \nu)$, which can be made arbitrarily small by fixing r and letting R tend to infinity. This proves that $\{\gamma_C^X : \gamma \in \mathrm{Opt}_\varepsilon(\mu, \nu)\}$ is tight. Moreover, since

$\mathrm{Opt}_\varepsilon(\mu,\nu) \subset \mathrm{Adm}(\mu,\nu)$, each γ_C^X has total mass $\mu(C) < \infty$, therefore $\{\gamma_C^X : \gamma \in \mathrm{Opt}_\varepsilon(\mu,\nu)\}$ has uniformly bounded total variation. By Theorem 2.4.11, the claim follows. □

As a consequence of the previous lemma, we obtain that the sets $\mathrm{Opt}_\varepsilon(\mu,\nu)$ are closed with respect to the vague topology. The proof follows along the same lines of those of [24, Proposition 3.2] and arguments in [27, p. 4]. Observe that the chain of equations in (4.8) below is the same as one in the proof of [24, Proposition 3.2], but we need Lemma 4.2.8 to justify it, even in the Euclidean setting. This is because, even when $f \in \mathcal{C}_c(\Omega)$, the composition $f \circ \pi^1 \colon \Omega \times X \to \mathbb{R}$ is not of compact support when X is not compact.

Corollary 4.2.9. *Let $\mu,\nu \in \mathcal{M}_p(X,A)$, $\varepsilon > 0$, and let $\mathrm{Opt}_\varepsilon(\mu,\nu)$ be defined as in Lemma 4.2.8. Then $\mathrm{Opt}_\varepsilon(\mu,\nu)$ is vaguely closed.*

Proof. Let $\{\gamma_k\}_{k\in\mathbb{N}} \subset \mathrm{Opt}_\varepsilon(\mu,\nu)$ be such that $\gamma_k \xrightarrow{v} \gamma$ for some $\gamma \in \mathcal{M}(E_\Omega)$. We first observe that $\gamma \in \mathrm{Adm}(\mu,\nu)$. Indeed, if $f \in \mathcal{C}_c(\Omega)$ then, by applying Lemma 4.2.8 with $C = \mathrm{supp}(f)$, we can assume that, up to passing to a subsequence, $\left\{(\gamma_k)^X_{\mathrm{supp}(f)}\right\}_{k\in\mathbb{N}}$ is weakly convergent to $\gamma^X_{\mathrm{supp}(f)}$. Therefore,

$$\int_\Omega f \, d\pi^1_\# \gamma = \int_{\Omega\times X} f\circ\pi^1 \, d\gamma = \lim_{k\to\infty} \int_{\Omega\times X} f\circ\pi^1 \, d\gamma_k = \int_\Omega f \, d\mu, \tag{4.8}$$

where the second equality follows from the fact that $f \circ \pi^1 \in \mathcal{C}_b(\Omega\times X)$. This implies that $\pi^1_\#\gamma|_\Omega = \mu$, and analogously we obtain $\pi^2_\#\gamma|_\Omega = \nu$.

Now, we prove that γ satisfies (4.5). Indeed, by Lemma 2.4.16 applied to the sequence $\{d(\cdot,\cdot)^p\gamma_k\}_{k\in\mathbb{N}}$, which is vaguely convergent to $d(\cdot,\cdot)^p\gamma$, and any bounded open set $U \subset E_\Omega$, we get

$$C(\gamma|_U) \le \liminf_{k\to\infty} C((\gamma_k)|_U) \le \mathrm{Wb}_p^p(\mu,\nu) + \varepsilon.$$

By the monotone convergence theorem,

$$C(\gamma) \le \mathrm{Wb}_p^p(\mu,\nu) + \varepsilon,$$

and the claim follows. □

Proof of Theorem 4.2.7. Let $\mu, \nu \in \mathcal{M}_p(X, A)$ and, for every $k \in \mathbb{N}$, let $\gamma_k \in \mathrm{Opt}_{1/k}(\mu, \nu)$, that is $\gamma_k \in \mathrm{Adm}(\mu, \nu)$ and

$$\mathrm{Wb}_p^p(\mu, \nu) \leq C(\gamma_k) \leq \mathrm{Wb}_p^p(\mu, \nu) + 1/k. \tag{4.9}$$

Remark 4.2.1 yields that $\{\gamma_k\}_{k\in\mathbb{N}}$ is a vaguely relatively compact subset of $\mathcal{M}(E_\Omega)$. Consequently, up to passing to a subsequence, we can assume that there exists $\gamma \in \mathcal{M}(E_\Omega)$ such that $\gamma_k \xrightarrow{v} \gamma$. Corollary 4.2.9 implies that $\gamma \in \mathrm{Opt}_{1/k}(\mu, \nu)$ for any $k \in \mathbb{N}$, which means that

$$\mathrm{Wb}_p^p(\mu, \nu) \leq C(\gamma) \leq \mathrm{Wb}_p^p(\mu, \nu) + \frac{1}{k}$$

for all $k \in \mathbb{N}$. Therefore, $\gamma \in \mathrm{Opt}(\mu, \nu)$.

Regarding the second part of the theorem, observe that if $\{\gamma_k\}_{k\in\mathbb{N}} \subset \mathrm{Opt}(\mu, \nu)$ then, by the previous arguments, up to passing to a subsequence, $\gamma_k \xrightarrow{v} \gamma$ for some $\gamma \in \mathcal{M}(E_\Omega)$. The fact that $\gamma \in \mathrm{Opt}(\mu, \nu)$ also follows from the arguments above. □

We now prove that Wb_p is a metric and that it is vaguely lower semi-continuous. This statement, and its proof, are adaptations of [27, Theorem 2.2] to the setting of proper metric spaces.

Theorem 4.2.10. *Let (X, A) be a metric pair and $p \in [1, \infty)$. Then the function Wb_p is a metric on $\mathcal{M}_p(X, A)$. Moreover, Wb_p is lower semi-continuous with respect to the vague topology.*

Proof. It is clear that Wb_p is symmetric. Moreover, due to Theorem 4.2.7, $\mathrm{Wb}_p(\mu, \nu) = 0$ if and only if $C(\gamma) = 0$ for some $\gamma \in \mathrm{Adm}(\mu, \nu)$, which is equivalent to

$$\int_{E_\Omega} d(x, y)^p \, d\gamma(x, y) = 0.$$

However, this is the same as $\mathrm{supp}(\gamma) \subset \Delta(X \times X)$, which in turn is equivalent to $\pi^1_\# \gamma = \pi^2_\# \gamma$. The latter implies that $\mu = \nu$. Conversely, if $\mu = \nu$ then $\gamma = (\mathrm{id}, \mathrm{id})_\# \mu \in \mathrm{Adm}(\mu, \nu)$ satisfies $C(\gamma) = 0$.

For the triangle inequality, we need Lemma 4.2.6. Indeed, let $\mu^1, \mu^2, \mu^3 \in \mathcal{M}_p(X, A)$ and choose $\gamma^{12} \in \mathrm{Opt}(\mu^1, \mu^2)$ and $\gamma^{23} \in \mathrm{Opt}(\mu^2, \mu^3)$ (which we can do thanks to Theorem 4.2.7). By Lemma 4.2.6, there exists $\gamma^{123} \in \mathcal{B}(X \times X \times X)$ such that

$$\pi^{12}_{\#}\gamma^{123} = \gamma^{12} + \sigma^{12},$$
$$\pi^{23}_{\#}\gamma^{123} = \gamma^{23} + \sigma^{23},$$

with $\mathrm{supp}(\sigma^{12})$, $\mathrm{supp}(\sigma^{23}) \subset \Delta(A \times A)$. In particular, $\pi^1_{\#}\gamma^{123}|_\Omega = \mu^1$ and $\pi^3_{\#}\gamma^{123}|_\Omega = \mu^3$, therefore $\pi^{13}_{\#}\gamma^{123}|_{E_\Omega} \in \mathrm{Adm}(\mu^1, \mu^3)$. This implies

$$\begin{aligned}
\mathrm{Wb}_p(\mu^1, \mu^3) &\leq C(\pi^{13}_{\#}\gamma^{123}|_{E_\Omega})^{1/p} \\
&= \left\| d \circ \pi^{13} \right\|_{L^p(\gamma^{123})} \\
&\leq \left\| d \circ \pi^{12} + d \circ \pi^{23} \right\|_{L^p(\gamma^{123})} \\
&\leq \left\| d \circ \pi^{12} \right\|_{L^p(\gamma^{123})} + \left\| d \circ \pi^{23} \right\|_{L^p(\gamma^{123})} \\
&= \|d\|_{L^p(\gamma^{12}+\sigma^{12})} + \|d\|_{L^p(\gamma^{23}+\sigma^{23})} \\
&= C(\gamma^{12})^{1/p} + C(\gamma^{23})^{1/p} \\
&= \mathrm{Wb}_p(\mu^1, \mu^2) + \mathrm{Wb}_p(\mu^2, \mu^3).
\end{aligned}$$

To prove that Wb_p is lower semi-continuous with respect to the vague topology, let $\mu_n \xrightarrow{v} \mu$ and $\nu_n \xrightarrow{v} \nu$. If $\liminf_{n\to\infty} \mathrm{Wb}_p(\mu_n, \nu_n) = \infty$, the result follows trivially. Otherwise, up to passing to a subsequence, we can assume that

$$\liminf_{n\to\infty} \mathrm{Wb}_p(\mu_n, \nu_n) = \lim_{n\to\infty} \mathrm{Wb}_p(\mu_n, \nu_n) < \infty.$$

For each $n \in \mathbb{N}$, take $\gamma_n \in \mathrm{Opt}(\mu_n, \nu_n)$. By similar arguments to those in the proofs of Lemma 4.2.8 and Corollary 4.2.9, up to passing to another subsequence, we can assume that $\gamma_n \xrightarrow{v} \gamma$ for some $\gamma \in \mathrm{Adm}(\mu, \nu)$. Lemma 2.4.16 applied to the measures $\{d(\cdot,\cdot)^p\gamma_n\}_{n\in\mathbb{N}}$, which vaguely converge to $d(\cdot,\cdot)^p\gamma$, and bounded open sets $U \subset E_\Omega$, yields

$$C(\gamma|_U) \leq \liminf_{n\to\infty} C((\gamma_n)|_U) \leq \lim_{n\to\infty} \mathrm{Wb}_p^p(\mu_n, \nu_n),$$

and by the monotone convergence theorem, the claim follows. □

Remark 4.2.11. Observe that Theorems 4.2.7 and 4.2.10 yield Theorem E.

4.3 Criteria for optimal partial transport plans

We now move on to study a characterisation of optimal partial transport plans analogous to Theorem 2.4.9. Let us define

$$c(x,y) = d(x,y)^p, \tag{4.10}$$

$$\widetilde{c}(x,y) = \min\{d(x,y)^p, d(x,A)^p + d(y,A)^p\}, \tag{4.11}$$

and

$$\mathcal{S} = \{(x,y) \in X \times X : \widetilde{c}(x,y) = c(x,y)\}. \tag{4.12}$$

The following theorem and its proof are adaptations of [27, Proposition 2.3] for proper metric spaces, included in the thesis for the sake of completeness.

Theorem 4.3.1. *Let $\gamma \in \mathcal{M}(E_\Omega)$ be a measure satisfying*

$$\int_{E_\Omega} d(x,A)^p + d(y,A)^p \; d\gamma(x,y) < \infty.$$

Then the following are equivalent:

1. $\gamma \in \mathrm{Opt}(\pi^1_\# \gamma|_\Omega, \pi^2_\# \gamma|_\Omega)$.
2. *γ is concentrated on $\mathcal{S}$ and the set $\mathrm{supp}(\gamma) \cup A \times A$ is $\widetilde{c}$-cyclically monotone.*
3. *there is a c-concave function ϕ such that both ϕ and ϕ^c are identically 0 on A and $\mathrm{supp}(\gamma) \subset \partial^c_+ \phi$.*

Moreover, $d(x,y) = d(x,A)$ for γ^A_Ω-a.e. (x,y) and $d(x,y) = d(y,A)$ for γ^Ω_A-a.e. (x,y), whenever γ is optimal.

Proof. We start proving that (1) implies (2). We denote $\overline{\mu} = \pi^1_\# \gamma$, $\overline{\nu} = \pi^2_\# \gamma$, $\mu = \overline{\mu}|_\Omega$ and $\nu = \overline{\nu}|_\Omega$, and define

$$\widetilde{\gamma} = \gamma|_{E_\Omega \cap \mathcal{S}} + (\pi^1, \mathrm{proj}_A \circ \pi^1)_\# \gamma|_{E_\Omega \setminus \mathcal{S}} + (\mathrm{proj}_A \circ \pi^2, \pi^2)_\# \gamma|_{E_\Omega \setminus \mathcal{S}}.$$

A straightforward computation shows that $\widetilde{\gamma} \in \mathrm{Adm}(\mu, \nu)$, and

$$C(\widetilde{\gamma}) = \int_{E_\Omega \cap \mathcal{S}} d(x,y)^p \; d\gamma(x,y) + \int_{E_\Omega \setminus \mathcal{S}} d(x,A)^p + d(y,A)^p \; d\gamma(x,y) \leq C(\gamma)$$

with strict inequality if and only if $\gamma(E_\Omega \setminus \mathcal{S}) > 0$. Since $\gamma \in \mathrm{Opt}(\mu, \nu)$, we get that $C(\widetilde{\gamma}) = C(\gamma)$, which implies that $\gamma(E_\Omega \setminus \mathcal{S}) = 0$, i.e. γ is concentrated on $\mathcal{S}$. In particular,

$$C(\gamma) = \int_{E_\Omega} \widetilde{c}(x,y) \; d\gamma(x,y),$$

where $\widetilde{c}$ is given by (4.11). Now suppose that a measure η on $X \times X$ satisfies $\pi^1_\# \eta = \overline{\mu}$ and $\pi^2_\# \eta = \overline{\nu}$. Then we can define $\widetilde{\eta}$ in analogy to how we defined $\widetilde{\gamma}$. Clearly $\widetilde{\eta} \in \mathrm{Adm}(\mu, \nu)$, and

$$C(\widetilde{\eta}) = \int_{E_\Omega} \widetilde{c}(x,y) \; d\eta(x,y).$$

In particular,

$$\int_{X \times X} \widetilde{c}(x,y) \; d\gamma(x,y) = C(\gamma) \leq C(\widetilde{\eta}) = \int_{X \times X} \widetilde{c}(x,y) \; d\eta(x,y)$$

for any η with the same marginals as γ. In other words, γ is an optimal plan between $\overline{\mu}$ and $\overline{\nu}$ with respect to the cost function $\widetilde{c}$ in the usual sense. Theorem 2.4.9 implies that $\mathrm{supp}(\gamma)$ is $\widetilde{c}$-cyclically monotone. Moreover, given $\{(x_i, y_i)\}_{i=1}^n \subset \mathrm{supp}(\gamma) \cup A \times A$ and $\sigma \in \Sigma_n$, we have

$$\begin{aligned}\sum_{i=1}^n \widetilde{c}(x_i, y_{\sigma(i)}) = & \sum_{\substack{(x_i,y_i) \in \mathrm{supp}(\gamma) \\ (x_{\sigma(i)}, y_{\sigma(i)}) \in \mathrm{supp}(\gamma)}} \widetilde{c}(x_i, y_{\sigma(i)}) + \sum_{\substack{(x_i,y_i) \in \mathrm{supp}(\gamma) \\ (x_{\sigma(i)}, y_{\sigma(i)}) \in A \times A \setminus \mathrm{supp}(\gamma)}} \widetilde{c}(x_i, y_{\sigma(i)}) \\ & + \sum_{\substack{(x_i,y_i) \in A \times A \setminus \mathrm{supp}(\gamma) \\ (x_{\sigma(i)}, y_{\sigma(i)}) \in \mathrm{supp}(\gamma)}} \widetilde{c}(x_i, y_{\sigma(i)}) + \sum_{\substack{(x_i,y_i) \in A \times A \setminus \mathrm{supp}(\gamma) \\ (x_{\sigma(i)}, y_{\sigma(i)}) \in A \times A \setminus \mathrm{supp}(\gamma)}} \widetilde{c}(x_i, y_{\sigma(i)})\end{aligned}$$

where the number of summands in the second and the third summations are the same, whereas the fourth summation vanishes since $\widetilde{c}(x,y) = 0$ for any $(x,y) \in A \times A$. Let $\{j_i\}_{i=1}^p$, $\{k_i\}_{i=1}^q$ and $\{l_i\}_{i=1}^q$ be the sets of indices for the first, second and third sums, respectively, and define a permutation $\widetilde{\sigma}$ of the indices $\{j_1, \ldots, j_p, k_1, \ldots, k_q\} = \{i : (x_i, y_i) \in \mathrm{supp}(\gamma)\}$ by $\widetilde{\sigma}(j_r) = \sigma(j_r)$ for $r = 1, \ldots, p$ and $\widetilde{\sigma}(k_s) = \sigma(l_s)$ for

$s = 1, \ldots, q$. Then

$$\begin{aligned}
\sum_{i=1}^{n} \widetilde{c}(x_i, y_{\sigma(i)}) &= \sum_{i=1}^{p} \widetilde{c}(x_{j_i}, y_{\sigma(j_i)}) + \sum_{i=1}^{q} \widetilde{c}(x_{k_i}, y_{\sigma(k_i)}) + \widetilde{c}(x_{l_i}, y_{\sigma(l_i)}) \\
&= \sum_{i=1}^{p} \widetilde{c}(x_{j_i}, y_{\sigma(j_i)}) + \sum_{i=1}^{q} \widetilde{c}(x_{k_i}, y_{\sigma(k_i)}) + \widetilde{c}(y_{\sigma(k_i)}, x_{l_i}) + \widetilde{c}(x_{l_i}, y_{\sigma(l_i)}) \\
&\geq \sum_{i=1}^{p} \widetilde{c}(x_{j_i}, y_{\sigma(j_i)}) + \sum_{i=1}^{q} \widetilde{c}(x_{k_i}, y_{\sigma(l_i)}) \\
&= \sum_{(x_i, y_i) \in \operatorname{supp}(\gamma)} \widetilde{c}(x_i, y_{\widetilde{\sigma}(i)}) \\
&\geq \sum_{(x_i, y_i) \in \operatorname{supp}(\gamma)} \widetilde{c}(x_i, y_i) \\
&= \sum_{i=1}^{n} \widetilde{c}(x_i, y_i)
\end{aligned}$$

where in the second and last equalities we have used the fact that $\widetilde{c}(x, y) = 0$ for any $(x, y) \in A \times A$, whereas for the first and second inequalities we have used the easily checked inequality $\widetilde{c}(x, y) + \widetilde{c}(z, y) \geq \widetilde{c}(x, z)$ that holds for any $x, z \in X$ and $y \in A$, and the fact that $\operatorname{supp}(\gamma)$ is $\widetilde{c}$-cyclically monotone. This shows (2).

Now, to prove that (2) implies (3), observe that, by Theorem 2.4.9, the $\widetilde{c}$-cyclical monotonicity of $\operatorname{supp}(\gamma) \cup A \times A$ implies that there exists a $\widetilde{c}$-concave function, say ϕ, such that $\operatorname{supp}(\gamma) \cup A \times A \subset \partial_{+}^{\widetilde{c}} \phi$. In particular,

$$\phi(x) + \phi^{\widetilde{c}}(y) = \widetilde{c}(x, y) = 0$$

for any $(x, y) \in A \times A$, which implies that both ϕ and $\phi^{\widetilde{c}}$ are constant on A. Since the $\widetilde{c}$-concavity is invariant under addition of constants, we can assume that $\phi = \phi^{\widetilde{c}} = 0$ on A.

On the other hand, we can see that ϕ is c-concave. Indeed, one can prove that, for any $y \in X$, the map $x \mapsto \widetilde{c}(x, y)$ is c-concave. Indeed, if we define $\psi_y \colon X \to \mathbb{R} \cup \{-\infty\}$ by

$$\psi_y(z) = \begin{cases} -\infty & \text{if } z \in X \setminus (A \cup \{y\}) \\ 0 & \text{if } z = y \\ -d(y, A)^p & \text{if } z \in A, \end{cases}$$

then it is clear that

$$\inf_{z\in X} c(x,z)-\psi_y(z)=\min\left\{d(x,y)^p, \inf_{z\in A} d(x,z)^p+d(y,A)^p\right\}=\widetilde{c}(x,y),$$

and since ϕ is $\widetilde{c}$-concave, there exists a function $\psi\colon X\to\mathbb{R}\cup\{-\infty\}$ such that $\phi=\psi^{\widetilde{c}}$, that is,

$$\begin{aligned}\phi(x)&=\inf_{y\in X}\widetilde{c}(x,y)-\psi(y)\\&=\inf_{y\in X}\inf_{z\in X} c(x,z)-\psi_y(z)-\psi(y)\\&=\inf_{z\in X}\inf_{y\in X} c(x,z)-\psi_y(z)-\psi(y)\\&=\inf_{z\in X} c(x,z)-\sup_{y\in X}\psi_y(z)+\psi(y)\\&=\inf_{z\in X} c(x,z)-\eta(z),\end{aligned}$$

where $\eta(z)=\sup_{y\in X}\psi_y(z)+\psi(y)$. Observe this is a well defined function $X\to\mathbb{R}\cup\{-\infty\}$ since, if $z\in A$ then

$$\sup_{y\in X}\psi_y(z)+\psi(y)=\sup_{y\in X}-d(y,A)^p+\psi(y)=-\inf_{y\in X}\widetilde{c}(z,y)-\psi(y)=-\phi(z)=0$$

and if $z\notin A$, then

$$\psi_y(z)+\psi(y)=\begin{cases}-\infty & \text{if } y\neq z\\ \psi(z) & \text{if } y=z\end{cases}$$

which implies that $\sup_{y\in X}\psi_y(z)+\psi(y)=\psi(z)$.

Moreover, if $(x,y)\in\partial_+^{\widetilde{c}}\phi\cap\mathcal{S}$ then

$$\phi^{\widetilde{c}}(y)=\widetilde{c}(x,y)-\phi(x)=c(x,y)-\phi(x)$$

whereas

$$\phi^{\widetilde{c}}(y)\leq\widetilde{c}(x',y)-\phi(x')\leq c(x',y)-\phi(x')$$

for any $x'\in X$, which means that $\phi^{\widetilde{c}}(y)=\phi^{c}(y)$, therefore $(x,y)\in\partial_+^{c}\phi$. In particular, since γ is concentrated on $\partial_+^{\widetilde{c}}\phi\cap\mathcal{S}$, we get that it is also concentrated on $\partial_+^{c}\phi$, which implies that $\operatorname{supp}(\gamma)\subset\partial_+^{c}\phi$.

On the other hand, if $x \in A$ then, since $\phi = \phi^{\widetilde{c}} = 0$ on A, we get

$$\phi(x) + \phi^{\widetilde{c}}(x) = \widetilde{c}(x,x) = c(x,x) = 0$$

which implies that $(x,x) \in \partial_+^{\widetilde{c}}\phi \cap \mathcal{S}$, therefore $(x,x) \in \partial_+^c\phi$. In particular,

$$\phi^c(x) = \phi(x) + \phi^c(x) = c(x,x) = 0$$

therefore $\phi^c = 0$ on A. This proves (3).

Finally, to prove that (3) implies (1), consider ϕ a c -concave function such that $\text{supp}(\gamma) \subset \partial_+^c\phi$ and $\phi = \phi^c = 0$ on A, and choose $\widetilde{\gamma} \in \text{Adm}(\mu,\nu)$. Then, since $\pi^1_\# \gamma|_\Omega = \pi^1_\# \widetilde{\gamma}|_\Omega$, we get

$$\int_X \phi \; d\pi^1_\# \gamma = \int_\Omega \phi \; d\pi^1_\# \gamma = \int_\Omega \phi \; d\pi^1_\# \widetilde{\gamma} = \int_X \phi \; d\pi^1_\# \widetilde{\gamma}.$$

Analogously with ϕ^c. Then, we get

$$\begin{aligned}
\int_{E_\Omega} d(x,y)^p \; d\gamma(x,y) &= \int_{E_\Omega} \phi(x) + \phi^c(y) \; d\gamma(x,y) \\
&= \int_X \phi(x) \; d\pi^1_\# \gamma(x) + \int_X \phi^c(y) \; d\pi^2_\# \gamma(y) \\
&= \int_X \phi(x) \; d\pi^1_\# \widetilde{\gamma}(x) + \int_X \phi^c(y) \; d\pi^2_\# \widetilde{\gamma}(y) \\
&= \int_{E_\Omega} \phi(x) + \phi^c(y) \; d\widetilde{\gamma}(x,y) \\
&\leq \int_{E_\Omega} d(x,y)^p \; d\widetilde{\gamma}(x,y)
\end{aligned}$$

where the last inequality is due to the inequality $\phi(x) + \phi^c(y) \leq c(x,y)$ that holds for general $(x,y) \in X \times X$. This argument implies that $\gamma \in \text{Opt}(\mu,\nu)$.

To prove the last part of the statement, we only need to observe that, whenever $\gamma \in \text{Opt}(\mu,\nu)$ for some $\mu,\nu \in \mathcal{M}_p(X,A)$, then by (2) we have

$$\text{supp}(\gamma_\Omega^A) \subset \text{supp}(\gamma) \cap \Omega \times A \subset \mathcal{S} \cap \Omega \times A$$

and for any $(x,y) \in \mathcal{S} \cap \Omega \times A$ we have

$$d(x,A)^p \leq d(x,y)^p = \min\{d(x,y)^p, d(x,A)^p\} \leq d(x,A)^p$$

which implies $d(x,y) = d(x,A)$ for any $(x,y) \in \text{supp}(\gamma_\Omega^A)$. Analogously with γ_A^Ω.

□

4.4 Completeness and separability

In this section we prove that $\mathcal{M}_p(X, A)$ inherits the properties of completeness and separability from the underlying space X. This statement and its proof are adaptations of [27, Proposition 2.7] for proper metric spaces.

Theorem 4.4.1. *The space $\mathcal{M}_p(X, A)$ is complete and separable.*

Proof. For the separability of $\mathcal{M}_p(X, A)$, if we choose a countable dense set $S \subset X$ and define,

$$F = \left\{ \sum_{i \in I} q_i \delta_{x_i} : x_i \in S \cap \Omega,\ q_i \in \mathbb{Q}_+,\ I \subset \mathbb{N} \text{ is finite} \right\},$$

it is easy to check that F is countable and dense in $\mathcal{M}_p(X, A)$.

To prove that $\mathcal{M}_p(X, A)$ is complete, we consider a Cauchy sequence $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{M}_p(X, A)$. Since $\{\mathrm{Wb}_p(\mu_n, 0)\}_{n \in \mathbb{N}}$ is a bounded sequence in $\mathbb{R}$, say by some $C > 0$, then for any compact $K \subset \Omega$ we have

$$\mu_n(K) \leq \frac{1}{r^p} \int_\Omega d(x, A)^p \, d\mu_n(x) = \frac{1}{r^p} \mathrm{Wb}_p^p(\mu_n, 0) \leq \frac{C^p}{r^p}$$

where $d(x, A) \geq r > 0$ for any $x \in K$. In particular

$$\sup\{\mu_n(K) : n \in \mathbb{N}\} < \infty$$

for any compact $K \subset \Omega$, which due to Lemma 2.4.15, implies that $\{\mu_n\}_{n \in \mathbb{N}}$ is vaguely precompact. Therefore it has a subsequence $\{\mu_{n_k}\}_{k \in \mathbb{N}}$ vaguely convergent to some μ. By the lower semi continuity of Wb_p, we get that

$$\mathrm{Wb}_p^p(\mu, 0) \leq \liminf_{k \to \infty} \mathrm{Wb}_p^p(\mu_{n_k}, 0) < \infty,$$

therefore $\mu \in \mathcal{M}_p(X, A)$. Moreover, for any $n \in \mathbb{N}$,

$$0 \leq \mathrm{Wb}_p(\mu_n, \mu) \leq \liminf_{k \to \infty} \mathrm{Wb}_p^p(\mu_n, \mu_{n_k}),$$

which implies

$$\lim_{n \to \infty} \mathrm{Wb}_p(\mu_n, \mu) = \lim_{n \to \infty} \liminf_{k \to \infty} \mathrm{Wb}_p^p(\mu_n, \mu_{n_k}) = 0$$

where the last equation comes from the fact that $\{\mu_n\}_{n\in\mathbb{N}}$ is Cauchy. Therefore $\mu_n \to \mu$ in $\mathcal{M}_p(X, A)$, which implies the completeness. □

Remark 4.4.2. Observe the notorious difference in the length of the proof of Theorem 3.4.1 and that of Theorem 4.4.1. This is due to the fact that in Theorem 3.4.1 we do not only prove that a Cauchy sequence of persistence diagrams converges to some measure (regarding persistence diagrams as Radon measures, as in Theorem 4.7.2 below), but that such limit measure is a persistence diagram itself. On the other hand, Theorems 3.4.1 and 4.7.2 yield that $\mathcal{D}_p(X, A)$ is a closed subset of $\mathcal{M}_p(X, A)$ for any $p \in [1, \infty)$.

Remark 4.4.3. Theorem 4.4.1 implies item 1 in Theorem F.

4.5 Geodesics

We now prove that $\mathcal{M}_p(X, A)$ is a geodesic space whenever X has this property. This is a generalisation of [27, Proposition 2.9] for proper metric spaces.

Theorem 4.5.1. *Let (X, A) be a metric pair such that X is geodesic. Then $\mathcal{M}_p(X, A)$ is geodesic as well. Furthermore, if $(\mu_t)_{t\in[0,1]}$ is a constant speed geodesic in $\mathcal{M}_p(X, A)$, then there exists a measure $\boldsymbol{\gamma} \in \mathcal{M}((e_0, e_1)^{-1}(E_\Omega))$ such that $(e_0, e_1)_\#\boldsymbol{\gamma} \in \mathrm{Opt}(\mu_0, \mu_1)$ and*

$$\mu_t = (e_t)_\#\boldsymbol{\gamma}|_\Omega$$

for any $t \in [0, 1]$.

For the proof of Theorem 4.5.1 we follow ideas from the proofs of [2, Theorem 2.10] and [27, Proposition 2.9].

First, we need the following technical lemma, which yields a measurable selection principle for geodesics.

Lemma 4.5.2. *There is a Borel measurable map* $\mathrm{GeoSel}\colon X \times X \to \mathrm{Geo}(X)$ *such that*

$$(e_0, e_1) \circ \mathrm{GeoSel} = \mathrm{id}_{X \times X}\,.$$

Proof. Since X and $\mathrm{Geo}(X)$ are proper metric spaces, the set

$$E = \{(x, y, \xi) \in X \times X \times \mathrm{Geo}(X) : \xi_0 = x,\ \xi_1 = y,\ d(x, y) = \mathcal{L}(\xi)\}$$

is σ-compact. Moreover, by the continuity of d, the lower semi-continuity of $\mathcal{L}$, the continuity of the evaluation maps e_0, e_1, and the fact that X is geodesic, it follows that E is closed and $\pi^{1,2}(E) = X \times X$, where $\pi^{1,2}\colon X \times X \times \mathrm{Geo}(X) \to X \times X$ is the projection onto the first two factors. Therefore, by Theorem 2.4.3, the claim follows. □

Remark 4.5.3. Observe that for any $(x, y) \in \mathcal{S}$, where $\mathcal{S}$ is given by (4.12), and any $t \in (0, 1)$ and $\xi \in \mathrm{Geo}(X)$ such that $(e_0, e_1)(\xi) = (x, y)$, we have $\xi_t \in \Omega$. Indeed, if this is not the case, then for some choice of $(x, y) \in \mathcal{S}$, $t \in (0, 1)$ and $\xi \in \mathrm{Geo}$ such that $(e_0, e_1)(\xi) = (x, y)$, we have $\xi_t \in A$. Therefore,

$$d(x, A)^p + d(y, A)^p \leq d(x, \xi_t)^p + d(y, \xi_t)^p = ((1 - t)^p + t^p) d(x, y)^p < d(x, y)^p$$

which contradicts the fact that $(x, y) \in \mathcal{S}$.

We also need the following lemma to prove the second part of Theorem 4.5.1. Intuitively speaking, this lemma allows us to construct a sequence of measures on $\mathrm{Geo}(X)$ that interpolate a given geodesic $(\mu_t)_{t \in [0,1]}$ in $\mathcal{M}_p(X, A)$ at arbitrarily fine dyadic rational parameters in $[0, 1]$.

Lemma 4.5.4. *Let* $(\mu_t)_{t \in [0,1]}$ *be a geodesic in* $\mathcal{M}_p(X, A)$. *Then, for any* $m \in \mathbb{N}$, *we can find* $\boldsymbol{\gamma}^m \in \mathcal{B}(\mathrm{Geo}(X))$ *such that*

$$(e_{j/2^m}, e_{k/2^m})_{\#} \boldsymbol{\gamma}^m|_{E_\Omega} \in \mathrm{Opt}(\mu_{j/2^m}, \mu_{k/2^m}),$$

$$\mathrm{supp}((e_{j/2^m}, e_{k/2^m})_{\#} \boldsymbol{\gamma}^m|_{A \times A}) \subset \Delta(A \times A),$$

for any $j, k \in \{0, \ldots, 2^m\}$.

Proof. For each $m \in \mathbb{N}$ and each $i \in \{0, \dots, 2^m - 1\}$, choose

$$\gamma^{i,m} \in \mathrm{Opt}(\mu_{i/2^m}, \mu_{(i+1)/2^m})$$

and apply Lemma 4.2.6 iteratively to get $\gamma^m \in \mathcal{B}(X^{2^m+1})$ such that

$$\pi^{i,i+1}_{\#}\gamma^m = \gamma^{i,m} + \sigma^{i,m},$$

where $\sigma^{i,m}$ is supported on $\Delta(A \times A)$. Then, by letting $G^m \colon X^{2^m+1} \to \mathcal{C}([0,1], X)$ be the Borel measurable map given by

$$(x_0, \dots, x_{2^m}) \mapsto \mathrm{GeoSel}(x_0, x_1) * \cdots * \mathrm{GeoSel}(x_{2^m}, x_{2^m+1}),$$

where $*$ denotes concatenation of paths, we can define

$$\boldsymbol{\gamma}^m = G^m_{\#}\gamma^m.$$

Now, for any $j, k \in \{0, \dots, 2^m\}$, we have

$$\begin{aligned}
\|d(e_{j/2^m}, e_{k/2^m})\|_{L^p(\boldsymbol{\gamma}^m)} &\leq \left\| \sum_{i=j}^{k-1} d(e_{i/2^m}, e_{(i+1)/2^m}) \right\|_{L^p(\boldsymbol{\gamma}^m)} \\
&\leq \sum_{i=j}^{k-1} \|d(e_{i/2^m}, e_{(i+1)/2^m})\|_{L^p(\boldsymbol{\gamma}^m)} \\
&= \sum_{i=j}^{k-1} \mathrm{Wb}_p(\mu_{i/2^m}, \mu_{(i+1)/2^m}) \\
&= \mathrm{Wb}_p(\mu_{j/2^m}, \mu_{k/2^m})
\end{aligned}$$

where the first inequality is given by the monotonicity of the integral; the second one is the triangle inequality in $L^p(\boldsymbol{\gamma}^m)$, and the last two lines are consequences of the definition of $\boldsymbol{\gamma}^m$ and the fact that $(\mu_t)_{t\in[0,1]}$ is a geodesic. As a consequence, we get that

$$(e_{j/2^m}, e_{k/2^m})_{\#}\boldsymbol{\gamma}^m|_{E_\Omega} \in \mathrm{Opt}(\mu_{j/2^m}, \mu_{k/2^m})$$

and

$$d(\xi_{j/2^m}, \xi_{k/2^m}) = \sum_{i=j}^{k-1} d(\xi_{i/2^m}, \xi_{(i+1)/2^m})$$

for $\boldsymbol{\gamma}^m$-a.e. $\xi \in \mathcal{C}([0,1], X)$, and the claim follows. □

The next results are similar in spirit to Lemma 4.2.8 and Corollary 4.2.9, and they will allow us to construct the limit measure we need for the second part of Theorem 4.5.1. To improve readability, let us define

$$\mathrm{OptGeo}(\mu,\nu) = \{\boldsymbol{\gamma} \in \mathcal{M}((e_0,e_1)^{-1}(E_\Omega)) : (e_0,e_1)_\# \boldsymbol{\gamma} \in \mathrm{Opt}(\mu,\nu)\}. \tag{4.13}$$

Lemma 4.5.5. *For any $\mu,\nu,\rho \in \mathcal{M}_p(X,A)$, any compact $C \subset \Omega$, and any $t \in [0,1]$, the set*

$$\left\{\boldsymbol{\gamma}|_{e_t^{-1}(C)} : \boldsymbol{\gamma} \in \mathrm{OptGeo}(\mu,\nu),\ (e_t)_\# \boldsymbol{\gamma}|_\Omega = \rho\right\} \tag{4.14}$$

is weakly relatively compact.

Proof. Let $p_0 \in X$ and $R > r > 0$ such that $C \subset \overline{B}_r(p_0) \subset \overline{B}_R(p_0)$. Then, an argument analogous to the proof of Lemma 4.2.8 shows that

$$\boldsymbol{\gamma}(e_t^{-1}(C) \setminus (e_0,e_1)^{-1}(\overline{B}_R(p_0) \times \overline{B}_R(p_0))) \le \frac{\mathrm{Wb}_p^p(\mu_0,\mu_1)}{(R-r)^p} \tag{4.15}$$

for any $\boldsymbol{\gamma} \in \mathrm{OptGeo}(\mu,\nu)$. Observe that $(e_0,e_1)^{-1}(\overline{B}_R(p_0) \times \overline{B}_R(p_0))$ is closed and, being the set of geodesics with endpoints in $\overline{B}_R(p_0)$, is contained in $\mathrm{Geo}(\overline{B}_{2R}(X))$, which is compact due to Arzelà–Ascoli theorem and the fact that X is proper. By fixing $r > 0$ and letting R tend to infinity, the right hand side of inequality (4.15) can be made arbitrarily small, uniformly over $\boldsymbol{\gamma}$, which implies tightness.

On the other hand, each $\boldsymbol{\gamma}|_{e_t^{-1}(C)}$ has total mass $\rho(C) < \infty$. Therefore we have uniformly bounded total variation, and the claim follows from Theorem 2.4.11. $\square$

Corollary 4.5.6. *For any $\mu,\nu,\rho \in \mathcal{M}_p(X,A)$ and any $t \in [0,1]$, the set*

$$\{\boldsymbol{\gamma} : \boldsymbol{\gamma} \in \mathrm{OptGeo}(\mu,\nu),\ (e_t)_\# \boldsymbol{\gamma}|_\Omega = \rho\} \tag{4.16}$$

is vaguely closed in $\mathcal{M}((e_0,e_1)^{-1}(E_\Omega))$.

Proof. Let $\{\boldsymbol{\gamma}^m\}_{m\in\mathbb{N}} \subset \mathrm{OptGeo}(\mu,\nu)$ be such that $(e_t)_\# \boldsymbol{\gamma}^m|_\Omega = \rho$ for all $m \in \mathbb{N}$, and assume that $\boldsymbol{\gamma}^m \xrightarrow{v} \boldsymbol{\gamma}$ for some $\boldsymbol{\gamma} \in \mathcal{M}((e_0,e_1)^{-1}(E_\Omega))$. Thanks to Lemma 4.5.5, for any compact $C \subset \Omega$, $\{\boldsymbol{\gamma}^m|_{e_t^{-1}(C)}\}_{m\in\mathbb{N}}$ is weakly relatively compact, therefore $\boldsymbol{\gamma}^m|_{e_t^{-1}(C)} \xrightarrow{w} \boldsymbol{\gamma}|_{e_t^{-1}(C)}$. In particular, if $f \in \mathcal{C}_c(\Omega)$ then

$$\begin{aligned}\int_\Omega f \, d(e_t)_\# \boldsymbol{\gamma} &= \int_{e_t^{-1}(\mathrm{supp}(f))} f \circ e_t \, d\boldsymbol{\gamma} \\ &= \lim_{m\to\infty} \int_{e_t^{-1}(\mathrm{supp}(f))} f \circ e_t \, d\boldsymbol{\gamma}^m \\ &= \lim_{m\to\infty} \int_{\mathrm{supp}(f)} f \, d(e_t)_\# \boldsymbol{\gamma}^m \\ &= \int_\Omega f \, d\mu_t,\end{aligned}$$

which implies $(e_t)_\# \boldsymbol{\gamma}|_\Omega = \rho$. Analogously, $(e_0, e_1)_\# \boldsymbol{\gamma} \in \mathrm{Adm}(\mu, \nu)$.

Now, we prove that $\boldsymbol{\gamma} \in \mathrm{OptGeo}(\mu, \nu)$. Indeed, by Lemma 2.4.16 applied to the sequence $\{d(e_0, e_1)^p \boldsymbol{\gamma}^m\}_{m\in\mathbb{N}}$, which is vaguely convergent to $d(e_0, e_1)^p \boldsymbol{\gamma}$, and any bounded open set $U \subset (e_0, e_1)^{-1}(E_\Omega)$, we get

$$C((e_0, e_1)_\# \boldsymbol{\gamma}|_U) \le \liminf_{m\to\infty} C((e_0, e_1)_\# \boldsymbol{\gamma}^m|_U) \le \mathrm{Wb}_p^p(\mu, \nu).$$

By the monotone convergence theorem, the claim follows. □

Proof of Theorem 4.5.1. Let $\mu^0, \mu^1 \in \mathcal{M}_p(X, A)$ and choose $\gamma \in \mathrm{Opt}(\mu^0, \mu^1)$. By Lemma 4.5.2 there is a Borel measurable map $\mathrm{GeoSel}\colon X \times X \to \mathrm{Geo}(X)$ such that $\mathrm{GeoSel}(x, y)$ is a constant speed geodesic joining x and y. We define $\boldsymbol{\gamma} \in \mathcal{B}(\mathrm{Geo}(X))$ by

$$\boldsymbol{\gamma} = \mathrm{GeoSel}_\# \, \gamma.$$

In particular, since γ is concentrated on $\mathcal{S}$ by item 2 in Theorem 4.3.1, then $\boldsymbol{\gamma}$ is concentrated on $(e_0, e_1)^{-1}(\mathcal{S})$, which implies that $(e_t)_\# \boldsymbol{\gamma}$ is concentrated on Ω, for any $t \in (0, 1)$, by Remark 4.5.3. Observe, however, that this is not necessarily the case for $t = 0$ and $t = 1$. Let μ_t be given by

$$\mu_t = (e_t)_\# \boldsymbol{\gamma}|_\Omega.$$

We claim that the curve $(\mu_t)_{t\in[0,1]}$ is a constant speed geodesic joining μ^0 and μ^1 in $\mathcal{M}_p(X, A)$. Indeed, since $(e_0, e_t) \circ \mathrm{GeoSel} = \mathrm{id}$ then

$$\mu^0 = \pi^1_\# \gamma|_\Omega = \pi^1_\# (e_0, e_1)_\# \, \mathrm{GeoSel}_\# \, \gamma|_\Omega = (e_0)_\# \boldsymbol{\gamma}|_\Omega = \mu_0$$

and, analogously, $\mu^1 = \mu_1$. Moreover, for any $s, t \in [0,1]$,

$$\begin{aligned}
\mathrm{Wb}_p^p(\mu_t, \mu_s) &= \mathrm{Wb}_p^p((e_t)_\# \boldsymbol{\gamma}|_\Omega, (e_s)_\# \boldsymbol{\gamma}|_\Omega) \\
&\leq C((e_t, e_s)_\# \boldsymbol{\gamma}|_{E_\Omega}) \\
&= \int_{E_\Omega} d(x,y)^p \, d(e_t, e_s)_\# \boldsymbol{\gamma}(x,y) \\
&= \int_{E_\Omega} d(\mathrm{GeoSel}(x,y)_t, \mathrm{GeoSel}(x,y)_s)^p \, d\gamma(x,y) \\
&= |t-s|^p \int_{E_\Omega} d(x,y)^p \, d\gamma(x,y) \\
&= |t-s|^p \mathrm{Wb}_p^p(\mu_0, \mu_1).
\end{aligned}$$

This argument implies both that $\mu_t \in \mathcal{M}_p(X,A)$ for any $t \in [0,1]$, by the triangle inequality, and that $(\mu_t)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{M}_p(X,A))$.

Now, for the second part of the theorem, let $(\mu_t)_{t\in[0,1]}$ be a geodesic in $\mathcal{M}_p(X,A)$. We want to construct $\boldsymbol{\gamma} \in \mathcal{M}((e_0,e_1)^{-1}(E_\Omega))$ such that $(e_0,e_1)_\# \boldsymbol{\gamma} \in \mathrm{Opt}(\mu_0,\mu_1)$ and $(e_t)_\# \boldsymbol{\gamma}|_\Omega = \mu_t$ for all $t \in [0,1]$. We will get such $\boldsymbol{\gamma}$ as a limit of a sequence of measures given by Lemma 4.5.4.

Indeed, for each $m \in \mathbb{N}$, let $\boldsymbol{\gamma}^m \in \mathcal{B}(\mathrm{Geo}(X))$ be as in Lemma 4.5.4. In particular, $(e_0,e_1)_\# \boldsymbol{\gamma}^m|_{E_\Omega} \in \mathrm{Opt}(\mu_0,\mu_1)$. Due to item 2 in Theorem 4.3.1, $\boldsymbol{\gamma}^m|_{(e_0,e_1)^{-1}(E_\Omega)}$ is concentrated on $(e_0,e_1)^{-1}(\mathcal{S} \cap E_\Omega)$. Additionally, Lemma 4.5.4 implies that $\boldsymbol{\gamma}^m|_{(e_0,e_1)^{-1}(A\times A)}$ is supported on constant geodesics. Therefore, without loss of generality, we can assume that $\boldsymbol{\gamma}^m = \boldsymbol{\gamma}^m|_{(e_0,e_1)^{-1}(E_\Omega)}$.

We now prove that $\{\boldsymbol{\gamma}^m\}_{m\in\mathbb{N}}$ is vaguely relatively compact in $\mathcal{M}((e_0,e_1)^{-1}(E_\Omega))$. Indeed, if $\mathcal{K} \subset (e_0,e_1)^{-1}(\mathcal{S} \cap E_\Omega)$ is a compact set, then $e_{1/2}(\mathcal{K}) \subset \Omega$ by Remark 4.5.3, and thanks to Lemma 4.5.4,

$$\boldsymbol{\gamma}^m(\mathcal{K}) \leq \boldsymbol{\gamma}^m((e_{1/2})^{-1}(e_{1/2}(\mathcal{K}))) = (e_{1/2})_\# \boldsymbol{\gamma}^m(e_{1/2}(\mathcal{K})) = \mu_{1/2}(e_{1/2}(\mathcal{K})).$$

Since $\mu_{1/2}$ is a Radon measure on Ω, and $e_{1/2}(\mathcal{K})$ is compact, we get that

$$\sup_{m\in\mathbb{N}} \boldsymbol{\gamma}^m(\mathcal{K}) < \infty,$$

which proves the claim, due to Remark 2.4.13 and Lemma 2.4.15, since $(e_0,e_1)^{-1}(E_\Omega)$ is an open subset of a separable, locally compact metric space.

Thus, we can assume, up to passing to a subsequence, that $\{\boldsymbol{\gamma}^m\}_{m\in\mathbb{N}}$ is vaguely convergent to some $\boldsymbol{\gamma} \in \mathcal{M}((e_0, e_1)^{-1}(E_\Omega))$. By Corollary 4.5.6, $(e_0, e_1)_\# \boldsymbol{\gamma} \in \mathrm{Opt}(\mu_0, \mu_1)$ and $(e_t)_\# \boldsymbol{\gamma}|_\Omega = \mu_t$ for any dyadic rational $t \in [0, 1]$.

Finally, for any other $t \in [0, 1]$, let $\{t_k\}_{k\in\mathbb{N}}$ and $\{t^l\}_{l\in\mathbb{N}}$ be two sequences of dyadic rational numbers converging to t, with $t_k \leq t \leq t^l$ for any $k, l \in \mathbb{N}$, and observe that $(e_{t_k}, e_t)_\# \boldsymbol{\gamma}|_{E_\Omega} \in \mathrm{Adm}(\mu_{t_k}, (e_t)_\# \boldsymbol{\gamma}|_\Omega)$. Therefore,

$$\begin{aligned}
\mathrm{Wb}_p^p(\mu_{t_k}, (e_t)_\# \boldsymbol{\gamma}|_\Omega) &\leq C((e_{t_k}, e_t)_\# \boldsymbol{\gamma}|_{E_\Omega}) \\
&\leq C((e_{t_k}, e_{t^l})_\# \boldsymbol{\gamma}|_{E_\Omega}) \\
&= \mathrm{Wb}_p^p(\mu_{t_k}, \mu_{t^l}).
\end{aligned}$$

By letting $k, l \to \infty$, we get that $(e_t)_\# \boldsymbol{\gamma}|_\Omega = \mu_t$ as claimed. □

We now prove that $\mathcal{M}_p(X, A)$ inherits the property of being non-branching, whenever $p > 1$. This is analogous to the second half of [27, Proposition 2.9], and the proof adapts ideas from [2, Proposition 2.16].

Theorem 4.5.7. *Let (X, A) be a metric pair such that X is geodesic and non-branching, and $p \in (1, \infty)$. Then $\mathcal{M}_p(X, A)$ is non branching as well. Furthermore, if $(\mu_t)_{t\in[0,1]} \subset \mathcal{M}_p(X, A)$ is a constant speed geodesic, then for any $t \in (0, 1)$ and any $\gamma \in \mathrm{Opt}(\mu_0, \mu_t)$, γ_X^Ω is unique and it is induced by a map.*

Proof. Let $(\mu_t)_{t\in[0,1]} \in \mathrm{Geo}(\mathcal{M}_p(X, A))$, $t \in (0, 1)$ and consider $\gamma^1 \in \mathrm{Opt}(\mu_0, \mu_t)$ and $\gamma^2 \in \mathrm{Opt}(\mu_t, \mu_1)$. By the proof of Lemma 4.5.4, there is $\gamma \in \mathcal{B}(X^3)$ such that

$$\pi^{12}_\# \gamma = \gamma^1 + \sigma^1 \quad \text{and} \quad \pi^{23}_\# \gamma = \gamma^2 + \sigma^2$$

for some $\sigma^1, \sigma^2 \in \mathcal{B}(X^2)$ supported on $\Delta(A \times A)$, and such that $\pi^{13}_\# \gamma \in \mathrm{Opt}(\mu_0, \mu_1)$. Moreover,

$$d(x, y) = td(x, z), \ d(y, z) = (1 - t)d(x, z)$$

for γ-a.e. (x, y, z).

Now consider $(x, y, z), (x', y, z') \in \operatorname{supp}(\gamma)$. By Proposition 4.3.1, we know that $\operatorname{supp}(\pi^{13}_{\#}\gamma)$ is in the superdifferential of a c-concave function (where c is given by (4.10)), which implies it is c-cyclically monotone. Therefore,

$$\begin{aligned} d(x,z)^p + d(x',z')^p &\le d(x,z')^p + d(x',z)^p \\ &\le (d(x,y) + d(y,z'))^p + (d(x',y) + d(y,z))^p \\ &= (td(x,z) + (1-t)d(x',z'))^p + (td(x',z') + (1-t)d(x,z))^p \\ &\le td(x,z)^p + (1-t)d(x',z')^p + td(x',z')^p + (1-t)d(x,z)^p \\ &= d(x,z)^p + d(x',z')^p \end{aligned}$$

where the last inequality is due to the convexity of $t \mapsto t^p$ for $p > 1$. Moreover, the strong convexity of the same function, and the fact the all the inequalities above are equations, imply that $d(x, z) = d(x', z')$ and

$$d(x,y) + d(y,z') = d(x,z').$$

In particular, $x,\ y,\ z'$ lie in a geodesic. Since X is non-branching, we get that $z = z'$, and analogously we get that $x = x'$. In other words, the map $\pi^2 \colon (x, y, z) \mapsto y$ is injective in $\operatorname{supp}(\gamma)$. In particular, if T is the inverse of $\pi^2|_{\operatorname{supp}(\gamma)}$, we get that

$$(\pi^1 \circ T, \mathrm{id})_{\#}\mu_t = (\gamma^1)^{\Omega}_X \quad \text{and} \quad (\mathrm{id}, \pi^3 \circ T)_{\#}\mu_t = (\gamma^2)^X_{\Omega}.$$

Therefore, $(\gamma^1)^{\Omega}_X$ and $(\gamma^2)^X_{\Omega}$ are induced by maps. Moreover, this also implies that $(\gamma^1)^{\Omega}_X$ is unique, because otherwise we could construct $\gamma \in \operatorname{Opt}(\mu_0, \mu_t)$ such that

$$\gamma^{\Omega}_X = \frac{1}{2}\left((\pi^1 \circ T, \mathrm{id})_{\#}\mu_t + (\pi^1 \circ T', \mathrm{id})_{\#}\mu_t\right),$$

which would not be induced by a map.

Finally, to prove that $\mathcal{M}_p(X, A)$ is non-branching, consider $t_0 \in (0, 1)$ and geodesics $(\mu_t)_{t\in[0,1]}, (\mu'_t)_{t\in[0,1]}$ such that $\mu_0 = \mu'_0$ and $\mu_{t_0} = \mu'_{t_0}$. Let $\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathcal{M}((e_0, e_1)^{-1}(E_\Omega))$ be such that $\mu_t = (e_t)_{\#}\boldsymbol{\gamma}|_\Omega$ and $\mu'_t = (e_t)_{\#}\boldsymbol{\gamma}'|_\Omega$ for all $t \in [0, 1]$. Then we have

$$(e_0, e_{t_0})_{\#}\boldsymbol{\gamma},\ (e_0, e_{t_0})_{\#}\boldsymbol{\gamma}' \in \operatorname{Opt}(\mu_0, \mu_{t_0})$$

which, due to our previous arguments, implies

$$(e_0, e_{t_0})_\# \boldsymbol{\gamma} = (e_0, e_{t_0})_\# \boldsymbol{\gamma}|_{X\times\Omega} = (e_0, e_{t_0})_\# \boldsymbol{\gamma}'|_{X\times\Omega} = (e_0, e_{t_0})_\# \boldsymbol{\gamma}',$$

where we have used that $\operatorname{supp}((e_{t_0})_\# \boldsymbol{\gamma}) \subset \Omega$ for $t_0 \in (0,1)$, thanks to Remark 4.5.3. Since X is non-branching, the map $(e_0, e_{t_0})\colon \operatorname{Geo}(X) \to X \times X$ is injective, which implies

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}',$$

therefore

$$\mu_1 = (e_1)_\# \boldsymbol{\gamma}|_\Omega = (e_1)_\# \boldsymbol{\gamma}'|_\Omega = \mu_1',$$

and the proposition follows. □

Remark 4.5.8. Theorems 4.5.1 and 4.5.7 imply items 3 and 4 in Theorem F

4.6 Non-negative curvature

In this section we prove that $\mathcal{M}_2(X, A)$ inherits the property of having non-negative curvature in the sense of Alexandrov. This provides a new way to construct Alexandrov spaces. The proof is an adaptation of that of [2, Theorem 2.20].

Theorem 4.6.1. *Assume that (X, A) is a metric pair such that X is a non-negatively curved Alexandrov space. Then $\mathcal{M}_2(X, A)$ is a non-negatively curved Alexandrov space.*

Proof. Since X is proper and geodesic, it follows that $\mathcal{M}_2(X, A)$ is complete and geodesic, due to Theorems 4.4.1 and 4.5.1. Now, let $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic in $\mathcal{M}_2(X, A)$. Let also $\nu \in \mathcal{M}_2(X, A)$ be some measure. By Theorem 4.5.1, we know there exists $\boldsymbol{\gamma} \in \mathcal{M}((e_0, e_1)^{-1}(E_\Omega))$ such that $(e_0, e_1)_\# \boldsymbol{\gamma} \in \operatorname{Opt}(\mu_0, \mu_1)$ and $(e_t)_\# \boldsymbol{\gamma}|_\Omega = \mu_t$ for all $t \in [0,1]$. Fix $t \in (0,1)$ and consider $\gamma \in \operatorname{Opt}(\mu_t, \nu)$. By observing that $(e_t)_\# \boldsymbol{\gamma} = \mu_t = \pi^1_\# \gamma|_\Omega$ and applying the gluing lemma (Theorem

2.4.7), we get a measure $\boldsymbol{\alpha} \in \mathcal{B}(\mathrm{Geo}(X) \times X)$ such that

$$\begin{aligned} \pi^{\mathrm{Geo}(X)}_{\#} \boldsymbol{\alpha} &= \boldsymbol{\gamma} \\ (e_t \circ \pi^{\mathrm{Geo}(X)}, \pi^X)_{\#} \boldsymbol{\alpha} &= \gamma^X_\Omega \end{aligned}$$

It is therefore easy to check that

$$\begin{aligned} (e_0 \circ \pi^{\mathrm{Geo}(X)}, \pi^X)_{\#} \boldsymbol{\alpha}|_{E_\Omega} + \gamma^\Omega_A \in \mathrm{Adm}(\mu_0, \nu), \\ (e_1 \circ \pi^{\mathrm{Geo}(X)}, \pi^X)_{\#} \boldsymbol{\alpha}|_{E_\Omega} + \gamma^\Omega_A \in \mathrm{Adm}(\mu_1, \nu) \end{aligned}$$

In particular,

$$\begin{aligned} &\mathrm{Wb}_2^2(\mu_t, \nu) = \int_{E_\Omega} d(x,z)^2 \, d\gamma(x,z) \\ &= \int_{\Omega \times X} d(x,z)^2 \, d\gamma(x,z) + \int_{A \times \Omega} d(x,z)^2 \, d\gamma(x,z) \\ &= \int_{\mathrm{Geo}(X) \times X} d(\xi_t, z)^2 \, d\boldsymbol{\alpha}(\xi, z) + \int_{A \times \Omega} d(x,z)^2 \, d\gamma(x,z) \\ &\geq \int_{\mathrm{Geo}(X) \times X} (1-t) d(\xi_0, z)^2 + t d(\xi_1, z)^2 - (1-t) t d(\xi_0, \xi_1)^2 \, d\boldsymbol{\alpha}(\xi, z) \\ &\qquad + \int_{A \times \Omega} d(x,z)^2 \, d\gamma(x,z) \\ &\geq (1-t) \left(\int_{E_\Omega} d(x,z)^2 \, d(e_0 \circ \pi^{\mathrm{Geo}(X)}, \pi^X)_{\#} \boldsymbol{\alpha}(x,z) + \int_{A \times \Omega} d(x,z)^2 \, d\gamma(x,z) \right) \\ &\qquad + t \left(\int_{E_\Omega} d(x,z)^2 \, d(e_1 \circ \pi^{\mathrm{Geo}(X)}, \pi^X)_{\#} \boldsymbol{\alpha}(x,z) + \int_{A \times \Omega} d(x,z)^2 \, d\gamma(x,z) \right) \\ &\qquad - (1-t) t \int_{E_\Omega} d(x,z)^2 \, d((e_0, e_1) \circ \pi^{\mathrm{Geo}(X)})_{\#} \boldsymbol{\alpha}(x,z) \\ &\geq (1-t) \mathrm{Wb}_2^2(\mu_0, \nu) + t \mathrm{Wb}_2^2(\mu_1, \nu) - (1-t) t \mathrm{Wb}_2^2(\mu_0, \mu_1), \end{aligned}$$

which proves the claim. □

Remark 4.6.2. Theorem 4.6.1 implies item 4 in Theorem F.

In analogy to Proposition 3.6.5, the following proposition shows that $\mathcal{M}_2(X, A)$ always has an extremal point at the zero measure.

Proposition 4.6.3. *The space of directions at the zero measure, $\Sigma_0(\mathcal{M}_2(X, A))$, has diameter no greater than $\pi/2$.*

Proof. Let $\mu, \nu \in \mathcal{M}_2(X, A)$. Then, we know that

$$\mathrm{Wb}_2(\mu, 0)^2 + \mathrm{Wb}_2(\nu, 0)^2 \geq \mathrm{Wb}_2(\mu, \nu)^2$$

since the transport plan $(\mathrm{id}, \mathrm{proj}_A)_\# \mu + (\mathrm{proj}_A, \mathrm{id})_\# \nu \in \mathrm{Adm}(\mu, \nu)$ is suboptimal. Therefore, if $\xi_1, \xi_2 \in \mathrm{Geo}(\mathcal{M}_2(X, A))$ are geodesics with $\xi_1(0) = \xi_2(0) = 0$, then

$$\cos \measuredangle(\xi_1, \xi_2) = \lim_{s,t \to 0} \frac{\mathrm{Wb}_2(\xi_1(s), 0)^2 + \mathrm{Wb}_2(\xi_2(t), 0)^2 - \mathrm{Wb}_2(\xi_1(s), \xi_2(t))^2}{2\mathrm{Wb}_2(\xi_1(s), 0)\mathrm{Wb}_2(\xi_2(t), 0)} \geq 0$$

which implies that $\measuredangle(\xi_1, \xi_2) \leq \pi/2$. □

4.7 Embedding of $\mathcal{D}_p(X, A)$ into $\mathcal{M}_p(X, A)$

In this final section, we generalise [24, Proposition 3.5], and prove that $\mathcal{D}_p(X, A)$ is isometrically embedded into $\mathcal{M}_p(X, A)$, which proves Theorem G. Indeed, there is a natural inclusion $\mathcal{D}_p(X, A) \hookrightarrow \mathcal{M}_p(X, A)$ given by

$$\sigma \mapsto \sum_{x \in \sigma|_{X \setminus A}} \delta_x.$$

The proofs of the following results are adaptations of those of [24, Lemma 3.4 and Proposition 3.5].

Proposition 4.7.1. *Let $\mu \in \mathcal{M}_p(X, A)$, $r > 0$ and $A_r = \{x \in X : d(x, A) \leq r\}$. Let $\mu^r = \mu|_{X \setminus A_r}$. Then $\mathrm{Wb}_p(\mu^r, \mu) \to 0$ when $r \to 0$. Similarly, if $\sigma \in \mathcal{D}_p(X, A)$, we have $d_p(\sigma^r, \sigma) \to 0$ as $r \to 0$.*

Proof. Let $\gamma \in \mathrm{Adm}(\mu, \mu^r)$ be the partial transport plan given by

$$\gamma = (\mathrm{id}, \mathrm{id})_\# \mu|_{X \setminus A_r} + (\mathrm{id}, \mathrm{proj}_A)_\# \mu|_{A_r}.$$

Therefore,

$$\mathrm{Wb}_p^p(\mu, \mu^r) \leq \int_{A_r} d(x, A)^p \, d\mu(x).$$

By the monotone convergence theorem applied to μ with the functions $f_r(x) = d(x, A)^p \cdot 1_{X \setminus A_r}(x)$, we conclude that $\mathrm{Wb}_p(\mu, \mu^r) \to 0$ as $r \to 0$. Similar arguments show that $d_p(\sigma, \sigma^r) \to 0$ as $r \to 0$. □

Theorem 4.7.2. *For $\sigma,\tau \in \mathcal{D}_p(X,A)$, $\mathrm{Wb}_p(\sigma,\tau) = d_p(\sigma,\tau)$.*

Proof. First we consider the case when both $\sigma \setminus A$ and $\tau \setminus A$ have finite cardinality (counting multiplicity), that is, $\sigma \setminus A = \{\{x_1,\dots,x_m\}\}$ and $\tau \setminus A = \{\{y_1,\dots,y_n\}\}$ for some $x_i, y_j \in \Omega$, $i = 1,\dots,m$, $j = 1,\dots,n$. Let us define

$$\begin{aligned}\widetilde{\sigma} &= \{\{x_1,\dots,x_m,\mathrm{proj}_A(y_1),\dots,\mathrm{proj}_A(y_n)\}\}\,,\\ \widetilde{\tau} &= \{\{y_1,\dots,y_n,\mathrm{proj}_A(x_1),\dots,\mathrm{proj}_A(x_m)\}\}\,.\end{aligned}$$

Then, it is clear that

$$d_p^p(\sigma,\tau) = d_p^p(\widetilde{\sigma},\widetilde{\tau}) = \min_P \langle P, C\rangle_{\mathrm{HS}},$$

where P runs over all permutation matrices of size $(m+n)\times(m+n)$, $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$ denotes the Hilbert–Schmidt inner product of square matrices, and

$$C_{ij} = \begin{cases} d(x_i,y_j)^p & \text{if } 1\le i\le m,\ 1\le j\le n \\ d(x_i,p_A(x_{j-n}))^p & \text{if } 1\le i\le m,\ n< j\le m+n \\ d(y_j,p_A(y_{i-m}))^p & \text{if } m< i\le m+n,\ 1\le j\le n \\ 0 & \text{if } m< i\le m+n,\ n< j\le m+n \end{cases}.$$

Similarly, it is clear that

$$\mathrm{Wb}_p^p(\sigma,\tau) = \min_M \langle M, C\rangle_{\mathrm{HS}} \tag{4.17}$$

where M runs over all matrices of size $(m+n)\times(m+n)$ such that $M_{ij}\ge 0$ and $\sum_{i=1}^{m+n} M_{ij} = \sum_{j=1}^{m+n} M_{ij} = 1$.

However, it is known that minimisers in equation (4.17) are permutation matrices (see [5, 48]). This proves the finite case.

For arbitrary $\sigma,\tau \in \mathcal{D}_p(X,A)$, consider $r>0$ and observe that both σ^r and τ^r contain finitely many points in Ω. Then, due to Proposition 4.7.1, we get that

$$\mathrm{Wb}_p(\sigma,\tau) = \lim_{r\to 0}\mathrm{Wb}_p(\sigma^r,\tau^r) = \lim_{r\to 0} d_p(\sigma^r,\tau^r) = d_p(\sigma,\tau).$$

□

Appendix A

Gromov–Hausdorff convergence of metric pairs

In this appendix, we present the theory of the Gromov–Hausdorff convergence for metric pairs, as introduced in [17] and developed in [1]. We show that this convergence can be metrised (Proposition A.1.11 in the compact case; Proposition A.1.15 in the case of proper length spaces; Proposition A.1.29 in the general case of proper metric spaces). We also prove the embedding, completeness and compactness theorems (Theorems A.2.1, A.2.5, and A.2.6), which are well-known results in the classical theory of Gromov–Hausdorff convergence. For further applications of this framework, we refer to [1]

For the convenience of the reader, we repeat the definition of the Gromov–Hausdorff convergence for metric pairs, as introduced in [17] and studied in [1], and already mentioned in this thesis in Definition 3.7.1.

Definition A.0.1. A sequence $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ *converges in the Gromov–Hausdorff topology to a metric pair* (X, A) if there exist sequences $\{\varepsilon_i\}_{i\in\mathbb{N}}$ and $\{R_i\}_{i\in\mathbb{N}}$ of positive numbers with $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$, and maps $\phi_i\colon \overline{B}_{R_i}(A_i) \to X$ satisfying the following three conditions:

1. $|d_{X_i}(x, y) - d_X(\phi_i(x), \phi_i(y)| \leq \varepsilon_i$ for any $x, y \in \overline{B}_{R_i}(A_i)$;

2. $d_{\mathsf{H}}^{d_X}(\phi_i(A_i), A) \le \varepsilon_i$;

3. $\overline{B}_{R_i}(A) \subset \overline{B}_{\varepsilon_i}(\phi_i(\overline{B}_{R_i}(A_i)))$.

We will denote the Gromov–Hausdorff convergence of metric pairs by

$$(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A).$$

A.1 Metrisability

We now consider the metrisability of the convergence of metric pairs. In order to do this, we consider three cases: compact spaces, proper length spaces, and general proper spaces.

A.1.1 Compact case

Let us first consider the case where the metric spaces are compact.

Definition A.1.1. Let (Z, δ) be a metric space, $X, Y \subset Z$ subsets and $A \subset X$, $B \subset Y$ non-empty closed subsets. The *Hausdorff distance* between (X, A) and (Y, B) is given by

$$d_{\mathsf{H}}^{\delta}((X, A), (Y, B)) = d_{\mathsf{H}}^{\delta}(X, Y) + d_{\mathsf{H}}^{\delta}(A, B)$$

Definition A.1.2. The *Gromov–Hausdorff distance* between two compact metric pairs (X, A) and (Y, B) is defined as

$$d_{\mathsf{GH}}((X, A), (Y, B)) = \inf\{d_{\mathsf{H}}^{\delta}((X, A), (Y, B)) : \delta \text{ admissible on } X \sqcup Y\}.$$

One tipically studies the Gromov–Hausdorff distance from a quantitative point of view through approximations. We now define the corresponding notion for metric pairs.

Definition A.1.3. Let X and Y be metric spaces and $\varepsilon > 0$. A pair of maps $f\colon X \to Y$ and $g\colon Y \to X$ (not necessarily continuous) is an *ε-(Gromov–Hausdorff) approximation* if for every $x, x_1, x_2 \in X$ and $y, y_1, y_2 \in Y$,

$$|d_X(x_1, x_2) - d_Y(f(x_1), f(x_2))| < \varepsilon, \qquad d_X(g \circ f(x), x) < \varepsilon,$$
$$|d_Y(y_1, y_2) - d_X(g(y_1), g(y_2))| < \varepsilon, \qquad d_Y(f \circ g(y), y) < \varepsilon.$$

The set of all such pairs is denoted by $\mathrm{Appr}_\varepsilon(X, Y)$. In the case of metric pairs, one restricts to pair maps as follows: For metric pairs (X, A) and (Y, B), we let

$$\mathrm{Appr}_\varepsilon((X, A), (Y, B)) = \left\{ (f, g) \in \mathrm{Appr}_\varepsilon(X, Y) : \begin{array}{c} d_{\mathsf{H}}(f(A), B) < \varepsilon, \\ d_{\mathsf{H}}(g(B), A) < \varepsilon \end{array} \right\}.$$

Remark A.1.4. In the literature, Gromov–Hausdorff approximations often are not defined as pairs of maps but as one map $f\colon X \to Y$ where f has *distortion* less than ε, i.e. for all $x_1, x_2 \in X$ the map f satisfies $|d_Y(f(x_1), f(x_2)) - d_X(x_1, x_2)| < \varepsilon$, and $B_\varepsilon(f(X)) = Y$ (compare with the maps ϕ_i in Definition 3.7.1). Observe that $(f, g) \in \mathrm{Appr}_\varepsilon(X, Y)$ already implies that f has these properties (for the same ε). In the following, we will see that Gromov–Hausdorff distance less than ε corresponds to the existence of ε-approximations (up to a factor). The next proposition shows that (up to another factor) the definition of Gromov–Hausdorff approximations used here can be replaced by the one described in this remark.

Proposition A.1.5. *Let $f\colon (X, d_X) \to (Y, d_Y)$ be a map between metric spaces with distortion smaller than $\varepsilon > 0$. Then there exists a map $g\colon f(X) \to X$ satisfying $(f, g) \in \mathrm{Appr}_\varepsilon(X, f(X))$. Moreover, if $Y = B_\varepsilon(f(X))$ and $d_{\mathsf{H}}(f(A), B) < \varepsilon$, then there exists a map $h\colon Y \to X$ such that $(f, h) \in \mathrm{Appr}_{3\varepsilon}((X, A), (Y, B))$.*

Proof. We define g choosing some $g(y) \in f^{-1}(y)$ for $y \in f(X)$. We note that $f \circ g = \mathrm{Id}|_{f(X)}$. For $y_1, y_2 \in f(X)$,

$$|d_X(g(y_1), g(y_2)) - d_Y(y_1, y_2)| = |d_X(g(y_1), g(y_2)) - d_Y(f(g(y_1)), f(g(y_2)))| < \varepsilon,$$

and for $x \in X$,

$$d_X(x, g \circ f(x)) = |d_X(x, g \circ f(x)) - d_Y(f(x), f(g \circ f(x)))| < \varepsilon.$$

These two inequalities are satisfied because f has distortion less than ε. The remaining two inequalities are satisfied trivially. Thus, $(f, g) \in \mathrm{Appr}_\varepsilon(X, f(X))$.

If $Y = B_\varepsilon(f(X))$, we define

$$h(y) = \begin{cases} g(y) & \text{if } y \in f(X), \\ g(y') & \text{if } y \notin f(X) \text{ and } y' \in f(X) \text{ is such that } d_Y(y, y') < \varepsilon. \end{cases}$$

We have that $h \circ f = g \circ f$ and then for all $x \in X$,

$$d_X(h \circ f(x), x) < \varepsilon.$$

Now for $y \in Y$, using $f \circ g = \mathrm{Id}|_{f(X)}$ or $f \circ h(y) = f \circ g(y') = y'$ for $y' \in f(X) \cap B_\varepsilon(y)$ as in the definition of h, we get

$$d_Y(f \circ h(y), y) = d_Y(y', y) < \varepsilon.$$

Regarding the distortion of h for every $y_1, y_2 \in Y$,

$$\begin{aligned} |d_X(h(y_1), h(y_2)) - d_Y(y_1, y_2)| &\leq |d_X(h(y_1), h(y_2)) - d_Y(f(h(y_1)), f(h(y_2)))| \\ &\quad + |d_Y(f(h(y_1)), f(h(y_2))) - d_Y(y_1, y_2)| \\ &< \varepsilon + d_Y(f \circ h(y_1), y_1) + d_Y(f \circ h(y_2), y_2) \\ &< 3\varepsilon. \end{aligned}$$

Finally we can prove that $d_{\mathsf{H}}(h(B), A) < 3\varepsilon$ as follows: if $b \in B$ then we know there is some $a \in A$ such that $d(f(a), b) < \varepsilon$ because $d_{\mathsf{H}}(f(A), B) < \varepsilon$, therefore we get

$$d(h(b), a) < \varepsilon + d(f \circ h(b), f(a)) \leq \varepsilon + d(f \circ h(b), b) + d(f(a), b) < 3\varepsilon$$

since f has distortion less than ε. On the other hand, if $a \in A$ then

$$d(a, h \circ f(a)) < \varepsilon$$

as we have seen previously. Thus, we have $h(B) \subset B_{3\varepsilon}(A)$ and $A \subset B_\varepsilon(h(B))$ which implies the claim. □

Proposition A.1.6. *Let (X, A) and (Y, B) be metric pairs and $\varepsilon > 0$. Then the following assertions hold:*

1. *If* $d_{\mathsf{GH}}((X,A),(Y,B)) < \varepsilon$, *then* $\mathrm{Appr}_{2\varepsilon}((X,A),(Y,B)) \neq \varnothing$.

2. *If* $\mathrm{Appr}_{\varepsilon}((X,A),(Y,B)) \neq \varnothing$, *then* $d_{\mathsf{GH}}((X,A),(Y,B)) \leq 4\varepsilon$.

Proof. To prove the first claim, we take a number θ such that $0 < \theta < \varepsilon - d_{\mathsf{GH}}((X,A),(Y,B))$. By the definition of infimum, we have an admissible metric δ with

$$d^{\delta}_{\mathsf{H}}((X,A),(Y,B)) < d_{\mathsf{GH}}((X,A),(Y,B)) + \theta < \varepsilon.$$

Then $d^{\delta}_{\mathsf{H}}(X,Y) < \varepsilon$ and $d^{\delta}_{\mathsf{H}}(A,B) < \varepsilon$. These inequalities imply the following:

1. For every $x \in X$, there exists $y_x \in Y$ such that $\delta(x, y_x) < \varepsilon$.

2. For every $a \in A$, there exists $b_a \in B$ such that $\delta(a, b_a) < \varepsilon$.

3. For every $y \in Y$, there exists $x_y \in X$ such that $\delta(y, x_y) < \varepsilon$.

4. For every $b \in B$, there exists $a_b \in A$ such that $\delta(b, a_b) < \varepsilon$.

With these properties in hand, we define $f\colon X \to Y$ and $g\colon Y \to X$ by setting

$$f(x) = \begin{cases} b_x & \text{if } x \in A, \\ y_x & \text{if } x \in X \setminus A, \end{cases}$$

$$g(y) = \begin{cases} a_y & \text{if } y \in B, \\ x_y & \text{if } y \in Y \setminus B. \end{cases}$$

By the definition of f, $\delta(f(x), x) < \varepsilon$ for every $x \in X$. Thus,

$$|d_Y(f(x), f(x')) - d_X(x, x')| \leq \delta(f(x), x) + \delta(f(x'), x') < 2\varepsilon$$

for every $x, x' \in X$. Analogously,

$$|d_X(g(y), g(y')) - d_Y(y, y')| < 2\varepsilon,$$

for every $y, y' \in Y$. Now,

$$\begin{aligned} d_X(g \circ f(x), x) &= \delta(g \circ f(x), x) \\ &\leq \delta(g \circ f(x), f(x)) + \delta(f(x), x) \\ &< 2\varepsilon, \end{aligned}$$

because $\delta(g(y), y) < \varepsilon$ by definition. Also, $d_Y(f \circ g(y), y) < 2\varepsilon$.

Finally, we notice that $f(A) \subset B \subset B_\varepsilon(B)$ and $B \subset B_{2\varepsilon}(f(A))$ because for any $b \in B$ we have $d_Y(b, f(g(b))) < 2\varepsilon$ due to the previous argument. Therefore, $d_{\mathsf{H}}(f(A), B) < 2\varepsilon$ and, in a similar way, we can prove that $d_{\mathsf{H}}(g(B), A) < 2\varepsilon$. Thus,

$$(f, g) \in \mathrm{Appr}_{2\varepsilon}((X, A), (Y, B)).$$

In order to prove the second claim, we take $(f, g) \in \mathrm{Appr}_\varepsilon((X, A), (Y B))$. We define an admissible metric $\delta\colon (X \bigsqcup Y) \times (X \bigsqcup Y) \to \mathbb{R}$ by setting

$$\delta(y, x) = \delta(x, y) = \begin{cases} d_X(x, y) & \text{if } x \in X,\ y \in X, \\ d_Y(x, y) & \text{if } x \in Y,\ y \in Y, \\ \frac{\varepsilon}{2} + \inf\{d_X(x, x') + d_Y(f(x'), y) : x' \in X\} & \text{if } x \in X,\ y \in Y. \end{cases}$$

By definition, δ is symmetric and positive definite. To prove the triangle inequality, first we take $x_1, x_2 \in X$ and $y \in Y$. Then

$$\begin{aligned} \delta(x_1, x_2) + \delta(x_2, y) &= d_X(x_1, x_2) + \frac{\varepsilon}{2} + \inf\{d_X(x_2, x') + d_Y(f(x'), y) : x' \in X\} \\ &= \frac{\varepsilon}{2} + \inf\{d_X(x_1, x_2) + d_X(x_2, x') + d_Y(f(x'), y) : x' \in X\} \\ &\geq \frac{\varepsilon}{2} + \inf\{d_X(x_1, x') + d_Y(f(x'), y) : x' \in X\} \\ &= \delta(x_1, y) \end{aligned}$$

and

$$\begin{aligned}
\delta(x_1, y) + \delta(y, x_2) &= \varepsilon + \inf \left\{ \begin{aligned} & d_X(x_1, x') + d_X(x_2, x'') \\ & + d_Y(f(x'), y) + d_Y(f(x''), y) \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \varepsilon + \inf \left\{ \begin{aligned} & d_X(x_1, x') + d_Y(f(x'), f(x'')) \\ & + d_X(x_2, x'') \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \varepsilon + \inf \left\{ \begin{aligned} & d_X(x_1, x') + (d_X(x', x'') - \varepsilon) \\ & + d_X(x_2, x'') \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \inf \left\{ d_X(x_1, x_2) : x', x'' \in X \right\} \\
&= \delta(x_1, x_2).
\end{aligned}$$

For $x \in X$ and $y_1, y_2 \in Y$,

$$\begin{aligned}
\delta(x, y_1) + \delta(y_1, y_2) &= \frac{\varepsilon}{2} + \inf \left\{ d_X(x, x') + d_Y(f(x'), y_1) : x' \in X \right\} + d_Y(y_1, y_2) \\
&= \frac{\varepsilon}{2} + \inf \left\{ d_X(x, x') + d_Y(f(x'), y_1) + d_Y(y_1, y_2) : x' \in X \right\} \\
&\geq \frac{\varepsilon}{2} + \inf \left\{ d_X(x, x') + d_Y(f(x'), y_2) : x' \in X \right\} \\
&= \delta(x, y_2)
\end{aligned}$$

and

$$\begin{aligned}
\delta(x, y_1) + \delta(x, y_2) &= \varepsilon + \inf \left\{ d_X(x, x') + d_Y(f(x'), y_1) : x' \in X \right\} \\
&\quad + \inf \left\{ d_X(x, x'') + d_Y(f(x''), y_2) : x'' \in X \right\} \\
&= \varepsilon + \inf \left\{ \begin{aligned} & d_X(x, x') + d_Y(f(x'), y_1) \\ & + d_X(x, x'') + d_Y(f(x''), y_2) \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \varepsilon + \inf \left\{ \begin{aligned} & d_X(x', x'') + d_Y(f(x'), y_1) \\ & + d_Y(f(x''), y_2) \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \varepsilon + \inf \left\{ \begin{aligned} & (d_Y(f(x'), f(x'')) - \varepsilon) \\ & + d_Y(f(x'), y_1) + d_Y(f(x''), y_2) \end{aligned} \quad : x', x'' \in X \right\} \\
&\geq \inf \left\{ d_Y(y_1, y_2) : x', x'' \in X \right\} \\
&= \delta(y_1, y_2).
\end{aligned}$$

Using the metric δ, we get, for $x \in X$,

$$\delta(x, f(x)) = \frac{\varepsilon}{2} + \inf \{d_X(x, x') + d_Y(f(x'), f(x)) : x' \in X\} = \frac{\varepsilon}{2}$$

using $x' = x$. For $y \in Y$, we have

$$\delta(y, g(y)) \leq \delta(y, f \circ g(y)) + \delta(f \circ g(y), g(y)) < \varepsilon + \frac{\varepsilon}{2} = \frac{3\,\varepsilon}{2}$$

by the previous inequality and the definition of an ε-approximation. We note that these two inequalities are true when we take $x \in A$ and $y \in B$, respectively. Since $d_{\mathsf{H}}(f(A), B) < \varepsilon$ and $d_{\mathsf{H}}(g(B), A) < \varepsilon$, we obtain $A \subset B^{\delta}_{\varepsilon/2}(f(A)) \subset B^{\delta}_{3\varepsilon/2}(B)$ and $B \subset B^{\delta}_{3\,\varepsilon/2}(g(B)) \subset B^{\delta}_{5\,\varepsilon/2}(A)$. It is also true that $X \subset B^{\delta}_{\varepsilon/2}(f(X)) \subset B^{\delta}_{\varepsilon/2}(Y)$ and $Y \subset B^{\delta}_{3\,\varepsilon/2}(X)$. Putting all together, we obtain

$$d_{\mathsf{GH}}((X, A), (Y, B)) \leq d^{\delta}_{\mathsf{H}}((X, A), (Y, B)) = d^{\delta}_{\mathsf{H}}(X, Y) + d^{\delta}_{\mathsf{H}}(A, B) \leq \frac{3\,\varepsilon}{2} + \frac{5\,\varepsilon}{2} = 4\,\varepsilon.$$

□

Definition A.1.7. Two metric pairs (X, A) and (Y, B) are *isometric* if there exists an isometry $f\colon X \to Y$ with $f(A) = B$.

Theorem A.1.8. *On the space of isometry classes of compact metric pairs, d_{GH} defines a metric.*

The proof of this theorem is the same as [34, Proposition 1.6] using Lemma A.1.6.

We can compare the usual Gromov–Hausdorff distance with its metric pair analogue as follows.

Proposition A.1.9. *Let (X, A) and (Y, B) be compact metric pairs. Then the following assertions hold:*

1. *$d_{\mathsf{GH}}(X, Y) \leq d_{\mathsf{GH}}((X, A), (Y, B))$.*

2. *For any non-empty closed subset $A \subset X$ and for any $n \in \mathbb{N}$, there exists $B_n \subset Y$ such that*

$$d_{\mathsf{GH}}((X, A), (Y, B_n)) \leq 2 d_{\mathsf{GH}}(X, Y) + \frac{2}{n}.$$

Proof. We get both statements from the definitions. For the first one, we take closed subsets $A \subset X$ and $B \subset Y$ and we obtain

$$\begin{aligned} d_{\mathsf{GH}}(X,Y) &= \inf\left\{d_{\mathsf{H}}^{\delta}(X,Y) : \delta \text{ is an admissible metric on } X \sqcup Y\right\} \\ &\leq \inf\left\{d_{\mathsf{H}}^{\delta}(X,Y) + d_{\mathsf{H}}^{\delta}(A,B) : \delta \text{ is an admissible metric on } X \sqcup Y\right\} \\ &= d_{\mathsf{GH}}((X,A),(Y,B)). \end{aligned}$$

Now, to prove the second assertion, we let $r = d_{\mathsf{GH}}(X,Y)$. For any $n \in \mathbb{N}$, there exists an admissible metric δ_n on $X \sqcup Y$ satisfying

$$d_{\mathsf{H}}^{\delta_n}(X,Y) < d_{\mathsf{GH}}(X,Y) + \frac{1}{n} = r + \frac{1}{n}.$$

Thus, $X \subset \overline{B}^{\delta_n}_{r+1/n}(Y)$. Now, if we fix a non-empty closed subset $A \subset X$, then for any $a \in A$ there exists $b_n^a \in Y$ such that

$$\delta_n(a, b_n^a) \leq r + \frac{1}{n}.$$

If $B_n = \{b_n^a : a \in A\}$, then

$$d_{\mathsf{H}}^{\delta_n}(A, B_n) \leq r + \frac{1}{n}.$$

Thus

$$\begin{aligned} d_{\mathsf{H}}^{\delta_n}((X,A),(Y,B_n)) &= d_{\mathsf{H}}^{\delta_n}(X,Y) + d_{\mathsf{H}}^{\delta_n}(A,B_n) \\ &\leq r + \frac{1}{n} + r + \frac{1}{n} \\ &= 2r + \frac{2}{n}. \end{aligned}$$

and

$$d_{\mathsf{GH}}((X,A),(Y,B_n)) \leq 2 d_{\mathsf{GH}}(X,Y) + \frac{2}{n}.$$

□

Corollary A.1.10. *Let X and X_i, $i \in \mathbb{N}$, be compact metric spaces.*

1. *If $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X,A)$ for some $A_i \subset X_i$ and $A \subset X$, then $X_i \xrightarrow{\mathsf{GH}} X$ as well.*
2. *If $X_i \xrightarrow{\mathsf{GH}} X$ and $A \subset X$, then there exist $A_i \subset X_i$ such that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$.*

We now prove that the Gromov–Hausdorff convergence of compact metric pairs can be metrised.

Proposition A.1.11. *If (X, A) and $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ are compact metric pairs, then*

$$(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$$

is equivalent to

$$\lim_{i\to\infty} d_{\mathsf{GH}}((X_i, A_i), (X, A)) = 0. \tag{A.1}$$

Proof. Let us assume that X is compact and $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$. Then we have $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$ and $f_i \colon \overline{B}_{R_i}(A_i) \to X$ as in Definition 3.7.1. Since X is compact, we know that $X = \overline{B}_{R_i}(A)$ for $i \in \mathbb{N}$ sufficiently large. By the triangle inequality and the conditions (1) and (2) in Definition 3.7.1, we get

$$|\operatorname{diam}(\overline{B}_{R_i}(A_i)) - \operatorname{diam}(\overline{B}_{R_i}(A))| \leq 3\varepsilon_i.$$

Thus, there exists $C > 0$ such that $\operatorname{diam}(\overline{B}_{R_i}(A_i)) < C$ for all $i \in \mathbb{N}$. This condition implies that $X_i = \overline{B}_{R_i}(A_i)$ for $i \in \mathbb{N}$ sufficiently large; otherwise, we would have that $\operatorname{diam}(X_i) > R_i$ for arbitrarily large $i \in \mathbb{N}$, which due to the fact that f_i has distortion less than ε_i implies $\operatorname{diam}(X) > R_i - \varepsilon_i$, and this is not possible if X is compact. In particular, $f_i \colon X_i \to X$ satisfies the hypothesis of Proposition A.1.5 for sufficiently large $i \in \mathbb{N}$, which implies that $\operatorname{Appr}_{3\varepsilon_i}((X_i, A_i), (X, A)) \neq \varnothing$. Thanks to Proposition A.1.6 we get $d_{\mathsf{GH}}((X_i, A_i), (X, A)) \to 0$.

Conversely, by (A.1) and Proposition A.1.6, we have $\operatorname{Appr}_{\varepsilon}((X_i, A_i), (X, A)) \neq \varnothing$ for any $\varepsilon > 0$ and sufficiently large $i \in \mathbb{N}$. In particular, if we take $R_i = \operatorname{diam}(X) + i$, any sequence $\varepsilon_i \searrow 0$ such that $d_{\mathsf{GH}}((X, A_i), (X, A)) \leq \varepsilon_i/2$, and $f_i \colon \overline{B}_{R_i}(A_i) \to X$ such that there exists $g_i \colon X \to X_i$ with $(f_i, g_i) \in \operatorname{Appr}_{\varepsilon_i}((X_i, A_i), (X, A))$, we get that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$. □

A.1.2 Proper length spaces

In general, the distance function d_{GH} described above is not well-defined for non-compact metric pairs. However, we can use the distance between metric pairs of the form $(\overline{B}_r(A), A)$ to study the convergence of proper length metric pairs. The following lemma is useful for this goal.

Lemma A.1.12. *Let (X, δ) be a proper length space, $A \subset X$ a closed subspace and $r, s > 0$. Then*

$$B_r(B_s(A)) = B_{r+s}(A).$$

Proof. Let $q \in B_r(B_s(A))$. There exists $x \in B_s(A)$ with $\delta(x, q) < r$. Then

$$\delta(q, A) \leq \delta(q, x) + \delta(x, A) < r + s.$$

Thus, $B_r(B_s(A)) \subset B_{r+s}(A)$.

Conversely, we take $q \in B_{r+s}(A)$. Because $B_s(A) \subset B_r(B_s(A))$, we can assume without loss of generality that $q \in B_{r+s}(A) \setminus B_s(A)$. We set $l = \delta(q, A)$ and we note that $s < l < r + s$. Since A is closed and X is proper, we can take a shortest geodesic γ from A to q, i.e. $\gamma\colon [0, l] \to X$ with $\gamma(0) \in A$ and $\gamma(l) = q$. We define

$$\varepsilon = \frac{1}{2} \min \{s, r + s - l\} > 0$$

and

$$t = s - \varepsilon \in (0, s) \subset [0, l].$$

Then $\delta(\gamma(t), A) = t < s$ and $\delta(\gamma(t), q) = l - t = l - s + \varepsilon < l - s + r + s - l = r$. Therefore, $\gamma(t) \in B_s(A)$ and $q \in B_r(\gamma(t))$, and finally, $B_{r+s}(A) \subset B_r(B_s(A))$. $\square$

Lemma A.1.13. *Let (X, δ) be a proper length space, $A, B \subset X$ be closed subsets, and let $r, s > 0$. Then*

$$d_{\mathsf{H}}^{\delta}(\overline{B}_r(A), \overline{B}_s(B)) \leq d_{\mathsf{H}}^{\delta}(A, B) + |r - s|.$$

Proof. We start by defining $\varepsilon = d_{\mathsf{H}}^{\delta}(A, B) + |r - s| \geq 0$. We have two cases.

If $\varepsilon = 0$, then $d_{\mathsf{H}}^{\delta}(A, B) = 0$ and $r = s$. Then, for any $\varepsilon' > 0$ we have

$$\overline{B}_r(A) \subset \overline{B}_{\varepsilon'+r}(B) = \overline{B}_{\varepsilon'}(\overline{B}_r(B))$$

and

$$\overline{B}_r(B) \subset \overline{B}_{\varepsilon'+r}(A) = \overline{B}_{\varepsilon'}(\overline{B}_r(A)),$$

so $d_{\mathsf{H}}^{\delta}(\overline{B}_r(A), \overline{B}_r(B)) = 0$ as well.

If $\varepsilon > 0$, we apply Lemma A.1.12, and obtain

$$B_r(A) \subset B_{d_{\mathsf{H}}^{\delta}(A,B)+r}(B) \subset B_{d_{\mathsf{H}}^{\delta}(A,B)+|r-s|+s}(B) \subset B_{\varepsilon+s}(B) \subset B_{\varepsilon}(B_s(B))$$

and

$$B_s(B) \subset B_{d_{\mathsf{H}}^{\delta}(A,B)+s}(A) \subset B_{d_{\mathsf{H}}^{\delta}(A,B)+|r-s|+r}(A) \subset B_{\varepsilon+r}(A) \subset B_{\varepsilon}(B_r(A)).$$

Therefore,

$$d_{\mathsf{H}}^{\delta}(\overline{B}_r(A), \overline{B}_s(B)) = d_{\mathsf{H}}^{\delta}(B_r(A), B_s(B)) \leq \varepsilon$$

since $d_{\mathsf{H}}(\overline{B}_r(A), B_r(A) = 0$. □

Corollary A.1.14. *Let (X, δ) be proper length spaces and $A, B \subset X$ be closed subsets. Then*

1. $d_{\mathsf{GH}}((\overline{B}_r(A), A), (\overline{B}_s(A), A)) \leq |r - s|$, *and*
2. $d_{\mathsf{GH}}((\overline{B}_r(A), A), (\overline{B}_r(B), B)) \leq 2d_{\mathsf{H}}^{\delta}(A, B)$.

Observe that Lemmas A.1.12 and A.1.13 and Corollary A.1.14 also hold if, instead of assuming that X is proper, one asks that the subspaces $A, B \subset X$ are compact.

Proposition A.1.15. *If (X, A) and $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ are proper metric spaces then*

$$\lim_{i\to\infty} d_{\mathsf{GH}}((\overline{B}_r(A_i), A_i), (\overline{B}_r(A), A)) = 0 \quad \textit{for all } r > 0. \tag{A.2}$$

implies $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$. If in addition $\{X_i\}_{i\in\mathbb{N}}$ and X are length spaces, then the converse also holds.

Proof. If condition (A.2) holds, then, by [34, Lemma 2.8], there exists $R_i \nearrow \infty$ such that

$$\sup\{d_{\mathsf{GH}}((\overline{B}_{R_i}(A_j), A_j), (\overline{B}_{R_i}(A), A)) : j \geq i\} \leq \frac{1}{R_i}.$$

Therefore, taking $\varepsilon_i = 2/R_i$, we have

$$d_{\mathsf{GH}}((\overline{B}_{R_i}(A_i), A_i), (\overline{B}_{R_i}(A), A)) \leq \frac{\varepsilon_i}{2}$$

which, due to Proposition A.1.6, implies that there exists some

$$(f_i, g_i) \in \mathrm{Appr}_{\varepsilon_i}((\overline{B}_{R_i}(A_i), A_i), (\overline{B}_{R_i}(A), A)).$$

Such choice of ε_i, R_i and f_i clearly satisfies Definition 3.7.1.

Let us now assume that X and $\{X_i\}_{i\in\mathbb{N}}$ are proper length spaces and such that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$. Then we have $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$ and $f_i \colon \overline{B}_{R_i}(A_i) \to X$ ε_i-approximations as in Definition 3.7.1. We can then define a metric δ_i on $\overline{B}_R(A_i) \sqcup \overline{B}_R(A)$ just as in the proof of Proposition A.1.6:

$$\delta_i(x,y) = \begin{cases} d_{X_i}(x,y), & x \in \overline{B}_R(A_i),\ y \in \overline{B}_R(A_i), \\ d_X(x,y), & x \in \overline{B}_R(A),\ y \in \overline{B}_R(A), \\ \dfrac{\varepsilon}{2} + \inf\limits_{x' \in \overline{B}_R(A_i)} \{d_{X_i}(x,x') + d_X(f_i(x'),y)\}, & x \in \overline{B}_R(A_i),\ y \in \overline{B}_R(A), \\ \delta_i(y,x), & x \in \overline{B}_R(A),\ y \in \overline{B}_R(A_i). \end{cases}$$

Clearly, δ_i is an admissible metric on $\overline{B}_R(A_i) \sqcup \overline{B}_R(A)$.

We can see that $A_i \subset \overline{B}^{\delta_i}_{3\varepsilon_i/2}(A)$ as follows: if $x \in A_i$ then $\delta_i(x, f_i(x)) = \varepsilon_i/2$, and we also know there is some $y \in A$ such that $\delta_i(f_i(x), y) \leq \varepsilon_i$, so $\delta_i(x,y) \leq 3\varepsilon_i/2$.

On the other hand, we can also check that $A \subset \overline{B}^{\delta_i}_{3\varepsilon_i/2}(A_i)$: if $y \in A$ then there is some $x \in A_i$ such that $\delta_i(y, f_i(x)) \leq \varepsilon_i$. Therefore $\delta_i(y,x) \leq \varepsilon_i/2 + \delta_i(f_i(x), y) \leq 3\varepsilon_i/2$.

Now we can see that $\overline{B}_R(A_i) \subset \overline{B}_{5\varepsilon_i/2}(\overline{B}_R(A))$: if $\delta_i(x, A_i) \leq R$ then, using the triangle inequality and the properties of f_i, we can verify that $\delta_i(f_i(x), A) \leq R + 2\varepsilon_i$, and since X is a length space, there is some $y \in \overline{B}_R(A)$ such that $\delta_i(f_i(x), y) \leq 2\varepsilon_i$, so

$$\delta_i(x,y) \leq \delta_i(x, f_i(x)) + \delta_i(f_i(x), y) \leq 5\varepsilon_i/2.$$

Let us now prove that $\overline{B}_R(A) \subset \overline{B}_{9\varepsilon_i/2}(\overline{B}_R(A_i))$. Therefore, for any $y \in \overline{B}_R(A)$ there is some $x \in \overline{B}_{R_i}(A_i)$ such that $d_X(y, f_i(x)) < \varepsilon_i$. Since $d_{\mathsf{H}}(f(A_i), A) \le \varepsilon_i$ and the distortion of f_i is less than ε_i, we get that

$$\begin{aligned}|d(x, A_i) - d(y, A)| &\le |d(x, A_i) - d(f(x), f(A_i))| + |d(f(x), f(A_i)) - d(y, f(A_i))| \\ &\quad + |d(y, f(A_i)) - d(y, A)| \\ &\le 3\varepsilon_i.\end{aligned}$$

Therefore, $d(x, A_i) \le d(y, A) + 3\varepsilon_i \le R + 3\varepsilon_i$, which due to the fact that X_i is a length space implies that there is some $x' \in \overline{B}_R(A_i)$ such that $\delta_i(x, x') \le 3\varepsilon_i$. Then

$$\delta_i(y, x') \le \frac{\varepsilon_i}{2} + d_{X_i}(x', x) + d_X(f_i(x), y) \le \frac{9\varepsilon_i}{2}.$$

We conclude then that $d_{\mathsf{H}}^{\delta_i}(A_i, A) \le 3\varepsilon_i/2$ $d_{\mathsf{H}}^{\delta_i}(\overline{B}_R(A_i), \overline{B}_R(A)) \le 9\varepsilon_i/2$. Therefore,

$$d_{\mathsf{GH}}((\overline{B}_R(A_i), A_i), (\overline{B}_R(A), A)) \le \frac{9\varepsilon_i}{2},$$

which proves that condition (A.2) holds. □

A.1.3 General case

The Gromov–Hausdorff distance between non-compact metric spaces is not well-defined in general. However, it is possible to define the Gromov–Hausdorff distance between non-compact pointed metric spaces (see [30] and cf. [12]), which is slightly different from the corresponding definition in the compact case. This notion is thoroughly studied in [32, 34]. We extend this definition to the setting of metric pairs.

Definition A.1.16. Given $\varepsilon > 0$ and metric pairs (X, A), (Y, B), an admissible distance function δ on $X \sqcup Y$ is $(\varepsilon; A, B)$-*admissible* provided

$$d_{\mathsf{H}}^{\delta}(A, B) < \varepsilon, \quad \overline{B}^{\delta}_{1/\varepsilon}(A) \subset B^{\delta}_{\varepsilon}(Y), \quad \overline{B}^{\delta}_{1/\varepsilon}(B) \subset B^{\delta}_{\varepsilon}(X).$$

Definition A.1.17. Let (X, A) and (Y, B) be metric pairs. The *Gromov–Hausdorff distance* between (X, A) and (Y, B) is

$$d_{\mathsf{GH}}((X,A),(Y,B)) = \min\left\{\frac{1}{2}, \widetilde{d}_{\mathsf{GH}}((X,A),(Y,B))\right\},$$

where

$$\widetilde{d}_{\mathsf{GH}}((X,A),(Y,B)) = \inf\left\{\varepsilon > 0 : \begin{array}{l}\text{there exists a } (\varepsilon; A, B)\text{-admissible} \\ \text{distance } \delta \text{ on } X \sqcup Y\end{array}\right\}$$

Remark A.1.18. Observe that both definitions of Gromov–Hausdorff distance between metric pair induce the same topology in the case of compact metric pairs.

Definition A.1.19. Let $f\colon X \to Y$ be a map between metric spaces, let $\varepsilon > 0$, and $A \subset X$, and $B \subset Y$ closed subsets. We say that f is an *ε-rough isometry from (X, A) to (Y, B)* if it satisfies $d_{\mathsf{H}}^{d_Y}(f(A), B) < \varepsilon$, it has distortion less than ε and $Y \subset B_{\varepsilon}^{d_Y}(f(X))$.

Lemma A.1.20. *Let (X, A) and (Y, B) be metric pairs.*

1. *If $d_{\mathsf{GH}}((X,A),(Y,B)) < \varepsilon < 1/2$, then there exists a 2ε-rough isometry $f\colon \overline{B}_{1/\varepsilon}^{d_X}(A) \to Y$ from $(\overline{B}_{1/\varepsilon}^{d_X}(A), A)$ to $(\overline{B}_{1/\varepsilon-2\varepsilon}^{d_Y}(B), B)$.*
2. *Conversely, let $R > \varepsilon > 0$ and suppose that there is an ε-rough isometry $f\colon \overline{B}_{R}^{d_X}(A) \to Y$ from $(\overline{B}_{R}^{d_X}(A), A)$ to $(\overline{B}_{R-\varepsilon}^{d_Y}(B), B)$. Then*
$$d_{\mathsf{GH}}((X,A),(Y,B)) < \max\left\{3\varepsilon, \frac{1}{R-\varepsilon}\right\}.$$

Proof. 1. We suppose that $d_{\mathsf{GH}}((X,A),(Y,B)) < \varepsilon < 1/2$ and we take δ a (ε, A, B)-admissible distance on $X \sqcup Y$. We define $f\colon \overline{B}_{1/\varepsilon}^{d_X}(A) \to Y$ by setting $f(x) \in Y$ with $\delta(x, f(x)) < \varepsilon$.

First, we prove that the distortion of f is less than 2ε. Let $x, y \in \overline{B}_{1/\varepsilon}^{d_X}(A)$. Then

$$\begin{aligned}|d_Y(f(x), f(y)) - d_X(x,y)| &\leq |\delta(f(x), x) + d_X(x,y) + \delta(y, f(y)) - d_X(x,y)| \\ &\leq \delta(f(x), x) + \delta(f(y), y) \\ &< 2\varepsilon.\end{aligned}$$

Let $y \in \overline{B}^{\delta}_{1/\varepsilon-2\varepsilon}(B)$. Since $\overline{B}^{\delta}_{1/\varepsilon}(B) \subset B^{\delta}_{\varepsilon}(X)$, we take x such that $\delta(x,y) < \varepsilon$. Then

$$\begin{aligned}\delta(x,A) &\leq \delta(x,y) + d_Y(y,B) + d^{\delta}_{\mathsf{H}}(A,B)\\ &< 2\varepsilon + d_Y(y,B)\\ &\leq 2\varepsilon + \frac{1}{\varepsilon} - 2\varepsilon\\ &= \frac{1}{\varepsilon}.\end{aligned}$$

Also,

$$d_Y(f(x),y) \leq \delta(x,f(x)) + \delta(x,y) < \varepsilon + \varepsilon = 2\varepsilon,$$

and, therefore,

$$\overline{B}^{d_Y}_{1/\varepsilon-2\varepsilon}(B) \subset B^{d_Y}_{2\varepsilon}(f(\overline{B}_{1/\varepsilon}(A))).$$

2. Let $R > \varepsilon > 0$ and let $f\colon \overline{B}^{d_X}_R(A) \to Y$ be an ε-rough isometry from $(\overline{B}^{d_X}_R(A), A)$ to $(\overline{B}^{d_Y}_{R-\varepsilon}(B), B)$. We define

$$\delta\colon X \sqcup Y \times X \sqcup Y \to \mathbb{R}$$

by

$$\delta(x,y) = \begin{cases} d_X(x,y) & \text{if } x \in X,\ y \in X,\\ d_Y(x,y) & \text{if } x \in Y,\ y \in Y,\\ \inf\limits_{\substack{u \in B^{d_X}_R(A),\, v \in Y\\ d_Y(v,f(u)) \leq \varepsilon}} \left\{ d_X(x,u) + \dfrac{3\varepsilon}{2} + d_Y(y,v) \right\} & \text{if } x \in X,\ y \in Y.\end{cases}$$

We will show that δ is a $(t; A, B)$-admissible distance on $X \sqcup Y$, where $t = \max\left\{3\varepsilon, \frac{1}{R-\varepsilon}\right\}$. Note that for $x \in \overline{B}^{d_X}_R(A)$ and $y \in Y$, we have $\delta(x,y) \leq 3\varepsilon/2 + d_Y(f(x),y)$.

It is clear that δ is symmetric and positive definite. The triangle inequality is valid where the three points lie in X and or in Y. Now, we have several cases to check. The first is where there is one point in X. Suppose that $x \in X$ and $y, z \in Y$. Let $u \in \overline{B}^{d_X}_R(A)$ and $v \in Y$ with $d_Y(v, f(u)) \leq \varepsilon$. Then, by definition,

$$\begin{aligned}\delta(x,z) &\leq d_X(x,u) + \frac{3\varepsilon}{2} + d_Y(v,z)\\ &\leq d_X(x,u) + \frac{3\varepsilon}{2} + d_Y(v,y) + d_Y(y,z)\\ &= d_X(x,u) + \frac{3\varepsilon}{2} + d_Y(v,y) + \delta(y,z).\end{aligned}$$

The preceding inequality implies, after taking the infimum over $u \in \overline{B}_R^{d_X}(A)$ and $v \in Y$ such that $d_Y(v, f(u)) < \varepsilon$, the triangle inequality $\delta(x,z) \leq \delta(x,y) + \delta(y,z)$.

Now suppose that $z \in X$ and $x, y \in Y$. Let $u \in \overline{B}_R^{d_X}(A)$ and $v \in Y$ with $d_Y(v, f(u)) \leq \varepsilon$. Then

$$\begin{aligned}\delta(z,x) &\leq d_X(z,u) + \frac{3\varepsilon}{2} + d_Y(v,x)\\ &\leq d_X(z,u) + \frac{3\varepsilon}{2} + d_Y(v,y) + d_Y(y,x)\\ &= d_X(z,u) + \frac{3\varepsilon}{2} + d_Y(v,y) + \delta(y,x).\end{aligned}$$

Taking the infimum over $u \in \overline{B}_R^{d_X}(A)$ and $v \in Y$ such that $d_Y(v, f(u)) < \varepsilon$, we have $\delta(z,x) \leq \delta(z,y) + \delta(y,x)$.

Suppose that $y \in X$ and $x, z \in Y$. Let $u, p \in \overline{B}_R^{d_X}(A)$ and $v, q \in Y$ such that $d_Y(v, f(u)) \leq \varepsilon$ and $d_Y(q, f(p)) \leq \varepsilon$. Then

$$\begin{aligned}\delta(x,z) &\leq d_Y(x,v) + d_Y(v,f(u)) + d_Y(f(u),f(p)) + d_Y(f(p),q) + d_y(q,z)\\ &\leq d_Y(x,v) + d_Y(f(u),f(p)) + 2\varepsilon + d_Y(q,z)\\ &\leq d_Y(x,v) + d_X(u,p) + |d_Y(f(u),f(p)) - d_X(u,p)| + 2\varepsilon + d_Y(q,z)\\ &\leq d_Y(x,v) + d_X(u,p) + 3\varepsilon + d_Y(q,z)\\ &\leq \left(d_X(y,u) + \frac{3\varepsilon}{2} + d_Y(x,v)\right) + \left(d_X(y,p) + \frac{3\varepsilon}{2} + d_Y(q,z).\right)\end{aligned}$$

Taking the infimum over $u \in \overline{B}_R^{d_X}(A)$ and $v \in Y$ such that $d_Y(v, f(u)) < \varepsilon$, we have $\delta(x,z) \leq \delta(y,x) + \delta(y,z)$.

The case where we have two points in X and therefore one point in Y is analogous.

We see now that it is admissible. Let $u \in A$ and $v \in B$, then $d_{\mathsf{H}}^{\delta}(A, B) < 5\varepsilon/2 < t$. If $x \in \overline{B}^{\delta}_{1/t}(A) \subset \overline{B}^{\delta}_{R}(A)$, then $\delta(x, f(x)) \leq 3\varepsilon/2 < t$. Thus,

$$\overline{B}^{\delta}_{\frac{1}{t}}(A) \subset B^{\delta}_{t}(Y).$$

Let $y \in \overline{B}^{d_Y}_{1/t}(B) \subset \overline{B}^{d_Y}_{R-\varepsilon}(B)$. Since $\overline{B}^{d_Y}_{R-\varepsilon}(B) \subset B^{d_Y}_{\varepsilon}(f(\overline{B}^{d_X}(A)))$, there exists $x \in \overline{B}^{d_X}_{R}(A)$ such that $d_Y(f(x), y) < \varepsilon$. Taking $x = u$ and $y = v$, we get $\delta(x, y) < 3\varepsilon/2 < t$.

Thus, $d_{\mathsf{GH}}((X, A), (Y, B)) < t$. □

Since we have a distance function for metric pairs, we can talk about convergence of sequences. We have the following characterisation.

Proposition A.1.21. *Let $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ be a sequence of metric pairs. The following are equivalent:*

1. $d_{\mathsf{GH}}((X_i, A_i), (X, A)) \to 0$.
2. *For all $R > 0$, there exist $R_i > R$, $\varepsilon_i > 0$ and maps $f \colon \overline{B}^{d_{X_i}}_{R_i}(A_i) \to X$ such that $R_i \to R$, $\varepsilon_i \to 0$, and f_i are ε_i-rough isometries from $\left(\overline{B}^{d_{X_i}}_{R_i}(A_i), A_i\right)$ to $\left(\overline{B}^{d_X}_{R}(A), A\right)$.*
3. *For all $R > \varepsilon > 0$ there is an $I \in \mathbb{N}$ such that for all $i \geq I$ there are maps $f_i \colon \overline{B}^{d_{X_i}}_{R}(A_i) \to X$ that are ε-rough isometries from $\left(\overline{B}^{d_{X_i}}_{R}(A_i), A_i\right)$ to $\left(\overline{B}^{d_X}_{R-\varepsilon}(A), A\right)$.*

Proof. From Lemma A.1.20, part (2), we have that assertion (3) implies assertion (1).

We suppose that assertion (1) holds. Let $R > 0$. We take $I \in \mathbb{N}$ such that

$$\varepsilon_i = 2\, d_{\mathsf{GH}}((X_i, A_i), (X, A)) < \min\left\{\frac{1}{2}, \frac{1}{R+1}\right\}$$

for every $i \geq I$. For $1 \leq i < I$, we define $R_i = R + 1$ and $t_i = 4R$. For $i \geq I$, we define $R_i = R + t_i$, where $t_i = 2\varepsilon_i$. Then $R_i \to R$ and $t_i \to 0$. Also, for $i \geq I$,

$$\frac{1}{\varepsilon_i} \geq R + 1 \geq R + 2\varepsilon_i = R + t_i = R_i$$

and there are $(\varepsilon_i; A_i, A)$-admissible distances δ_i on $X_i \sqcup Z$.

For $1 \leq i < I$, we define the constant maps $f_i \colon \overline{B}^{d_X}_{R_i}(A_i) \to X$ by $f_i(x) = a \in A$. For $i \geq I$, let $f_i(x) \in A$ be any point with $\delta_i(x, f_i(x)) < \varepsilon_i$ for $x \in A_i$ and $f_i(x) \in X \smallsetminus A$ be any point with $\delta_i(x, f_i(x)) < t_i$ with $x \in \overline{B}^{d_{X_i}}_{R_i}(A_i) \smallsetminus A$. These points always exist because

$$\overline{B}^{\delta_i}_{R_i}(A_i) \subset \overline{B}^{\delta_i}_{1/\varepsilon_i}(A_i) \subset B^{\delta_i}_{\varepsilon_i}(X)$$

and $d^{\delta_i}_{\mathsf{H}}(A_i, A) < \varepsilon_i$.

The maps f_i are clearly t_i-rough isometries for $0 \leq i < I$. We assume that $i \geq I$. Since $\delta_i(x, f_i(x)) < \varepsilon_i$ for all $x \in \overline{B}^{d_{X_i}}_{R_i}(A_i)$, we have

$$|d_X(f_i(x), f_i(y)) - d_{X_i}(x, y)| \leq \delta_i(f_i(x), x) + \delta_i(f_i(y), y) \leq 2\varepsilon_i = t_i.$$

Since $\overline{B}^{\delta_i}_R(A) \subset \overline{B}^{\delta_i}_{1/\varepsilon_i}(A) \subset B^{\delta_i}_{\varepsilon_i}(X_i)$, if we take $y \in \overline{B}^{\delta_i}_R(A)$, there exists a point $x \in X_i$ such that $\delta_i(x, y) < \varepsilon_i$. Then

$$d_{X_i}(x, A_i) \leq \delta_i(x, y) + d_X(y, A) + d^{\delta_i}_{\mathsf{H}}(A, A_i) < R + 2\varepsilon_i = R_i.$$

Therefore, $x \in \overline{B}^{d_{X_i}}_{R_i}(A_i)$ and

$$d_X(f_i(x), y) \leq \delta_i(x, f_i(x)) + \delta_i(x, y) < 2\varepsilon_i = 2t_i.$$

Finally, we suppose that assertion (2) holds and let $R > \varepsilon > 0$. We choose $R_i > R$, t_i and f_i as in assertion (2). Therefore, we have maps $f \colon \overline{B}^{d_{X_i}}_{R_i}(A_i) \to X$ whose distortion is less than t_i and $d^{d_X}_{\mathsf{H}}(f_i(A_i), A) < t_i$. We choose $I \in \mathbb{N}$ such that, for all $i \geq I$, $t_i < \varepsilon/3$.

We take $i \geq I$. We have to see that f_i is a ε-rough isometry from $(\overline{B}^{d_{X_i}}_R(A_i), A_i)$ to $(\overline{B}^{d_X}_{R-\varepsilon}(A), A)$ and it is left to prove that $\overline{B}^{d_X}_{R-\varepsilon}(A) \subset B^{d_X}_{\varepsilon}(f(\overline{B}^{d_{X_i}}_R(A_i)))$. Let $y \in B^{d_X}_{R-\varepsilon}(A)$. Then

$$y \in B^{d_X}_R(A) \subset B^{d_X}_{t_i}(f_i(\overline{B}^{d_{X_i}}_{R_i}(A_i)).$$

Hence, there exists $x \in \overline{B}^{d_{X_i}}_{R_i}(A_i)$ with $d_X(f_i(x), y) < t_i$. Thus,

$$\begin{aligned} d_{X_i}(x, A_i) &< d_X(f_i(x), f_i(A)) + t_i \\ &\leq d_X(f_i(x), y) + d_X(y, A) + d_{\mathsf{H}}^{d_X}(A, f_i(A_i)) + t_i \\ &< 3t_i + (R - \varepsilon) \\ &< R. \end{aligned}$$

Therefore, $y \in B^{d_X}_{\varepsilon}(f(B^{d_{X_i}}_R(A_i)))$. □

The following lemma provides another useful method for estimating the truncated Gromov–Hausdorff distance between metric pairs.

Lemma A.1.22 (cf. Lemma 3.3 in [32])**.** *Let* (X, A), (Y, B) *be metric pairs and* $\varepsilon > 0$. *Suppose there are* $\{x_1, \ldots, x_n\} \subset X$ *and* $\{y_1, \ldots, y_n\} \subset Y$ *such that*

$$\begin{aligned} B_{\frac{1}{2\varepsilon}}(A) &\subset \bigcup_{i=1}^{n} B_\varepsilon(x_i), \\ B_{\frac{1}{2\varepsilon}}(B) &\subset \bigcup_{i=1}^{n} B_\varepsilon(y_i), \\ A &\subset \bigcup_{i=1}^{k} B_\varepsilon(x_i), \\ A \cap B_\varepsilon(x_i) &\neq \varnothing \ \forall\ 1 \leq i \leq k, \\ B &\subset \bigcup_{i=1}^{k} B_\varepsilon(y_i), \\ B \cap B_\varepsilon(y_i) &\neq \varnothing \ \forall\ 1 \leq i \leq k, \text{ and} \\ \text{for all } \ i, j, \ |d_X(x_i, x_j) - d_Y(y_i, y_j)| &\leq \varepsilon. \end{aligned}$$

Then $d_{\mathsf{GH}}((X, A), (Y, B)) \leq 3\varepsilon$.

Proof. We define an admissible metric on $X \sqcup Y$ by setting

$$\delta(x, y) = \delta(y, x) = \min_{1 \leq i \leq n} \{d_X(x, x_i) + d_Y(y, y_i)\} + \varepsilon.$$

This is an actual metric and the proof is the same as in [32, Lemma 3.3]. Moreover, if $x \in A$, then there is some $i \in \{1, \ldots, k\}$ such that $x \in B_\varepsilon(x_i)$. Then, for any

$y \in B \cap B_\varepsilon(y_i)$, we have

$$\delta(x, y) \le d_X(x, x_i) + d_Y(y, y_i) + \varepsilon < 3\varepsilon,$$

which implies that $A \subset B^\delta_{3\varepsilon}(B)$. Analogously, we have $B \subset B^\delta_{3\varepsilon}(A)$, thus $d^\delta_H(A, B) < 3\varepsilon$. Finally, $\overline{B}_{\frac{1}{3\varepsilon}}(A) \subset B^\delta_{3\varepsilon}(Y)$ and $\overline{B}_{\frac{1}{3\varepsilon}}(B) \subset B^\delta_{3\varepsilon}(X)$ can be easily verified by an argument analogous to the proof of [32, Lemma 3.3]. □

We introduce the following definitions to understand several results from now on.

Definition A.1.23. Let X be a metric space. A subset $S \subset X$ is *ε-separated* if its cardinality is greater than 1 and for all distinct $x, y \in S$, we have $d_X(x, y) \ge \varepsilon$. For $A \subset X$ and $r > 0$ we define *inner and outer covering* numbers via

$$M(r, A) = \min \left\{ m \in \mathbb{N} : \begin{array}{c} \text{there exist } x_1, \dots, x_m \in X \\ \text{such that } A \subset B_r(x_1) \cup \dots \cup B_r(x_m) \end{array} \right\},$$

$$N(r, A) = \min \left\{ n \in \mathbb{N} : \begin{array}{c} \text{there exist } a_1, \dots, a_n \in A \\ \text{such that } A \subset B_r(a_1) \cup \dots \cup B_r(a_n) \end{array} \right\},$$

and *packing* and *separation* numbers via

$$P(r, A) = \max \left\{ p \in \mathbb{N} : \begin{array}{c} \text{there exist } a_1, \dots, a_p \in A \\ \text{such that } B_r(a_1), \dots, B_r(a_m) \text{ are disjoint} \end{array} \right\},$$

$$S(r, A) = \max \left\{ s \in \mathbb{N} : \begin{array}{c} \text{there exists an } r\text{-separated set} \\ \{a_1, \dots, a_s\} \subset A \end{array} \right\}.$$

The proof of the following lemma is analogous that of [32, Lemma 3.9].

Lemma A.1.24. *Let (X, A) and (Y, B) be metric pairs with $d_{\mathsf{GH}}((X, A), (Y, B)) < \varepsilon < 1/2$. Then for any $(\varepsilon; A, B)$-admissible metric δ on $X \sqcup Y$ and all $R > 0$ and $r > 0$:*

$$R \le 1/\varepsilon \Rightarrow M(r + 2\varepsilon, \overline{B}_R(B) \cap Y) \le N(r, \overline{B}_R(A) \cap X) \text{ and}$$

$$R + r \le 1/\varepsilon \Rightarrow P(r + 2\varepsilon, \overline{B}_{R-2\varepsilon}(B) \cap Y) \le P(r, \overline{B}_R(A) \cap X)$$

We also get an analogous version of [32, Corollary 3.10].

Corollary A.1.25. *Let $d_{\mathsf{GH}}((X_i, A_i), (X, A)) \to 0$. If each X_i is a proper space and X is complete, then X is proper too.*

The proof of the previous corollary is the same as that of [32, Corollary 3.10] after fixing some $a \in A$ and observing that whenever $d_{\mathsf{GH}}((X_i, A_i), (X, A)) < \varepsilon < 1/2$ then we can find δ_i a $(\varepsilon; A_i, A)$-admissible metric on $X_i \sqcup X$ and $a_i \in A_i$ such that $\delta_i(a_i, a) < \varepsilon$.

Corollary A.1.26. *Let X be a proper metric space and Y be a complete metric space such that $d_{\mathsf{GH}}((X, A), (Y, B)) = 0$. Then Y is proper.*

Proposition A.1.27. *Let (X, A) and (Y, B) be metric pairs. Suppose that one space is proper and the other is complete. Then*

$$d_{\mathsf{GH}}((X, A), (Y, B)) = 0$$

if and only if (X, A) and (Y, B) are isometric.

The proof of the preceding proposition is the same as that of [32, Proposition 3.12]. We only notice that the balls $\overline{B}_r(A)$ are separable since $\overline{B}_r(A)$ is proper and is the countable union of compact balls $\overline{B}_s(p)$. This fact allows us to construct the isometry between (X, A) and (Y, B) along the lines of the construction in [32].

Corollary A.1.28. *Let collection of all isometry classes of proper metric pairs (X, A) endowed with d_{GH} is a metric space.*

Proof. It is clear that d_{GH} is symmetric and non-negative, and satisfies the triangle inequality. From Proposition A.1.27, d_{GH} is positive definite. □

Proposition A.1.29. *Let $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ and (X, A) proper metric pairs. Then*

$$(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$$

is equivalent to

$$\lim_{i\to\infty} d_{\mathsf{GH}}((X_i, A_i), (X, A)) = 0.$$

Proof. Let us assume that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (X, A)$, that is, we have $\varepsilon_i \searrow 0$, $R_i \nearrow \infty$ and maps $f_i \colon \overline{B}_{R_i}(A_i) \to X$ as in Definition 3.7.1. If $R > \varepsilon > 0$, take $i \in \mathbb{N}$ sufficiently large such that $R_i > R > \varepsilon > 3\varepsilon_i$. It is clear that the restriction of f_i gives an ε_i-rough isometry from $(\overline{B}_{R_i}(A_i), A_i)$ to $(\overline{B}_{R_i}(A), A)$. Now, this implies that f_i restricted to $\overline{B}_R(A_i)$ still has distortion less than ε_i and $d_{\mathsf{H}}(f_i(A_i), A) \leq \varepsilon_i$. Moreover, for any $y \in \overline{B}_{R-\varepsilon}(A) \subset \overline{B}_{R_i}(A)$, we know there is some $x \in \overline{B}_{R_i}(A_i)$ such that $d_X(y, f_i(x)) < \varepsilon_i$, and by the triangle inequality and the fact that f_i has distortion less than ε_i and $d_{\mathsf{H}}^{d_X}(f_i(A_i), A) \leq \varepsilon_i$, we have

$$|d_{X_i}(x, A_i) - d_X(f_i(x), A)| \leq 2\varepsilon_i,$$

which in turn implies

$$d_{X_i}(x, A_i) \leq d_X(f_i(x), A) + 2\varepsilon_i \leq d_X(f_i(x), y) + d_X(y, A) + 2\varepsilon_i \leq 3\varepsilon_i + R - \varepsilon < R.$$

Thus, $\overline{B}_{R-\varepsilon}(A) \subset B_\varepsilon(f_i(\overline{B}_R(A)))$. This means that f_i induces an ε-rough isometry from $(\overline{B}_R(A_i), A_i)$ to $(\overline{B}_{R-\varepsilon}(A), A)$ for sufficiently large $i \in \mathbb{N}$. Using Proposition A.1.21, we conclude that

$$d_{\mathsf{GH}}((X_i, A_i), (X, A)) \to 0.$$

Conversely, if we assume that

$$d_{\mathsf{GH}}((X_i, A_i), (X, A)) \to 0$$

and we consider sequences $\varepsilon_i \searrow 0$ and $R_i \nearrow \infty$ with $R_i > \varepsilon_i$, then we have ε_i-rough isometries f_i from $(\overline{B}_{R_i}^{d_{X_i}}(A_i), A_i)$ to $(\overline{B}_{R_i-\varepsilon_i}^{d_X}(A), A)$, by assertion (3) of Proposition A.1.21. This is exactly Definition 3.7.1. □

A.2 Embedding, completeness and compactness theorems

The embedding, completeness, and compactness theorems are fundamental results in the classical theory of Gromov–Hausdorff convergence [12, 30]. This section is devoted to prove versions of these theorems for the class of metric pairs. These results are the counterparts of the main results in [32]. The proofs are natural generalisations of the arguments given in [32] but we include most of the details for the sake of completeness. Moreover, we establish analogous results for a larger class, namely, the class of metric tuples.

Theorem A.2.1. *Let $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ be a sequence of proper metric pairs. Suppose that*

$$\sum_{i=1}^{\infty} d_{\mathsf{GH}}((X_i, A_i), (X_{i+1}, A_{i+1})) < \infty.$$

Then there exist a non-complete locally complete metric space Y and a closed subset $W \subset \overline{Y} \setminus Y$, where $\overline{Y}$ is the metric completion of Y, with the following properties:

1. *For each i, the space X_i naturally isometrically embeds into Y.*
2. *The space $\overline{Y}$ is proper.*
3. *$(X_i, A_i) \xrightarrow{\mathsf{GH}} (Z, W)$, where $Z = \overline{Y} \setminus Y$.*

Proof. We can construct the space Y and prove it is a non-complete and locally complete metric space satisfying item 1 just as in the proof of the Embedding Theorem in [32]. Moreover, we define ε_n, R_n and δ_n in the same way as in the proof of the Embedding Theorem in [32]. Namely, we choose $\varepsilon_n > d_{\mathsf{GH}}((X_n, A_n), (X_{n+1}, A_{n+1}))$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. We also set $R_n = 1/\varepsilon_n$ and choose δ_n a $(\varepsilon_n; A_n, A_{n+1})$-admissible metric on $X_n \sqcup X_{n+1}$.

It is clear that any sequence $\{a_i\}_{i\in\mathbb{N}} \subset Y$ such that $a_i \in A_i$ and $d(a_i, a_{i+1}) < \varepsilon_i$ for sufficiently large $i \in \mathbb{N}$ is a Cauchy sequence, therefore it converges to some $a \in \overline{Y}$. We can then define

$$W = \left\{ \lim_{i\to\infty} a_i \in \overline{Y} : \{a_i\}_{i\in\mathbb{N}} \subset Y,\ a_i \in A_i \text{ and } d(a_i, a_{i+1}) < \varepsilon_i \right\}.$$

This set is non-empty, since each A_i is non-empty and we can construct at least one limit of a sequence as in the definition of W. It is also a closed subset of $\overline{Y}$ by a standard diagonal argument.

We will now prove that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (Z, W)$. The argument is very similar to the one used in [32]. We give the details for the convenience of the reader.

Fix $\varepsilon \in (0, 1/2)$ and set $R = 1/\varepsilon$. Take $N \in \mathbb{N}$ such that $\sum_{i=n}^{\infty} \varepsilon_i < \varepsilon/4$ for $n \geq N$.

Claim A.2.2. *$d_{\mathsf{H}}^d(A_n, W) < \varepsilon$ for $n \geq N$.*

If $a \in W$ then choose any sequence $\{a_i\}_{i\in\mathbb{N}} \subset Y$ such that $a_i \in A_i$, $d(a_i, a_{i+1}) < \varepsilon_i$ and $\lim_{i\to\infty} a_i = a$. In particular, if $n \geq N$ then

$$d(a_n, a) = \lim_{i\to\infty} d(a_n, a_i) \leq \lim_{i\to\infty} \sum_{k=n}^{i-1} d(a_k, a_{k+1}) \leq \sum_{k=n}^{\infty} \varepsilon_k < \frac{\varepsilon}{4} < \varepsilon.$$

Therefore $W \subset B_\varepsilon^d(A_n)$.

On the other hand, if $a_n \in A_n$ then we can inductively construct a sequence $\{a_i\}_{i=n}^{\infty} \subset Y$ such that $a_i \in A_i$ and $d(a_i, a_{i+1}) < \varepsilon_i$ for all $i \geq n$, therefore this sequence is convergent in $\overline{Y}$, with some limit $a \in W$ and clearly $d(a_n, a) < \varepsilon$. Thus $A_n \subset B_\varepsilon^d(W)$, which proves the claim.

Claim A.2.3. *$\overline{B}_R^d(A_n) \cap X_n \subset B_\varepsilon^d(Z)$ for $n \geq N$.*

The following is a simple consequence of the definition of ε_n:

$$\overline{B}_R^{\delta_n}(A_n) \cap X_n \subset B_{\varepsilon_n}^{\delta_n}(\overline{B}_{R+2\varepsilon_n}^{\delta_n}(A_{n+1}) \cap X_{n+1}),$$
$$\overline{B}_R^{\delta_n}(A_{n+1}) \cap X_{n+1} \subset B_{\varepsilon_n}^{\delta_n}(\overline{B}_{R+2\varepsilon_n}^{\delta_n}(A_n) \cap X_n)$$

for any $R \in (0, R_n]$.

Given any $x_n \in \overline{B}_R^d(A_n) \cap X_n$, we can then construct a sequence $\{x_i\}_{i=n}^{\infty}$ with $x_i \in X_i$ and $d(x_i, x_{i+1}) < \varepsilon_i$ just as in [32]. Such a sequence is Cauchy and converges to some $x \in Z$ with $d(x_n, x) < \varepsilon$. This implies the claim.

Claim A.2.4. *$\overline{B}_R^d(W) \cap Z \subset B_\varepsilon^d(X_n)$ for $n \geq N$.*

By an analogous argument to the one in [32, Section 4.1.3], we can prove the following Engulfing Conditions: for any $T > 0$ and $N \in \mathbb{N}$ such that

$$T + 2\sum_{k=N}^{\infty} \varepsilon_k < R_n$$

for any $n \geq N$, we have

$$\overline{B}^d_T(A_m) \cap X_m \subset \begin{cases} B^d_{\sum_{k=m}^{n-1} \varepsilon_k}(X_n) & \text{if } n > m \geq N, \\ B^d_{\sum_{k=n}^{m-1} \varepsilon_k}(X_n) & \text{if } m > n \geq N. \end{cases}$$

In particular, if $T = R + 2\varepsilon$ and $N \in \mathbb{N}$ is such that $\sum_{k=N}^{\infty} \varepsilon_k < \varepsilon/4$ then

$$T + 2\sum_{k=N}^{\infty} \varepsilon_k < \frac{1}{\varepsilon} + 2\varepsilon + \frac{\varepsilon}{2} < \frac{7}{2\varepsilon} < R_n$$

for any $n \geq N$. Therefore we have

$$\overline{B}^d_{R+2\varepsilon}(A_m) \cap X_m \subset B^d_{\varepsilon/2}(X_n)$$

for any $m, n \geq N$. In particular, since $\overline{B}^d_{R+\varepsilon}(A_N) \subset \overline{B}^d_{R+2\varepsilon}(A_m)$ for $m \geq N$, which can be easily verified by the definition of N, we get that

$$\overline{B}^d_{R+\varepsilon}(A_N) \cap X_m \subset B^d_{\varepsilon/2}(X_n)$$

for $m, n \geq N$, therefore

$$\overline{B}^d_{R+\varepsilon}(A_N) \cap \bigsqcup_{m=N}^{\infty} X_m \subset B^d_{\varepsilon/2}(X_n)$$

for $n \geq N$. This implies

$$B^d_{R+\varepsilon}(A_N) \cap Z \subset B^d_{R+\varepsilon}(A_N) \cap \overline{\left(\bigsqcup_{m=N}^{\infty} X_m\right)} \subset B^d_{\varepsilon}(X_n).$$

Now if we fix $z \in \overline{B}^d_R(W) \cap Z$ then there is some $w \in W$ such that $d(z, w) \leq R$, and since $w \in W$ then $w = \lim_{n\to\infty} a_n$ for some sequence $\{a_n\}_{n\in\mathbb{N}}$ with $a_n \in A_n$ and $d(a_n, a_{n+1}) < \varepsilon_n$. In particular,

$$d(z, a_N) \leq d(z, w) + d(w, a_N) \leq R + \lim_{m\to\infty} \sum_{k=N}^{m-1} d(a_k, a_{k+1}) \leq R + \sum_{k=N}^{\infty} \varepsilon_k < R + \varepsilon.$$

Therefore $\overline{B}^d_R(W) \cap Z \subset B^d_{R+\varepsilon}(A_N) \cap Z \subset B^d_{\varepsilon}(X_n)$ and the claim follows.

Combining claims 1, 2 and 3 we can conclude that $(X_i, A_i) \xrightarrow{\mathsf{GH}} (Z, W)$. We prove the properness of $\overline{Y}$ by applying the same argument as in the proof of the Embedding Theorem in [32] after fixing some $w \in W$ and some sequence $\{a_i\}_{i\in\mathbb{N}}$ such that $a_i \in A_i$, $d(a_i, a_{i+1}) < \varepsilon_i$ and $w = \lim_{i\to\infty} a_n$, and observing that $(X_i, a_i) \xrightarrow{\mathsf{GH}} (Z, w)$ in the sense of [32]. □

Theorem A.2.5. *The metric d_{GH} in the class of proper metric pairs is complete.*

Proof. Let $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ be a Cauchy sequence of proper metric pairs an take a subsequence $\{(X_{i_k}, A_{i_k})\}_{k\in\mathbb{N}}$ such that

$$\sum_{k=1}^{\infty} d_{\mathsf{GH}}((X_{i_k}, A_{i_k}), (X_{i_{k+1}}, A_{i_{k+1}})) < \infty.$$

Then Theorem A.2.1 implies that $(X_{i_k}, A_{i_k}) \xrightarrow{\mathsf{GH}} (Z, W)$ for some proper metric pair (Z, W). □

Theorem A.2.6. *For any collection $\mathcal{X}$ of (isometry classes of) proper metric pairs that is uniformly bounded in the sense of pairs, i.e. if there exists some $C > 0$ such that $\operatorname{diam}(A) \leq C$ for any $(X, A) \in \mathcal{X}$, the following assertions are equivalent:*

1. *$\mathcal{X}$ is precompact with respect to d_{GH}.*
2. *There exists $\pi\colon (0,\infty) \to (0,\infty)$ such that for all $\varepsilon > 0$,*

   $$P(\varepsilon, \overline{B}_{1/\varepsilon}(A)) \leq \pi(\varepsilon)$$

   for all $(X, A) \in \mathcal{X}$.
3. *There exists $\nu\colon (0,\infty) \to (0,\infty)$ such that for all $\varepsilon > 0$,*

   $$N(\varepsilon, \overline{B}_{1/\varepsilon}(A)) \leq \nu(\varepsilon)$$

   for all $(X, A) \in \mathcal{X}$.

Proof. As in the preceding theorems in this section, the proof is analogous to the one in the pointed case (see [32]). We give the details for convenience of the reader. The implication (2) ⇒ (3) follows directly from the fact that

$$N(\varepsilon, \overline{B}_{1/\varepsilon}(A)) \leq P(\varepsilon/2, \overline{B}_{1/\varepsilon}(A)) \leq P(\varepsilon/2, \overline{B}_{2/\varepsilon}(A)) \leq \pi(\varepsilon/2),$$

so we can define $\nu(\varepsilon) = \pi(\varepsilon/2)$.

On the other hand, if we assume (1) and $\mathcal{X}$ is precompact with respect to d_{GH} and uniformly bounded in the sense of pairs, then in particular $\mathcal{X}$ is totally bounded. Therefore for a fixed $\varepsilon \in (0,1)$ there is a minimal $N \in \mathbb{N}$ (which only depends on ε) such that there exist $(X_1, A_1), \ldots, (X_N, A_N) \in \mathcal{X}$ that make up a $(\varepsilon/5)$-net for $\mathcal{X}$. We then define

$$\pi(\varepsilon) = \max_{1 \leq n \leq N} \{P(\varepsilon/2, \overline{B}_{2/\varepsilon}(A_n) \cap X_n)\},$$

which is finite since each A_n is compact. We can verify that $\pi(\varepsilon)$ satisfies item (2) by applying Lemma A.1.24 and an analogous argument to the one used in [32].

Finally, we prove that (3) $\Rightarrow$ (1). Let us assume there is some function ν as in (3) and fix $\varepsilon > 0$ and a sequence $\{(X_n, A_n)\}_{n\in\mathbb{N}} \subset \mathcal{X}$. We will prove that there is a subsequence $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ such that $d_{\mathsf{GH}}((X_i, A_i), (X_k, A_k)) < 2\varepsilon$ for all $i, k \in \mathbb{N}$ and the conclusion follows as in [32].

As in [32], we first get some $N \in \mathbb{N} \cap (0, \nu(\varepsilon/2)]$ and a subsequence $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ of $\{(X_n, A_n)\}_{n\in\mathbb{N}}$ such that

$$N(\varepsilon/2, \overline{B}_{2/\varepsilon}(A_i)) = N$$

for all $i \in \mathbb{N}$. In particular, there exist distinct points $\{x_{i1}, \ldots, x_{iN}\} \subset \overline{B}_{2/\varepsilon}(A_i)$ such that

$$\overline{B}_{2/\varepsilon}(A_i) \subset \bigcup_{j=1}^{N} B_{\varepsilon/2}(x_{ij})$$

for all $i \in \mathbb{N}$. Moreover, up to taking a subsequence and relabelling, we may assume there is some $1 \leq k \leq N$ such that

$$A_i \subset \bigcup_{j=1}^{k} B_{\varepsilon/2}(x_{ij})$$

and $A_i \cap B_{\varepsilon/2}(x_{ij}) \neq \varnothing$ for all $1 \leq j \leq k$ and for all $i \in \mathbb{N}$.

Now, by observing that

$$d_{X_i}(x_{im}, x_{in}) \leq \operatorname{diam}(\overline{B}_{2\varepsilon}(A_i)) \leq C + \frac{4}{\varepsilon}$$

for all $i \in \mathbb{N}$ and $1 \leq m < n \leq N$ (where $C > 0$ comes from the fact that $\mathcal{X}$ is uniformly bounded in the sense of pairs) we can apply the same argument as in [32, Lemma 3.3] to extract a subsequence $\{(X_j, A_j)\}_{j\in\mathbb{N}}$ of $\{(X_i, A_i)\}_{i\in\mathbb{N}}$ (therefore a subsequence of $\{(X_n, A_n)\}_{n\in\mathbb{N}}$) such that, for all $j, k \in \mathbb{N}$ and all $1 \leq m < n \leq N$,

$$|d_{X_j}(x_{jm}, x_{jn}) - d_{X_k}(x_{km}, x_{kn})| < \varepsilon/2.$$

Applying Lemma A.1.22, we get that $d_{\mathsf{GH}}((X_j, A_j), (X_k, A_k)) < 2\varepsilon$. □

Bibliography

[1] Andrés Ahumada Gómez and Mauricio Che. Gromov-Hausdorff convergence of metric pairs and metric tuples. *Differential Geom. Appl.*, 94:Paper No. 102135, 29, 2024. doi:10.1016/j.difgeo.2024.102135.

[2] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013. doi:10.1007/978-3-642-32160-3_1.

[3] Edward A. Azoff. Borel measurability in linear algebra. *Proc. Amer. Math. Soc.*, 42:346–350, 1974. doi:10.2307/2039503.

[4] David Bate and Ana Lucia Garcia Pulido. Bi-lipschitz embeddings of the space of unordered m-tuples with a partial transportation metric. *Mathematische Annalen*, 2024. doi:10.1007/s00208-024-02831-x.

[5] Garrett Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A.*, 5:147–151, 1946.

[6] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. (Polar decomposition and increasing rearrangement of vector fields). *Comptes Rendus de l'Académie des Sciences. Série I*, 305:805–808, 1987.

[7] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991. doi:10.1002/cpa.3160440402.

[8] Peter Bubenik and Alex Elchesen. Universality of persistence diagrams and the bottleneck and Wasserstein distances. *Computational Geometry. Theory and Applications*, 105/106:Paper No. 101882, 18, 2022. doi:10.1016/j.comgeo.2022.101882.

[9] Peter Bubenik and Alex Elchesen. Virtual persistence diagrams, signed measures, Wasserstein distances, and Banach spaces. *Journal of Applied and Computational Topology*, 6(4):429–474, 2022. doi:10.1007/s41468-022-00091-9.

[10] Peter Bubenik and Irina Hartsock. Topological and metric properties of spaces of generalized persistence diagrams. *Journal of Applied and Computational Topology*, 2024. doi:10.1007/s41468-023-00157-2.

[11] Peter Bubenik and Alexander Wagner. Embeddings of persistence diagrams into Hilbert spaces. *Journal of Applied and Computational Topology*, 4(3):339–351, 2020. doi:10.1007/s41468-020-00056-w.

[12] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. doi:10.1090/gsm/033.

[13] Yuri Burago, Mikhail Gromov, and Grigori Perel'man. A. D. Aleksandrov spaces with curvatures bounded below. *Uspekhi Matematicheskikh Nauk*, 47(2(284)):3–51, 222, 1992. doi:10.1070/RM1992v047n02ABEH000877.

[14] Luis A. Caffarelli and Robert J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Ann. of Math. (2)*, 171(2):673–730, 2010. doi:10.4007/annals.2010.171.673.

[15] Prof. Cayley. On Monge's 'Memoire sur la Theorie des Deblais et des Remblais.'. *Proc. Lond. Math. Soc.*, 14:139–142, 1882/83. doi:10.1112/plms/s1-14.1.139.

[16] Mauricio Che. Optimal partial transport for metric pairs, 2024. arXiv:2406.17674.

[17] Mauricio Che, Fernando Galaz-García, Luis Guijarro, and Ingrid Membrillo Solis. Metric geometry of spaces of persistence diagrams. *Journal of Applied and Computational Topology*, 2024. doi:https://doi.org/10.1007/s41468-024-00189-2.

[18] Mauricio Che, Fernando Galaz-García, Luis Guijarro, Ingrid Membrillo Solis, and Motiejus Valiunas. Basic metric geometry of the bottleneck distance. *Proc. Amer. Math. Soc.*, 152(8):3575–3591, 2024. doi:10.1090/proc/16776.

[19] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: dynamic and Kantorovich formulations. *J. Funct. Anal.*, 274(11):3090–3123, 2018. doi:10.1016/j.jfa.2018.03.008.

[20] Samir Chowdhury. Geodesics in persistence diagram space, 2019. arXiv:1905.10820.

[21] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Inventiones Mathematicae*, 146(2):219–257, 2001. doi:10.1007/s002220100160.

[22] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série VI*, 15(4):613–635, 2006. doi:10.5802/afst.1132.

[23] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 14(5):1550066, 8, 2015. doi:10.1142/S0219498815500668.

[24] Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5(1):1–53, 2021. doi:10.1007/s41468-020-00061-z.

[25] Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010. An introduction. doi:10.1090/mbk/069.

[26] Alessio Figalli. The optimal partial transport problem. *Arch. Ration. Mech. Anal.*, 195(2):533–560, 2010. doi:10.1007/s00205-008-0212-7.

[27] Alessio Figalli and Nicola Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées. Neuvième Série*, 94(2):107–130, 2010. doi:10.1016/j.matpur.2009.11.005.

[28] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[29] Detlef Gromoll, Wilhelm Klingenberg, and Wolfgang Meyer. *Riemannsche Geometrie im Grossen*, volume No. 55 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York, 1968.

[30] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, english edition, 2007. Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[31] Leonid G. Hanin. Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces. *Proc. Amer. Math. Soc.*, 115(2):345–352, 1992. doi:10.2307/2159251.

[32] David A. Herron. Gromov–Hausdorff distance for pointed metric spaces. *The Journal of Analysis*, 24(1):1–38, 2016. doi:10.1007/s41478-016-0001-x.

[33] Thomas W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. Reprint of the 1974 original.

[34] Dorothea Jansen. Notes on pointed Gromov–Hausdorff convergence, 2017. arXiv:1703.09595.

[35] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998. doi:10.1137/S0036141096303359.

[36] Olav Kallenberg. *Random measures*. Akademie-Verlag, Berlin; Academic Press, Inc., London, fourth edition, 1986.

[37] Olav Kallenberg. *Random measures, theory and applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017. doi:10.1007/978-3-319-41598-7.

[38] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117, 2018. doi:10.1007/s00222-017-0759-8.

[39] John Lott and Cédric Villani. Ricci curvature for metric measure spaces via optimal transport. *Annals of Mathematics*, 169:903–991, 2009.

[40] Wolfgang Meyer. Toponogov's Theorem and Applications. Lecture Notes, College on Differential Geometry, Trieste, 1989.

[41] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods*

and Computerized Inversion of Data, 27(12):124007, 22, 2011. doi:10.1088/0266-5611/27/12/124007.

[42] Atish Mitra and Žiga Virk. The space of persistence diagrams on n points coarsely embeds into Hilbert space. *Proceedings of the American Mathematical Society*, 149(6):2693–2703, 2021. doi:10.1090/proc/15363.

[43] Ayato Mitsuishi. A splitting theorem for infinite dimensional Alexandrov spaces with nonnegative curvature and its applications. *Geom. Dedicata*, 144:101–114, 2010. doi:10.1007/s10711-009-9390-1.

[44] James R. Munkres. *Topology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2000. Second edition of [MR0464128].

[45] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000. doi:10.1006/jfan.2000.3557.

[46] Anton Petrunin. Semiconcave functions in Alexandrov's geometry. In *Surveys in differential geometry. Vol. XI*, volume 11 of *Surv. Differ. Geom.*, pages 137–201. Int. Press, Somerville, MA, 2007. doi:10.4310/SDG.2006.v11.n1.a6.

[47] Conrad Plaut. Metric spaces of curvature $\geq k$. In *Handbook of geometric topology*, pages 819–898. North-Holland, Amsterdam, 2002.

[48] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling. doi:10.1007/978-3-319-20828-2.

[49] Giuseppe Savaré and Giacomo Enrico Sodini. A relaxation viewpoint to Unbalanced Optimal Transport: Duality, optimality and Monge formulation. *J. Math. Pures Appl. (9)*, 188:114–178, 2024. doi:10.1016/j.matpur.2024.05.009.

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

[51] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. *Acta Mathematica*, 196:133–177, 2006.

[52] Katharine Turner. Medians of populations of persistence diagrams. *Homology, Homotopy and Applications*, 22(1):255–282, 2020. doi:10.4310/hha.2020.v22.n1.a15.

[53] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry. An International Journal of Mathematics and Computer Science*, 52(1):44–70, 2014. doi:10.1007/s00454-014-9604-7.

[54] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new. doi:10.1007/978-3-540-71050-9.

[55] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Communications on Pure and Applied Mathematics*, 58(7):923–940, 2005. doi:10.1002/cpa.20060.

[56] Alexander Wagner. Nonembeddability of persistence diagrams with $p > 2$ Wasserstein metric. *Proceedings of the American Mathematical Society*, 149(6):2673–2677, 2021. doi:10.1090/proc/15451.

[57] Takumi Yokota. A rigidity theorem in Alexandrov spaces with lower curvature bound. *Mathematische Annalen*, 353(2):305–331, 2012. doi:10.1007/s00208-011-0686-8.

[58] Takumi Yokota. On the spread of positively curved Alexandrov spaces. *Math. Z.*, 277(1-2):293–304, 2014. doi:10.1007/s00209-013-1255-5.