Gromov-Wasserstein Distances and their Lower Bounds

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm

"Mathematical Sciences (SMS)"

der Georg-August University School of Science (GAUSS)

vorgelegt von

Christoph Alexander Weitkamp

aus Minden-Lübbecke

Göttingen, 2022

Thesis Committee:

Prof. Dr. Axel Munk
Institute for Mathematical Stochastics, University of Göttingen

Dr. Katharina Proksch
Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente

Prof. Dr. Dominic Schuhmacher
Institute for Mathematical Stochastics, University of Göttingen

Members of the Examination Board:

Reviewer:
Prof. Dr. Axel Munk
Institute for Mathematical Stochastics, University of Göttingen

Second Reviewer:
Dr. Katharina Proksch
Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente

Further Members of the Examination Board:

Prof. Dr. Bernhard Schmitzer
Institute for Computer Science, University of Göttingen

Dr. Benjamin Eltzner
Research Group Computational Biomolecular Dynamics, Max Planck Institute for Multidisciplinary Sciences

Prof. Dr. Max Wardetzky,
Institute for Numerical and Applied Mathematics, University of Göttingen

Date of the Oral Examination: July 7, 2022

Preface

In various applications in biochemistry, computer vision and machine learning, it is of great interest to compare general objects in a pose invariant manner. Recently, the following approach has received increased attention: Model the objects considered as metric measure spaces and compare them with the Gromov-Wasserstein distance. While this distance has many theoretically appealing properties and is a natural distance concept in numerous frameworks, it is NP-hard to compute. In consequence, several alternatives to the precise determination of this distance have been proposed. On the one hand, it is possible to approximate local optima of the minimization problem corresponding to the calculation of the Gromov-Wasserstein distance by conditional gradient descent. On the other hand, one can work with efficiently computable surrogates and lower bounds for the previously mentioned distance. In this PhD-thesis, we follow the second approach and investigate the statistical potential of some of the meaningful known lower bounds and propose a new surrogate for comparing ultrametric measure spaces.

This dissertation is a compilation of the results of the three articles Weitkamp et al. (2020), Mémoli et al. (2021a) and Weitkamp et al. (2022) which can be found in Chapter A, B and C in the addenda. In Chapter 1, we concisely illustrate the ideas leading to the definition of Gromov-Wasserstein distance and detail several meaningful, polynomial time computable lower bounds that are related to other approaches for pose invariant object matching proposed in the literature. Chapter 2 briefly summarizes the results of Weitkamp et al. (2020) and discusses related work. Chapters 3 and 4 are structured analogously and present the results of Mémoli et al. (2021a) and Weitkamp et al. (2022), respectively.

Own Contributions

- Weitkamp et al. (2020) (Addendum A) was written jointly with K. Proksch, C. Tameling and A. Munk. The limit behavior of the test statistic under the hypothesis was derived in close collaboration with K. Proksch and mostly by me under the alternative. C. Tameling and A. Munk contributed with helpful suggestions concerning the manuscript.

- Mémoli et al. (2021a) (Addendum B) is the joint work of all authors. The main ideas were developed by F. Mémoli, A. Munk and me during a meeting in Göttingen. The precise derivation of the results and the preparation of the manuscript was mostly done by Z. Wan and me with helpful comments and suggestions from F. Mémoli and A. Munk.

- Weitkamp et al. (2022) (Addendum C) is based on equal contributions of K. Proksch and me. Some technical details in the proof of the main statement were thoroughly discussed with T. Staudt. B. Lelandais and C. Zimmer significantly improved the application part of the paper.

Acknowledgments

I am grateful to ***Axel Munk*** for giving me the opportunity to become his PhD student. His support and guidance were fundamental to my development as researcher. Further, he helped me cope with the losses of my favorite football team ***Armina Bielefeld*** ("Na, da hat Bielefeld am Wochenende ja wieder ganz schön verkackt." -A. M.). I would also like to thank my second supervisor ***Katharina Proksch*** for the fruitful discussions and great collaboration in two different projects. Her ideas and positive energy were often essential for progress. I am also indebted to ***Dominic Schumacher*** and ***Anja Sturm*** for their interest in my research as well as various helpful comments and suggestions. Moreover, I would like to thank ***Facundo Mémoli*** and ***Zhengchao Wan*** for the great collaboration which gave me a completely new mathematical perspective.

Furthermore, I would like to acknowledge the support and work of the ***Research Training Group 2088***. The colloquia, yearly retreats and financial support for stays abroad diversified my research interests and facilitated my collaborations.

A special thanks also goes to all my amazing colleagues who helped to create a great atmosphere in the institute, in particular to ***Christian Böhm***, who made sure that my PhD did not fail because of technical problems, and the ***Optimal Transport Crew***: ***Marcel "German Cubano" Klatt***, ***Thomas "The Gecko" Staudt***, ***Shayan "Efficiency Personified" Hundrieser***, ***Giacomo "Mr. Hungry" Nies***, ***Gilles "Terrible Jokes" Mordant*** and ***Florian "I-will-not-Join-for-Lunch" Heinemann***.

I would also like to thank my former math teacher ***Bernhard Waldmüller*** who sparked and reinforced my interest in mathematics and prepared me to fight their challenges.

Finally, I would like to thank ***Marieke Enders***, who had to endure all of my mathematical rehearsal talks, for always being there for me. Another big thanks goes to my family and friends for their support and all the good times we had. Ceterum censeo Carthaginem esse delendam.

Contents

1 Introduction **1**
- 1.1 The Gromov-Hausdorff Distance 1
- 1.2 The Gromov-Wasserstein Distance 3
- 1.3 Lower Bounds 6

2 Distribution of Distances based Object Matching **9**
- 2.1 Main Results 10
- 2.2 Discussion and Related Work 12

3 The Ultrametric Gromov-Wasserstein Distance **15**
- 3.1 Main Results 17
- 3.2 Discussion and Related Work 19

4 Distance-to-Measure Density based Geometric Analysis of Complex Data **21**
- 4.1 Main Results 22
- 4.2 Discussion and Related Work 25

Bibliography **27**

Addenda **37**

A Distribution of Distances based Object Matching: Asymptotic Inference **39**

B The Ultrametric Gromov-Wasserstein Distance **179**

C From Small Scales to Large Scales: Distance-to-Measure Density based Geometric Analysis of Complex Data **271**

List of Symbols

$\mathbb{N}$	Set of non-negative integers
$\mathbb{R}_{\geq 0}$	Set of non-negative real numbers
$\mathbb{R}^d$	Euclidean space of dimension d
$\Rightarrow$	Classical weak convergence
$\rightsquigarrow$	Weak convergence in the sense of Hoffman-Jørgensen
$(\mathbb{X}, d_{\mathbb{X}}), (\mathbb{Y}, d_{\mathbb{Y}}), (\mathbb{Z}, d_{\mathbb{Z}})$	Metric spaces
$\mathcal{M}$	Collection of isometry classes of compact metric spaces
d_{GH}	Gromov-Hausdorff distance (see (1.1))
$\mathcal{S}(\mathbb{Z})$	Collection of compact sets of the space $(\mathbb{Z}, d_{\mathbb{Z}})$
$d_{\mathrm{H}}^{(\mathbb{Z},d_{\mathbb{Z}})}$	Hausdorff distance on $\mathcal{S}(\mathbb{Z})$ (see (1.2))
$\mathcal{R}(\mathbb{X}, \mathbb{Y})$	Set of correspondences of $\mathbb{X}$ and $\mathbb{Y}$
$(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}}), (\mathcal{Z}, d_{\mathcal{Z}}, \mu_{\mathcal{Z}})$	Metric measure spaces
$\mathcal{M}^w$	Collection of isomorphism classes of metric measure spaces
$\mathcal{P}(\mathbb{Z})$	Set of probability measures on $(\mathbb{Z}, d_{\mathbb{Z}})$
$d_{\mathrm{W},p}^{(\mathbb{Z},d_{\mathbb{Z}})}$	Wasserstein distance on $\mathcal{P}(\mathbb{Z})$
$C(\mu, \nu)$	Set of couplings of the probability measures μ and ν
supp	Support of a measure or a function
$d_{\mathrm{GW},p}^{\mathrm{sturm}}$	Sturm's Gromov-Wasserstein distance (see (1.7))
$\mathrm{dis}_p(\pi)$	p-distortion of the measure π (see (1.8))
$d_{\mathrm{GW},p}$	Gromov-Wasserstein distance (see (1.9))
$\mathbf{FLB}_p$	First lower bound (see (1.13))
μ^U, μ^V, U, V	Distribution of distances of the metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ and the corresponding distribution functions
$\mathbf{SLB}_p$	Second lower bound (see (1.14))

$\mathbf{TLB}_p$	Third lower bound (see (1.15))
$DoD_p, DoD_{(\beta)}$	See (2.2) and (2.10)
$\widehat{DoD}_p, \widehat{DoD}_{(\beta)}$	See (2.4) and (2.6)
U_n, V_m	Empirical c.d.f.'s of all pairwise distances of samples from $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ (see (2.5))
Ξ	Limit distribution of $\widehat{DoD}_{(\beta)}$ under the hypothesis H_0 (see (2.7))
$\Gamma_{d_{\mathcal{X}}}$	Term in the definition of the covariance of the Gaussian process $\mathbb{G}$ (see (2.8))
$\mathrm{d}_{\mathcal{X},\kappa}$	Distance-to-Measure (DTM) function (see (2.12) or (4.1))
$\widehat{d}_{\mathcal{X},\kappa}$	Empirical DTM-function (see (2.13) or (4.2))
$D_{\mathcal{X},\kappa}, D_{\mathcal{Y},\kappa}$	DTM-Signature of the metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$
$(\mathbb{X}, u_{\mathbb{X}}), (\mathbb{Y}, u_{\mathbb{Y}}), (\mathbb{Z}, u_{\mathbb{Z}})$	Ultrametric spaces
$\mathcal{U}$	Collection of isometry classes of compact ultrametric spaces
u_{GH}	Ultrametric Gromov-Hausdorff distance (see (3.4))
Λ_∞	A specific ultrametric on $\mathbb{R}_{\geq 0}$ (see (3.3))
$(\mathcal{X}, u_{\mathcal{X}}, \mu_{\mathcal{X}}), (\mathcal{Y}, u_{\mathcal{Y}}, \mu_{\mathcal{Y}})$	Ultrametric measure spaces
$\mathcal{U}^w$	Collection of isomorphism classes of ultrametric measure spaces
$u_{\mathrm{GW},p}^{\mathrm{sturm}}$	Sturm's ultrametric Gromov-Wasserstein distance (see (3.5))
$\mathrm{dis}_p^{\mathrm{ult}}(\pi)$	p-ultra-distortion of the measure π (see (3.6))
$u_{\mathrm{GW},p}$	Ultrametric Gromov-Wasserstein distance (see (3.7))
$\mathbf{SLB}_p^{\mathrm{ult}}$	Ultrametric Second Lower Bound (see (3.9))
$\mathbf{TLB}_p^{\mathrm{ult}}$	Ultrametric Third Lower Bound (see (3.8))
$f_{\mathrm{d}_{\mathcal{X},\kappa}}$	DTM-density
$\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}$	Oracle DTM-density estimator (see (4.4))
$\widehat{f}_{\widehat{d}_{\mathcal{X},\kappa}}$	DTM-density estimator (see (4.5))
$\mathscr{H}^d$	The d-dimensional Hausdorff measure

CHAPTER 1

Introduction

Over the last two decades the acquisition of complex data, structures and shapes has increased drastically. In order to analyze the ever growing collections of data, meaningful methods for comparing general objects are needed. In various fields, e.g., computer vision (Jain and Dorai, 2000; Lowe, 2001) electrical engineering (Papazov et al., 2012; Kuo et al., 2014) and biochemistry (Holm and Sander, 1993; Kufareva and Abagyan, 2011; Brown et al., 2016), one aims to differentiate between objects in a pose invariant manner, i.e., instances of the same object in different spatial orientations should be considered as equal. Furthermore, the need to compare more general, non-Euclidean objects such as graphs, trees, ultrametric spaces and networks, where only the connectivity structure matters, has become apparent (Chen and Safro, 2011; Dong and Sawin, 2020). Numerous approaches to these problems have been studied in the literature and the majority of them is *signature based*. This means that a given point cloud, object or graph is reduced to a comparatively simple signature that is then used for the comparison (see Bustos et al. (2005); Veltkamp and Latecki (2006); Wills and Meyer (2020) for an overview). Examples for signatures used for pose invariant object discrimination are the *shape distributions* that are distributions of angles, volumes and lengths (Osada et al., 2002) or *reduced size functions* (d'Amico et al., 2010), that count the connected components of certain lower level sets. A popular signature for the comparison of graphs are the *spectral distances* (Wilson and Zhu, 2008) that compare the eigenvalues of graph Laplacians. It is important to note that while these signature based approaches often perform reasonably well in applications, the reduction to a specific feature is usually not injective, i.e., different objects are potentially mapped to the same feature and hence cannot be distinguished.

1.1 The Gromov-Hausdorff Distance

One possibility to circumvent the distinguishability problem and to compare objects in a pose invariant manner is to model them as metric spaces $\mathbb{X} = (\mathbb{X}, d_{\mathbb{X}})$ and $\mathbb{Y} = (\mathbb{Y}, d_{\mathbb{Y}})$ and to consider them as elements of the set of isometry classes of compact metric spaces $\mathcal{M}$ (two compact metric spaces are in the same class if and only if there exists an isometry between them). It is well known that $\mathcal{M}$ can be turned into a metric space by equipping it with the *Gromov-Hausdorff distance*

(Gromov et al., 1999), which is for $\mathbb{X}, \mathbb{Y} \in \mathcal{M}$ defined as

$$d_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) := \inf_{\mathbb{Z},\phi,\psi} d_{\mathrm{H}}^{(\mathbb{Z},d_{\mathbb{Z}})}(\phi(\mathbb{X}), \psi(\mathbb{Y})), \tag{1.1}$$

where $\phi : \mathbb{X} \to \mathbb{Z}$ and $\psi : \mathbb{Y} \to \mathbb{Z}$ are isometric embeddings into a metric space $\mathbb{Z} = (\mathbb{Z}, d_{\mathbb{Z}})$ and $d_{\mathrm{H}}^{(\mathbb{Z},d_{\mathbb{Z}})}$ denotes the *Hausdorff distance in* $\mathbb{Z}$. The Hausdorff distance is a metric on the collection of compacts sets in a given metric space $(\mathbb{Z}, d_{\mathbb{Z}})$ which is denoted as $\mathcal{S}(\mathbb{Z})$ throughout the following. For $A, B \in \mathcal{S}(\mathbb{Z})$, this metric is defined as

$$d_{\mathrm{H}}^{(\mathbb{Z},d_{\mathbb{Z}})}(A, B) := \max\left(\sup_{a\in A}\inf_{b\in B} d_{\mathbb{Z}}(a, b),\ \sup_{b\in B}\inf_{a\in A} d_{\mathbb{Z}}(a, b)\right). \tag{1.2}$$

There is an alternative formulation of the Gromov-Hausdorff distance that is slightly more practical. It has been shown that $d_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ can be rewritten as an optimization problem over $\mathcal{R}(\mathbb{X}, \mathbb{Y})$, which stands for the *set of correspondences* of $\mathbb{X}$ and $\mathbb{Y}$. The set $\mathcal{R}(\mathbb{X}, \mathbb{Y})$ contains all $R \subset \mathbb{X} \times \mathbb{Y}$ such that

1. For all $x \in \mathbb{X}$ there exists $y \in \mathbb{Y}$ such that $(x, y) \in R$;
2. For all $y \in \mathbb{Y}$ there exists $x \in \mathbb{X}$ such that $(x, y) \in R$.

It is important to note that $\mathcal{R}(\mathbb{X}, \mathbb{Y}) \neq \emptyset$, since we always have that $\mathbb{X} \times \mathbb{Y} \in \mathcal{R}(\mathbb{X}, \mathbb{Y})$. By Burago et al. (2001, Sec. 7), we obtain that

$$d_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) = \frac{1}{2} \inf_{R\in\mathcal{R}(\mathbb{X},\mathbb{Y})} \sup_{\substack{x_1,x_2\in\mathbb{X}\\ y_1,y_2\in\mathbb{Y}\\ s.t.\ (x_i,y_i)\in R}} |d_{\mathbb{X}}(x_1, x_2) - d_{\mathbb{Y}}(y_1, y_2)| \tag{1.3}$$

$$= \frac{1}{2} \inf_{R\in\mathcal{R}(\mathbb{X},\mathbb{Y})} \|d_{\mathbb{X}}(\cdot,\cdot) - d_{\mathbb{Y}}(\cdot,\cdot)\|_{L^\infty(R\times R)}\,. \tag{1.4}$$

While the Gromov-Hausdorff distance has been successfully applied for specific shape and data analysis tasks (Mémoli and Sapiro, 2004; Bronstein et al., 2006a,b, 2009a,b; Chazal et al., 2009; Bronstein et al., 2010; Carlsson and Mémoli, 2010), its widespread usage is severely hindered by two facts. On the one hand, the practical computation of the Gromov-Hausdorff distance is extremely complicated (Mémoli, 2007) and procedures for estimating it based on point clouds have only been developed under certain smoothness conditions on the underlying objects (Mémoli and Sapiro, 2005; Bronstein et al., 2006a). On the other hand, it is not clear to

Figure 1.1: **Metric Measure Spaces I:** Two metric measure spaces that are isometric but not isomorphic.

what extend the similarity/dissimilarity captured by the Gromov-Hausdorff distance coincides with the human intuition, as it is sensitive to noise (due to the use of the L^∞-norm in (1.4)) and has not been related to many other approaches based on signatures. In consequence, other representations of objects and metrics have been studied.

1.2 The Gromov-Wasserstein Distance

It turns out that it is generally more convenient to model objects with more structure and to consider them as *metric measure spaces*. A metric measure space $\mathcal{X} = (\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ is a triple, where $(\mathcal{X}, d_{\mathcal{X}})$ stands for a compact metric space and $\mu_{\mathcal{X}}$ denotes a probability measure that is fully supported on $\mathcal{X}$. The additional probability measure can be interpreted as marker for the importance of various regions of the considered object. In order to apply this representation for pose invariant object matching, it is important to identify metric measure spaces in a meaningful way. Two metric measure spaces $\mathcal{X} = (\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ are *isomorphic* (i.e., equal), denoted as $\mathcal{X} \cong \mathcal{Y}$, if there there exists an isometry $\phi : \mathcal{X} \to \mathcal{Y}$ such that $\phi\#\mu_{\mathcal{X}} = \mu_{\mathcal{Y}}$. Here, $\phi\#\mu_{\mathcal{X}}$ denotes the pushforward measure. Figure 1.1 illustrates an example of two isometric but non-isomorphic metric measure spaces. Throughout the following, $\mathcal{M}^w$ denotes the collection of isomorphism classes of metric measure spaces.

The additional probability measure allows us to regard objects as compactly supported probability measures instead of compact sets. Thus, it is possible to replace the Hausdorff metric in (1.1) with a relaxed notion of proximity, namely the *Wasserstein distance* (Vaserstein, 1969). This distance is fundamental to various mathematical developments and also known as Kantorovich-Rubinstein distance (Kantorovich and Rubinstein, 1958), Mallows distance (Mallows, 1972) or as the Earth Mover's distance (Rubner et al., 2000). Given a metric space $(\mathbb{Z}, d_{\mathbb{Z}})$, let $\mathcal{P}(\mathbb{Z})$ denote the set of probability measures on $\mathbb{Z}$. Then, the Wasserstein distance of order $1 \le p < \infty$ between $\mu, \nu \in \mathcal{P}(\mathbb{Z})$ is defined as

$$d_{\mathrm{W},p}^{(\mathbb{Z},d_{\mathbb{Z}})}(\mu,\nu) = \left(\inf_{\pi \in C(\mu,\nu)} \int_{\mathbb{Z}\times\mathbb{Z}} d_{\mathbb{Z}}^p(z,z')\, d\pi(z,z') \right)^{\frac{1}{p}}, \tag{1.5}$$

and for $p = \infty$ as

$$d_{\mathrm{W},\infty}^{(\mathbb{Z},d_{\mathbb{Z}})}(\mu,\nu) := \inf_{\pi \in C(\mu,\nu)} \sup_{(z,z')\in \mathrm{supp}(\pi)} d_{\mathbb{Z}}(z,z'), \tag{1.6}$$

where supp (π) denotes the support of the measure π and $C(\mu, \nu)$ denotes the set of all couplings of μ and ν, i.e., the set of all measures π on $\mathbb{Z} \times \mathbb{Z}$ such that

$$\pi(A \times \mathbb{Z}) = \mu(A) \text{ and } \pi(\mathbb{Z} \times B) = \nu(B)$$

for all measurable subsets A and B of $\mathbb{Z}$. For $1 \leq p < \infty$, the Wasserstein distance of order p defines a metric on

$$\mathcal{P}_p(\mathbb{Z}) = \left\{ \mu \in \mathcal{P}(\mathbb{Z}) \,\middle|\, \int_{\mathbb{Z}} d_{\mathbb{Z}}^p(z_0, z)\, d\mu(z) < \infty,\ z_0 \in \mathbb{Z} \text{ arbitrary} \right\}$$

and metrizes weak convergence together with convergence of moments of order p (Villani, 2008, Chap. 6).

Sturm (2006) demonstrates that replacing the Hausdorff distance $d_{\mathrm{H}}^{(\mathbb{Z},d_{\mathbb{Z}})}$ in (1.1) by the Wasserstein distance yields a meaningful metric on $\mathcal{M}^w$, namely *Sturm's Gromov-Wasserstein distance*. Let $\mathcal{X} = (\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ be two metric measure spaces. Then, Sturm's Gromov-Wasserstein distance of order $1 \leq p \leq \infty$ between $\mathcal{X}$ and $\mathcal{Y}$ is given by

$$d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) := \inf_{\mathbb{Z},\phi,\psi} d_{\mathrm{W},p}^{(\mathbb{Z},d_{\mathbb{Z}})}(\phi_{\#}\mu_{\mathcal{X}}, \psi_{\#}\mu_{\mathcal{Y}}), \tag{1.7}$$

where $\phi : \mathcal{X} \to \mathbb{Z}$ and $\psi : \mathcal{Y} \to \mathbb{Z}$ are isometric embeddings into the metric space $(\mathbb{Z}, d_{\mathbb{Z}})$.

Mémoli (2007, 2011) observed that the alternative formulation of the Gromov-Hausdorff distance in (1.4) can also be used to define a metric on $\mathcal{M}^w$. By replacing the set of correspondences $\mathcal{R}(\mathbb{X}, \mathbb{Y})$ by the set of couplings $C(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ and the L^∞- by an L^p-norm, the aforementioned author derives a metric on $\mathcal{M}^w$ that is topologically equivalent to $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, but computationally more tractable, namely the *Gromov-Wasserstein distance*. The *p-distortion*, $1 \leq p \leq \infty$, of a coupling $\pi \in C(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ is given for $1 \leq p < \infty$ as

$$\mathrm{dis}_p(\pi) := \left(\iint_{\mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{Y}} \left| d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \right|^p \pi(dx \times dy)\, \pi(dx' \times dy') \right)^{1/p} \tag{1.8}$$

and for $p = \infty$ it is given as

$$\mathrm{dis}_\infty(\pi) := \sup_{\substack{x,x' \in \mathcal{X},\, y,y' \in \mathcal{Y} \\ s.t.\ (x,y),(x',y') \in \mathrm{supp}(\pi)}} \left| d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \right|.$$

Based on this, the Gromov-Wasserstein distance of order p, $1 \leq p \leq \infty$, is defined as

$$d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \inf_{\pi \in C(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \mathrm{dis}_p(\pi). \tag{1.9}$$

In general it holds that $d_{\mathrm{GW},p} \leq d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and in particular there are instances where the inequality is strict (Mémoli, 2011).

In order to illustrate, why the calculation of the distance $d_{\mathrm{GW},p}$ is in many situations more practical than the calculation of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ consider two finite metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ with $\mathcal{X} = \{x_1, \ldots, x_m\}$ and $\mathcal{Y} = \{y_1, \ldots, y_n\}$. Then, it follows (see Mémoli (2011, Rem. 7.1)) that the calculation of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ boils down to solving

$$\min_{(\pi,d) \in C(\mu_{\mathcal{X}},\mu_{\mathcal{Y}}) \times \mathcal{D}(d_{\mathcal{X}},d_{\mathcal{Y}})} \sum_{i=1}^m \sum_{j=1}^n d_{ij}^p \pi_{ij}, \tag{1.10}$$

where

$$C(\mu_{\mathcal{X}},\mu_{\mathcal{Y}}) = \left\{ \pi \in \mathbb{R}^{m\times n}_{\geq 0} \;\middle|\; 0 \leq \pi_{ij} \leq 1, \text{ where } \begin{array}{c} \sum_j \pi_{ij} = \mu_{\mathcal{X}}(x_i) \text{ for all } 1 \leq i \leq m \\ \text{and} \\ \sum_i \pi_{ij} = \mu_{\mathcal{Y}}(y_j) \text{ for all } 1 \leq j \leq n \end{array} \right\}$$

and

$$\mathcal{D}(d_{\mathcal{X}},d_{\mathcal{Y}}) = \left\{ d \in \mathbb{R}^{m\times n}_{\geq 0} \;\middle|\; \begin{array}{c} |d_{ij} - d_{i'j}| \leq d_{\mathcal{X}}(x_i,x_{i'}) \leq d_{ij} + d_{i'j} \text{ for all } 1 \leq i,i' \leq m \\ \text{and} \\ |d_{ij} - d_{ij'}| \leq d_{\mathcal{Y}}(y_j,y_{j'}) \leq d_{ij} + d_{ij'} \text{ for all } 1 \leq j,j' \leq n \end{array} \right\}.$$

In consequence, the calculation of $d^{\text{sturm}}_{\text{GW},p}$ boils down to solving a bilinear program (which is a special case of a nonconvex quadratic program). We note that the number of variables of this program is given by $2mn$ and the number of constraints by $2\left(m\binom{m}{2} + n\binom{n}{2}\right) + m + n$. On the other hand, the determination of $d_{\text{GW},p}$ is in the current setting equivalent to solving the following quadratic program

$$\min_{\pi\in C(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \sum_{i,i'=1}^{m} \sum_{j,j'=1}^{n} |d_{\mathcal{X}}(x_i,x_{i'}) - d_{\mathcal{Y}}(y_j,y_{j'})|^p \pi_{ij}\pi_{i'j'}. \tag{1.11}$$

It is clear that solving (1.11) only requires mn variables and $m + n$ constraints which makes $d_{\text{GW},p}$ from a practical point of view the more convenient choice.

We stress that both the bilinear problem defined in (1.10) as well as the quadratic program defined in (1.11) constitute non-convex optimization problems that are in general NP-hard to solve (Pardalos and Vavasis, 1991). However, it is possible to approximate local minima of (1.11) by conditional gradient descent (Mémoli, 2011; Peyré et al., 2016, see Algorithm 1 for a simple variant of this approach). This has led to various applications of the Gromov-Wasserstein distance in biochemistry (Nitzan et al., 2019; Demetci et al., 2020) and machine learning (Alvarez-Melis and Jaakkola, 2018; Bunne et al., 2019) as well as to numerous extensions of this distance (Vayer et al., 2019; Chapel et al., 2020; Chowdhury and Needham, 2021; Scetbon et al., 2021; Séjourné et al., 2021).

Algorithm 1 Conditional gradient descent for the approximation of $d_{\text{GW},p}$

input: Two finite metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$
$\pi^{(0)} = \mu_{\mathcal{X}} \otimes \mu_{\mathcal{Y}}$
for $j = 1, 2\ldots$ **do**
  $J\left(\pi^{(j-1)}\right) = \frac{1}{2}\nabla_\pi \text{dis}_p\left(\pi^{(j-1)}\right)$
  $\tilde{\pi}^{(j)}$= Optimal coupling of the Optimal Transport problem with ground loss $J\left(\pi^{(j-1)}\right)$
  $\gamma^{(j)} = \frac{2}{j+2}$ *//Alt. find* $\gamma \in [0,1]$ *that minimizes* $\text{dis}_p\left(\pi^{(j-1)} + \gamma\left(\tilde{\pi}^{(j)} - \pi^{(j-1)}\right)\right)$
  $\pi^{(j)} = \left(1 - \gamma^{(j)}\right)\pi^{(j-1)} + \gamma^{(j)}\tilde{\pi}^{(j)}$
end for
return $\frac{1}{2}\text{dis}_p\left(\pi^{(j)}\right)$

1.3 Lower Bounds

There exist a number of lower bounds of $d_{\mathrm{GW},p}$ that are closely related to signatures used for pose invariant object matching. On the one hand, this suggests that the similarity captured by the Gromov-Wasserstein distance is more natural than that of the Gromov-Hausdorff distance. On the other hand, this yields an alternative to the approximation of (local minima of) $d_{\mathrm{GW},p}$ via conditional gradient descent: Directly working with one of the polynomial time computable lower bounds. This approach has for example been pursued in Gellert et al. (2019) to compare the isosurfaces of various proteins. In what follows, we restrict ourselves to $1 \leq p < \infty$ and present three meaningful and efficiently computable lower bounds.

First Lower Bound (FLB): The *p-eccentricity function* of a metric measure space $\mathcal{X}$, denoted as $s_{\mathcal{X},p}$, is defined as

$$s_{\mathcal{X},p} : \mathcal{X} \to \mathbb{R}_+, \ x \mapsto \left(\int_{\mathcal{X}} d_{\mathcal{X}}^p(x, x')\, d\mu_{\mathcal{X}}(x') \right)^{1/p}. \tag{1.12}$$

It is shown in Mémoli (2011, Prop. 6.1) that p-eccentricity functions can be used to lower bound the Gromov-Wasserstein distance as follows

$$d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{FLB}_p\,(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \left(\int_{\mathcal{X}\times\mathcal{Y}} \left| s_{\mathcal{X},p}(x) - s_{\mathcal{X},p}(y) \right|^p d\pi(x, y) \right)^{1/p}.$$

Remarkably, it is possible to rewrite $\mathbf{FLB}_p$ in terms of the Wasserstein distance between to measures on the real line. Let $S^{-1}_{\mathcal{X},p}$ denote the quantile function corresponding to $\mu^{s_{\mathcal{X},p}} = s_{\mathcal{X},p}\#\mu_{\mathcal{X}}$ and let $S^{-1}_{\mathcal{Y},p}$ as well as $\mu^{s_{\mathcal{Y},p}}$ be defined analogously. Then, Theorem 24 in Chowdhury and Mémoli (2019) yields that

$$\mathbf{FLB}_p\,(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} d_{\mathrm{W},p}^{(\mathbb{R},|\cdot|)}\,(\mu^{s_{\mathcal{X},p}}, \mu^{s_{\mathcal{Y},p}}) = \frac{1}{2} \left(\int_0^1 \left| S^{-1}_{\mathcal{X},p}(t) - S^{-1}_{\mathcal{Y},p}(t) \right|^p dt \right)^{1/p}. \tag{1.13}$$

We observe that the eccentricity functions of the spaces $\mathcal{X}$ and $\mathcal{Y}$ are closely related to the *geodesic shape distribution* which is a signature for pose invariant object matching proposed by Hamza and Krim (2003). Furthermore, it has been shown in Weitkamp et al. (2020) that FLB_1 coincides with the Wasserstein distance between the *Distance-to-Measure signatures* of $\mathcal{X}$ and $\mathcal{Y}$ (with mass parameter one) proposed by Brécheteau (2019).

Second Lower Bound (SLB): Proposition 6.2 in Mémoli (2011) yields that

$$d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p\,(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \inf_{\pi \in \widetilde{\mathcal{C}}} \left(\int_{\mathcal{X}^2\times\mathcal{Y}^2} \left| d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \right|^p d\pi(x, x'y, y') \right)^{\frac{1}{p}},$$

where $\widetilde{\mathcal{C}} := \mathcal{C}(\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}, \mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Y}})$. Let $\mu^U = d_{\mathcal{X}}\#\,(\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}})$ and let $\mu^V = d_{\mathcal{Y}}\#\,(\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Y}})$. Then, we call μ^U and μ^V the *Distribution of the (pairwise) Distances (DoD)* of $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$, respectively. Chowdhury and Mémoli (2019, Thm. 24) prove that

$$\mathbf{SLB}_p(\mathcal{X},\mathcal{Y}) = \frac{1}{2} d_{\mathrm{W},p}^{(\mathbb{R},|\cdot|)}\left(\mu^U,\mu^V\right) = \frac{1}{2}\left(\int_0^1 \left|U^{-1}(t) - V^{-1}(t)\right|^p dt\right)^{1/p}, \tag{1.14}$$

where U^{-1} and V^{-1} are the quantile functions of μ^U and μ^V, respectively. Indeed, the distribution of pairwise distances was proposed as a signature itself for pose invariant object matching and was shown to work well in various examples (Osada et al., 2002; Brinkman and Olver, 2012; Berrendero et al., 2016; Gellert et al., 2019). Furthermore, the discriminative abilities of this signature are well investigated theoretically (Boutin and Kemper, 2004; Mémoli and Needham, 2021).

Third Lower Bound (TLB): Let $\Omega_p : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be given as

$$\Omega_p(x,y) := \inf_{\pi \in \mathcal{C}(\mu_\mathcal{X},\mu_\mathcal{Y})} \left(\int_{\mathcal{X}\times\mathcal{Y}} |d_\mathcal{X}(x,x') - d_\mathcal{Y}(y,y')|^p \, d\pi(x',y')\right)^{1/p}.$$

Then, it follows by Proposition 6.3 in Mémoli (2011) that

$$d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq \mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) := \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_\mathcal{X},\mu_\mathcal{Y})} \left(\int_{\mathcal{X}\times\mathcal{Y}} \Omega_p^p(x,y)\, d\pi(x,y)\right)^{1/p}.$$

It is important to note that this lower bound can be reformulated in terms of the *local distribution of distances* of the metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$, i.e., in terms of the sets of measures $\{d_\mathcal{X}(x,\cdot)\#\mu_\mathcal{X}\}_{x\in\mathcal{X}}$ and $\{d_\mathcal{Y}(y,\cdot)\#\mu_\mathcal{Y}\}_{y\in\mathcal{Y}}$. More precisely, let F_x^{-1} denote the quantile function of $d_\mathcal{X}(x,\cdot)\#\mu_\mathcal{X}$, $x \in \mathcal{X}$, and G_y the one of $d_\mathcal{Y}(y,\cdot)\#\mu_\mathcal{Y}$, $y \in \mathcal{Y}$. Then, it follows by Theorem 24 in Chowdhury and Mémoli (2019) that

$$\mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) = \frac{1}{2} \inf_{\pi \in \mathcal{C}(\mu_\mathcal{X},\mu_\mathcal{Y})} \left(\int_{\mathcal{X}\times\mathcal{Y}} \int_0^1 \left|F_x^{-1}(t) - G_y^{-1}(t)\right|^p dt\, d\pi(x,y)\right)^{1/p}. \tag{1.15}$$

The local distributions of distances are closely related to a signature called *shape context* that has been investigated empirically in Shi et al. (2007) and Ruggeri and Saupe (2008). Further, the local distributions of distances of various structures are studied theoretically in Mémoli and Needham (2021).

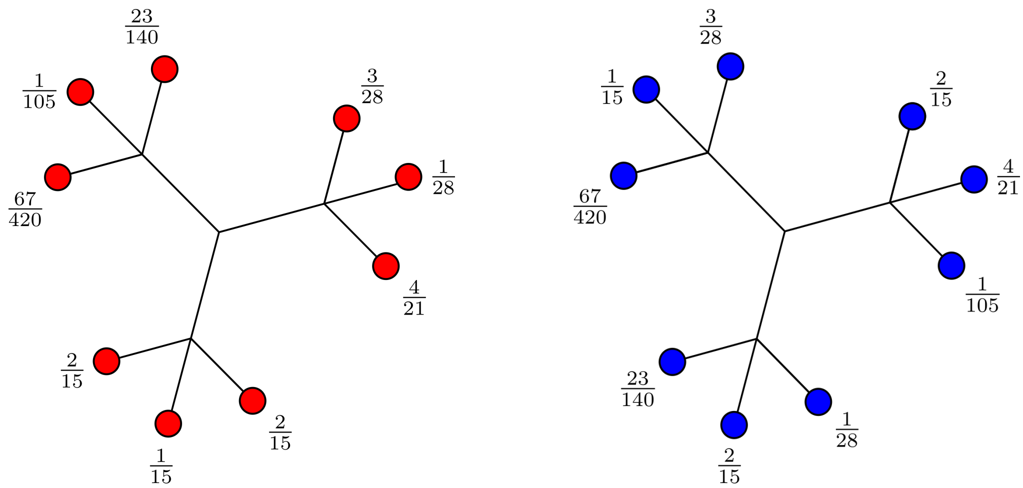

Figure 1.2: **Metric Measure Spaces II:** Illustration of two finite, non-isomorphic metric measure spaces $\mathcal{X}$ (red) and $\mathcal{Y}$ (blue) with $\mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) = 0$. The numbers represent the mass assigned to each node and the length of all edges is $\frac{1}{2}$. Note that by construction, the sum of mass of any three nodes in a branch is $\frac{1}{3}$.

All in all, we obtain the following relation between $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, $d_{\mathrm{GW},p}$ and the lower bounds introduced in this section

$$d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \geq d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq \mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) \geq \begin{cases} \mathbf{SLB}_p(\mathcal{X},\mathcal{Y}) \\ \mathbf{FLB}_p(\mathcal{X},\mathcal{Y}) \end{cases},$$

where all inequalities are strict (Mémoli, 2011; Mémoli and Needham, 2021). While the computation of the lower bounds presented boils down to solving various optimal transport problems and can be done in polynomial time (Mémoli, 2011), it is important to note that they are not able to discriminate between all elements of $\mathcal{M}^w$, i.e., there exist metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ with $\mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) = 0$ but $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) > 0$. One example for such a pair of spaces is illustrated in Figure 1.2.

In conclusion, we have seen that both versions of the Gromov-Wasserstein distance constitute topologically equivalent metrics on the collection of (isomorphism classes of) metric measure spaces. They admit several lower bounds that are strongly connected to signatures for pose invariant object matching proposed in the literature. This suggests that the distances captured by both metrics are natural. Unfortunately, both $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$ are NP-hard to compute. In applications, it is possible to either approximate $d_{\mathrm{GW},p}$ by conditional gradient descent or to directly work with one of its lower bounds.

CHAPTER 2

Distribution of Distances based Object Matching

In this chapter, we summarize and discuss the main results of Paper A. To this end, we first recall the aim and setting of the paper.

In many applications, it is of great interest to decide on the basis of random samples, whether two objects considered are equal or not. By modeling the objects considered as metric measure spaces this problem can be reformulated as follows: Given independent samples $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, \ldots, Y_m\}$ from two metric measure spaces $\mathcal{X} = (\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$, we aim to test the null hypothesis that $\mathcal{X}$ and $\mathcal{Y}$ are isomorphic against the alternative that they are not, i.e.,

$$H_0^* : \mathcal{X} \cong \mathcal{Y} \quad \text{vs} \quad H_1^* : \mathcal{X} \ncong \mathcal{Y}. \tag{2.1}$$

Due to the computational complexity of $d_{\mathrm{GW},p}$, it is not possible to construct a computationally feasible test for H_0^* based on the Gromov-Wasserstein distance. Hence, we propose in Paper A to consider the lower bound $\mathbf{SLB}_p(\mathcal{X}, \mathcal{Y})$, $1 \leq p < \infty$, for constructing such a test. More precisely, we propose to work with

$$DoD_p(\mathcal{X}, \mathcal{Y}) := 2\mathbf{SLB}_p^p(\mathcal{X}, \mathcal{Y}) = \int_0^1 \left| U^{-1}(t) - V^{-1}(t) \right|^p dt, \tag{2.2}$$

where U^{-1} and V^{-1} are defined as in (1.14). As already discussed in Section 1.3, $\mathbf{SLB}_p$ is not able to discriminate between all metric measure spaces in $\mathcal{M}^w$. Hence, the testing problem

$$H_0 : DoD_p(\mathcal{X}, \mathcal{Y}) = 0 \quad \text{vs} \quad H_1 : DoD_p(\mathcal{X}, \mathcal{Y}) > 0 \tag{2.3}$$

is not equivalent to the one defined in (2.1). However, it is important to note that if $DoD_p(\mathcal{X}, \mathcal{Y}) > 0$ the same holds for $d_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$. Consequently, any level α test for H_0 automatically provides a level α test for H_0^*.

A natural, efficiently computable test statistic for the hypothesis H_0 is given by

$$\widehat{DoD}_p = \widehat{DoD}_p\left(\mathcal{X}_n, \mathcal{Y}_m\right) := \int_0^1 \left| U_n^{-1}(t) - V_m^{-1}(t) \right|^p \, dt, \tag{2.4}$$

where, for $t \in \mathbb{R}$, U_n and V_m are defined as the empirical c.d.f.'s of all pairwise distances of the samples $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively, i.e.,

$$U_n(t) := \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\le t\}} \text{ and } V_m(t) := \frac{2}{m(m-1)} \sum_{1\le k<l\le m} \mathbb{1}_{\{d_{\mathcal{Y}}(Y_k,Y_l)\le t\}}. \tag{2.5}$$

Furthermore, U_n^{-1} and V_m^{-1} denote the corresponding empirical quantile functions. It is important to note that the calculation of $\widehat{DoD}_p$ boils down to calculating and sorting the distances $\{d_{\mathcal{X}}(X_i, X_j)\}_{i,j=1}^n$ and $\{d_{\mathcal{Y}}(Y_k, Y_l)\}_{k,l=1}^m$ and no formal integration is necessary. The computational complexity of $\widehat{DoD}_p$ is, up to log factors, given as $O\left((n \vee m)^2\right)$.

2.1 Main Results

The main contributions of Paper A are distributional limits for (trimmed variants of) the statistic $\widehat{DoD}_2$ under the hypothesis H_0 as well as under the alternative H_1 (see (2.3)), i.e., we focus on the asymptotic behavior of

$$\widehat{DoD}_{(\beta)} := \int_{\beta}^{1-\beta} \left(U_n^{-1}(t) - V_m^{-1}(t)\right)^2 dt, \tag{2.6}$$

where $\beta \in [0, 1/2)$ denotes a trimming parameter. It is important to note that many of our arguments and derivations can be generalized to $p \in [1, \infty)$. Based on the distributional limits derived, we design an asymptotic test for H_0 which we study empirically and successfully apply for protein structure comparison.

Assumptions: Before we come to the statements of the distributional limit theorems, we briefly discuss the corresponding assumptions. The first assumption we make is that the distributions of distances μ^U and μ^V admit Lebesgue densities u and v. Considering (2.6), we realize that $\widehat{DoD}_{(\beta)}$ is based on empirical U-quantile functions. It is well known that trimming generally simplifies the derivation of the asymptotics of quantile processes (Czado and Munk, 1998; Alvarez-Esteban et al., 2008). Hence, we distinguish between the cases $\beta \in (0, 1/2)$ and $\beta = 0$ in the following. The subsequent assumptions imply Hadamard differentiability of the inversion functional $\phi_{inv} : F \mapsto F^{-1}$ as a map from the set of restricted distribution functions into the space of all bounded functions on $[\beta, 1-\beta]$.

Condition 2.1. *Let $\beta \in (0, 1/2)$ and let U be continuously differentiable on an interval $[C_1, C_2] = [U^{-1}(\beta) - \epsilon, U^{-1}(1-\beta) + \epsilon]$ for some $\epsilon > 0$ with strictly positive derivative u and let the analogous assumption hold for V and its derivative v.*

Figure 2.1 illustrates the densities of the distributions of distances for several different metric measure spaces. We observe that these densities vanish at the boundaries of their support. Hence, it is not possible to derive the asymptotics of $\widehat{DoD}_{(0)}$ via the Hadamard differentiability of the inversion functional as done in the trimmed case. Instead, it is important to control how fast the densities u and v of U and V, respectively, vanish. This is achieved by the subsequent assumptions

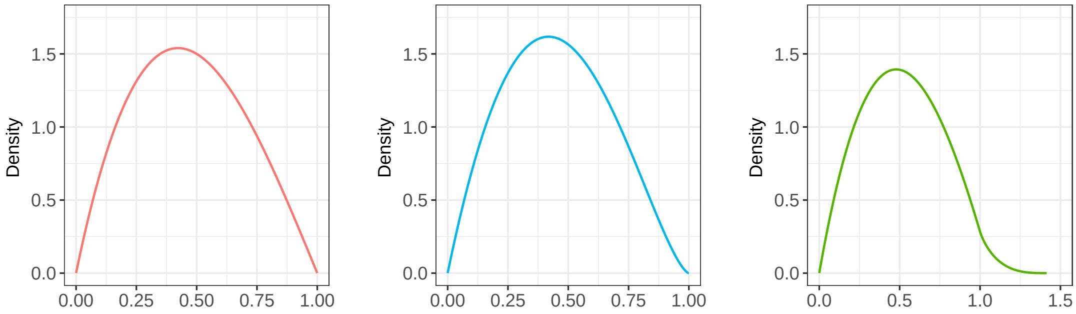

Figure 2.1: **Distribution of Distances:** The densities of the distribution of distances of $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ (red), where $\mathcal{X}$ denotes the disc in $\mathbb{R}^2$ with radius 0.5, $d_{\mathcal{X}}$ the Euclidean distance and $\mu_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$, $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ (blue), where $\mathcal{Y} = [0,1]^2$, $d_{\mathcal{Y}}$ corresponds to the supremum norm and $\mu_{\mathcal{Y}}$ is the uniform distribution on $\mathcal{Y}$, and $(\mathcal{Z}, d_{\mathcal{Z}}, \mu_{\mathcal{Z}})$ (green), where $\mathcal{Z} = [0,1]^2$, $d_{\mathcal{Z}}$ corresponds to the Euclidean distance and $\mu_{\mathcal{Z}}$ denotes the uniform distribution on $\mathcal{Z}$.

which are reminiscent of the ones employed in Mason (1984).

Condition 2.2. *Let U be continuously differentiable on its support. Further, assume there exist constants $-1 < \gamma_1, \gamma_2 < \infty$ and $c_U > 0$ such that $\left|(U^{-1})'(t)\right| \leq c_U t^{\gamma_1}(1-t)^{\gamma_2}$ for $t \in (0,1)$ and let the analogous assumptions hold for V and $(V^{-1})'$.*

Limit Distribution under H_0: Recall that the null hypothesis H_0 implies $\mu^U = \mu^V$. Under Condition 2.1 we obtain for $\beta \in (0, 1/2)$ (resp. under Condition 2.2 for $\beta = 0$) and $n, m \to \infty$ that

$$\frac{nm}{n+m}\widehat{DoD}_{(\beta)} \rightsquigarrow \Xi = \Xi(\beta) := \int_{\beta}^{1-\beta} (\mathbb{G}(t))^2 \, dt, \tag{2.7}$$

where "$\rightsquigarrow$" denotes weak convergence in the sense of Hoffman-Jørgensen (see Van der Vaart and Wellner (1996, Part 1)) and $\mathbb{G}$ denotes a centered Gaussian process with covariance given for $t, t' \in (0,1)$ by

$$\mathrm{Cov}\left(\mathbb{G}(t), \mathbb{G}(t')\right) = \frac{4}{(u \circ U^{-1}(t))(u \circ U^{-1}(t'))} \Gamma_{d_{\mathcal{X}}}(U^{-1}(t), U^{-1}(t')).$$

Here,

$$\begin{aligned}\Gamma_{d_{\mathcal{X}}}(t, t') = \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \leq t\}} \, d\mu_{\mathcal{X}}(y) \int \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \leq t'\}} \, d\mu_{\mathcal{X}}(y) \, d\mu_{\mathcal{X}}(x) \\ - \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \leq t\}} \, d\mu_{\mathcal{X}}(y) \, d\mu_{\mathcal{X}}(x) \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \leq t'\}} \, d\mu_{\mathcal{X}}(y) \, d\mu_{\mathcal{X}}(x).\end{aligned} \tag{2.8}$$

Limit Distribution under H_1: Given Condition 2.1 we obtain for $\beta \in (0, 1/2)$ (resp. given Condition 2.2 for $\beta = 0$) and $m, n \to \infty$ with $\frac{n}{n+m} \to \lambda \in (0,1)$ that

$$\sqrt{\frac{nm}{n+m}}\left(\widehat{DoD}_{(\beta)} - DoD_{(\beta)}\right) \rightsquigarrow N(0, \sigma^2_{U,V,\lambda}), \tag{2.9}$$

where

$$\sigma^2_{U,V,\lambda} = 16\lambda \int\limits_{U^{-1}(\beta)}^{U^{-1}(1-\beta)} \int\limits_{U^{-1}(\beta)}^{U^{-1}(1-\beta)} (x - V^{-1}(U(x)))(y - V^{-1}(U(y)))\Gamma_{d_{\mathcal{X}}}(x,y)\, dxdy$$
$$+16(1-\lambda) \int\limits_{V^{-1}(\beta)}^{V^{-1}(1-\beta)} \int\limits_{V^{-1}(\beta)}^{V^{-1}(1-\beta)} (U^{-1}(V(x)) - x)(U^{-1}(V(y)) - y)\Gamma_{d_{\mathcal{Y}}}(x,y)\, dxdy.$$

Here, $\Gamma_{d_{\mathcal{X}}}(x,y)$ is defined as in (2.8) and $\Gamma_{d_{\mathcal{Y}}}(x,y)$ is defined analogously. It is important to note that this normal distribution is degenerate if

$$DoD_{(\beta)} := \int_{\beta}^{1-\beta} \left(U^{-1}(t) - V^{-1}(t)\right)^2 dt = 0. \tag{2.10}$$

In this case, the asymptotic behavior of $\widehat{DoD}_{(\beta)}$ is given by (2.7).

Application: From (2.7) it immediately follows that the decision rule given by rejecting H_0 (see (2.3)) if

$$\frac{nm}{n+m}\widehat{DoD}_{(\beta)} > \xi_{1-\alpha}, \tag{2.11}$$

where $\xi_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of Ξ, defines a robust, asymptotic level α test for H_0 against H_1. In Paper A, we empirically illustrate its power in various settings. In particular, we showcase that it can be applied for 3D protein structure comparison, which is fundamental for developing an understanding of the functional and evolutionary relationships among proteins (Kolodny et al., 2005; Srivastava et al., 2016).

2.2 Discussion and Related Work

The idea to construct a test for the hypothesis H_0^* defined in (2.1) based on a polynomial time computable lower bound has already been pursued by Brécheteau (2019). In this work, the author defines the *Distance-to-Measure signature (DTM-signature)* as follows: Consider a metric measure space $\mathcal{X} = (\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and define $G_x(t) = \mathbb{P}(d_{\mathcal{X}}(x, X) \le t)$, where $x \in \mathcal{X}$ and $X \sim \mu_{\mathcal{X}}$. Then, its *Distance-to-Measure function (DTM-function) with mass parameter* $\kappa \in [0,1]$ is for $x \in \mathcal{X}$ given as

$$\mathrm{d}_{\mathcal{X},\kappa}(x) := \frac{1}{\kappa}\int_0^{\kappa} G_x^{-1}(l)\, dl. \tag{2.12}$$

To give some intuition, we remark that the value of the DTM-function can be interpreted as the mean distance of x to its "$100 \cdot \kappa$% nearest neighbors" in $\mathcal{X}$. Based on this, the DTM-signature of $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ is defined as the random variable

$$D_{\mathcal{X},\kappa}(\mu_{\mathcal{X}}) = D_{\mathcal{X},\kappa} := \mathrm{d}_{\mathcal{X},\kappa}(X),$$

where $X \sim \mu_{\mathcal{X}}$. Consequently, the DTM-signature corresponds to a distribution of mean distances and the parameter κ can be interpreted as a scale parameter. For small κ the signature $D_{\mathcal{X},\kappa}$ only incorporates local information whereas for $\kappa = 1$ also global distances are considered.

It turns out that also the DTM-singature is stable with respect to $d_{\mathrm{GW},1}$. More precisely, we demonstrate in Section B.7 in the supplement of Paper A that

$$\frac{2}{\kappa} d_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y}) \geq \frac{2}{\kappa}\mathbf{TLB}_1(\mathcal{X},\mathcal{Y}) \geq T_\kappa(\mathcal{X},\mathcal{Y}) := d_{\mathrm{W},1}^{(\mathbb{R},|\cdot|)}(D_{\mathcal{X},\kappa}, D_{\mathcal{Y},\kappa}).$$

Furthermore, we prove that the lower bound T_κ is closely related to $\mathbf{FLB}_1$ and in particular that $T_1(\mathcal{X},\mathcal{Y}) = 2\mathbf{FLB}_1(\mathcal{X},\mathcal{Y})$ (see Sec. B.7 in the supplement of Paper A). Given two samples $\mathcal{X}_n = \{X_1,\dots,X_n\}$ and $\mathcal{Y}_n = \{Y_1,\dots,Y_n\}$ from the spaces $\mathcal{X}$ and $\mathcal{Y}$, Brécheteau (2019) derives the distributional limit of

$$\sqrt{n_S}\,\widehat{T}_\kappa(\mathcal{X}_n,\mathcal{Y}_n) := \sqrt{n_S}\, d_{\mathrm{W},1}^{(\mathbb{R},|\cdot|)}\left(\frac{1}{n_S}\sum_{i=1}^{n_S}\delta_{\widehat{d}_{\mathcal{X},\kappa}(X_i)}, \frac{1}{n_S}\sum_{i=1}^{n_S}\delta_{\widehat{d}_{\mathcal{Y},\kappa}(Y_i)}\right),$$

under H_0^* as $n \to \infty$, where $n_S \in o(n)$,

$$\widehat{d}_{\mathcal{X},\kappa}(x) := \frac{1}{\kappa}\int_0^\kappa \widehat{G}_{n,x}^{-1}(l)\, dl \tag{2.13}$$

and $\widehat{d}_{\mathcal{Y},\kappa}$ is defined analogously. Here,

$$\widehat{G}_{n,x}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{d_{\mathcal{X}}(x,X_i)\leq t\}}.$$

In Paper A, we demonstrate that, while the additional subsampling simplifies the derivations of the asymptotics of $\widehat{T}_\kappa$, it reduces the power of the corresponding test. In particular, we showcase empirically that the test based on (2.7), where we handle the occurring dependencies carefully, obtains more power in various settings. From a more practical side, Paper A is strongly related to Gellert et al. (2019). In their paper, the authors successfully apply the lower bounds $\mathbf{FLB}_p$, $\mathbf{SLB}_p$ and $\mathbf{TLB}_p$ defined in Section 1.3 for the comparison of the isosurfaces of various proteins.

Reconsidering the definition of $\widehat{DoD}_{(\beta)}$ in (2.6), we observe that the proposed test statistic can also be interpreted as one dimensional (trimmed) Wasserstein distance between the empirical measures μ^{U_n} and μ^{V_m} based on $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\leq i<j\leq n}$ and $\{d_{\mathcal{Y}}(Y_k,Y_l)\}_{1\leq k<l\leq m}$, respectively. Distributional limits for the one dimensional empirical Wasserstein distance have already been derived in a variety of settings (Munk and Czado, 1998; del Barrio et al., 1999, 2005; Dede, 2009; Bobkov and Ledoux, 2019; Dedecker and Merlevede, 2017; Hundrieser et al., 2022). However, it is important to note that the random variables in the samples $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\leq i<j\leq n}$ and $\{d_{\mathcal{Y}}(Y_k,Y_l)\}_{1\leq k<l\leq m}$ exhibit a specific dependency structure that has, to the best of our knowledge, so far not been considered in the context of empirical Wasserstein distances. This dependency structure is related to U- and U-quantile processes (Nolan and Pollard, 1988; Arcones and Giné, 1994; Wendler, 2012). Indeed, the theory

developed by Nolan and Pollard (1988) allows to immediately derive the asymptotics of $\widehat{DoD}_p$ in the special case $p = 1$. Furthermore, the results of Wendler (2012) on U-quantile processes can be used to derive the asymptotics of $\widehat{DoD}_{(\beta)}$ for $\beta > 0$ under slightly stronger assumptions. For the theoretically more involved case $\beta = 0$, the results derived by Wendler (2012) are not applicable and we pursue completely different approaches for the derivation of (2.7) and (2.9).

Conclusion: We study the asymptotic behavior of $\widehat{DoD}_{(\beta)}$, $\beta \in [0, 1/2)$, under the hypothesis H_0 as well as under the alternative H_1. This leads to the construction of an asymptotic level α test for H_0^* against H_1^* (see (2.1)). This test is studied empirically and shows promising performance. Our findings suggest that, in many applications, it is sufficient to work with the computationally more tractable lower bounds of the Gromov-Wasserstein distance instead of said metric. This opens many possible directions of research. In particular, it might be of interest to investigate the asymptotics of the slightly stronger (but computationally more demanding) lower bound $\mathbf{TLB}_p$, to use them to construct a test for H_0^* and to compare the practical performance of this test to the one proposed in (2.11).

CHAPTER 3

The Ultrametric Gromov-Wasserstein Distance

The main results of Paper B are discussed in this chapter. For this purpose, we briefly recall and motivate the issues considered.

As mentioned in Section 1.2, the Gromov-Wasserstein is a useful tool for various data analysis tasks. Unfortunately, its precise determination is in general infeasible. However, the collection of all isomorphism classes of metric measure spaces $\mathcal{M}^w$ contains a huge variety of spaces and there might be a subclass $\mathcal{O}^w \subset \mathcal{M}^w$ for which it is possible to determine (variants of) $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ or $d_{\mathrm{GW},p}$ in polynomial time. Further, it might be possible to adjust the definitions of Sturm's/the Gromov-Wasserstein distance to $\mathcal{O}^w$ in order to obtain more informative metrics on this subclass. Naturally, it is of great interest to identify such subclasses and adjustments.

Similar ideas have been applied in the study of the Gromov-Hausdorff distance and led to the definition of the *ultrametric Gromov-Hausdorff distance* on the collection of compact *ultrametric spaces* (Zarichnyi, 2005; Qiu, 2009; Mémoli et al., 2021b). Recall that a metric space $(\mathbb{X}, u_{\mathbb{X}})$ is denoted as ultrametric, if for all $x, x', x'' \in \mathbb{X}$ the subsequent relation is fulfilled

$$u_{\mathbb{X}}(x, x'') \leq \max(u_{\mathbb{X}}(x, x'), u_{\mathbb{X}}(x', x'')). \tag{3.1}$$

Throughout the following, $\mathcal{U}$ denotes the isometry classes of compact ultrametric measure spaces. Ultrametric spaces arise naturally in various applications, for instance as metric encodings of dendrograms (Jardine and Sibson, 1971; Carlsson and Mémoli, 2010), in the context of phylogenetic trees (Semple et al., 2003), in the probabilistic approximation of finite metric spaces (Bartal, 1996; Fakcharoenphol et al., 2004) or in the context of a mean-field theory of spin glasses (Mézard et al., 1987; Rammal et al., 1986).
Reconsidering the definition of the Gromov-Hausdorff distance in (1.1) it is clear that the ultrametric structure of two ultrametric spaces is lost, if we minimize over all possible embeddings into a general metric space $(\mathbb{Z}, d_{\mathbb{Z}})$. In order to preserve this structure, the ultrametric Gromov-Hausdorff distance only considers embeddings into a common ultrametric space $(\mathbb{Z}, u_{\mathbb{Z}})$, i.e., for two ultrametric spaces

$(\mathbb{X}, u_{\mathbb{X}})$ and $(\mathbb{Y}, u_{\mathbb{Y}})$ it is defined as

$$u_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) := \inf_{\mathbb{Z},\phi,\psi} d_{\mathrm{H}}^{(\mathbb{Z},u_{\mathbb{Z}})}(\phi(\mathbb{X}), \psi(\mathbb{Y})), \tag{3.2}$$

where $\phi : \mathbb{X} \to \mathbb{Z}$ and $\psi : \mathbb{Y} \to \mathbb{Z}$ are isometric embeddings into a common ultrametric space $(\mathbb{Z}, u_{\mathbb{Z}})$ and $d_{\mathrm{H}}^{(\mathbb{Z},u_{\mathbb{Z}})}$ denotes the Hausdorff distance on $\mathbb{Z}$. Similar as for d_{GH} it is possible to prove that u_{GH} constitutes an ultrametric on the space $\mathcal{U}$. Furthermore, it is important to note that it is possible to reformulate u_{GH}, just as d_{GH}, as an optimization problem over the set of metric couplings (Mémoli et al., 2021b). More precisely, define for $a, b \in \mathbb{R}$

$$\Lambda_\infty(a, b) := \begin{cases} \max(a, b) & \text{if } a \neq b, \\ 0 & \text{if } a = b. \end{cases} \tag{3.3}$$

Then, an equivalent formulation of u_{GH} between the compact ultrametric spaces $\mathbb{X}$ and $\mathbb{Y}$ is given as

$$u_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) = \inf_{R \in \mathcal{R}(\mathbb{X},\mathbb{Y})} \sup_{\substack{x_1,x_2 \in \mathbb{X} \\ y_1,y_2 \in \mathbb{Y} \\ s.t.\ (x_i,y_i) \in R}} \Lambda_\infty\left(u_{\mathbb{X}}(x_1, x_2), u_{\mathbb{Y}}(y_1, y_2)\right). \tag{3.4}$$

Based on this reformulation it is possible to derive a polynomial time algorithm for the computation of u_{GH} (Mémoli et al., 2021b).

Inspired by the results on the ultrametric Gromov-Hausdorff distance and the relation between d_{GH} and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ as well as $d_{\mathrm{GW},p}$, we aim to study ultrametric variants of these distances on the collection of the isomorphism classes of ultrametric measure measure spaces $\mathcal{U}^w \subset \mathcal{M}^w$. Here, an ultrametric measure space $\mathcal{X} = (X, u_X, \mu_X)$ is a metric measure space, where the corresponding metric space (X, u_X) is ultrametric. Reconsidering the definition of the ultrametric Gromov-Hausdorff distance in (3.2), we propose to mimic its construction and to only infimize over ultrametric spaces $(\mathbb{Z}, u_{\mathbb{Z}})$ instead of all possible metric spaces in (1.7). Thus, we define for $p \in [1, \infty]$ *Sturm's ultrametric Gromov-Wasserstein distance* of order p as

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) := \inf_{\mathbb{Z},\phi,\psi} d_{\mathrm{W},p}^{(\mathbb{Z},d_{\mathbb{Z}})}(\phi_\# \mu_X, \psi_\# \mu_Y), \tag{3.5}$$

where $\phi : X \to \mathbb{Z}$ and $\psi : Y \to \mathbb{Z}$ are isometric embeddings into an ultrametric space $(\mathbb{Z}, u_{\mathbb{Z}})$.

Furthermore, we realize that it is also possible to redo the construction of the Gromov-Wasserstein distance based on the equivalent formulation of u_{GH} in (3.4) (see Section 1.2 for an illustration of this construction). This leads to the definition of the *p-ultra-distortion* of a coupling $\pi \in C(\mu_X, \mu_Y)$ that is for $1 \leq p < \infty$ given as

$$\mathrm{dis}_p^{\mathrm{ult}}(\pi) := \left(\iint_{X \times Y \times X \times Y} \left(\Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right)^p \pi(dx \times dy)\, \pi(dx' \times dy') \right)^{1/p} \tag{3.6}$$

and for $p = \infty$ as

$$\operatorname{dis}_{\infty}^{\text{ult}}(\pi) := \sup_{\substack{x,x' \in \mathcal{X},\, y,y' \in \mathcal{Y} \\ s.t.\ (x,y),(x',y') \in \operatorname{supp}(\pi)}} \Lambda_{\infty}(u_{\mathcal{X}}(x,x'), u_{\mathcal{Y}}(y,y')).$$

Based on this, the *ultrametric Gromov-Wasserstein distance* of order $p \in [1, \infty]$, is defined as

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) := \inf_{\pi \in C(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} \operatorname{dis}_p^{\text{ult}}(\pi). \tag{3.7}$$

3.1 Main Results

The main contributions of Paper B are the proofs that both Strum's ultrametric Gromov-Wasserstein distance and the ultrametric Gromov-Wasserstein distance constitute metrics on $\mathcal{U}^w$, the derivation of several properties of these metrics, a polynomial time algorithm for the calculation of $u_{\mathrm{GW},\infty}^{\text{sturm}} = u_{\mathrm{GW},\infty}$, the derivation of polynomial time computable lower bounds for $u_{\mathrm{GW},p}$, $1 \leq p < \infty$, as well as an empirical illustration of the distance captured by $u_{\mathrm{GW},p}$ for synthetic and real data.

Properties of the Metrics: Recall that a metric space $(\mathbb{X}, d_{\mathbb{X}})$ is called *p-metric space* for $p \in [1, \infty)$ if it fulfills the subsequent stronger version of the triangle inequality:

$$d_{\mathbb{X}}(x, x'') \leq \left(d_{\mathbb{X}}(x, x')^p + d_{\mathbb{X}}(x', x'')^p\right)^{1/p},$$

for all $x, x', x'' \in \mathbb{X}$. Note that ultrametric spaces can be understood as a limit case of p-metric spaces ($p \to \infty$). Our first main result is the fact that both $u_{\mathrm{GW},p}^{\text{sturm}}$ and $u_{\mathrm{GW},p}$ are p-metrics on $\mathcal{U}^w$ that induce different topologies than $d_{\mathrm{GW},p}^{\text{sturm}}$ and $d_{\mathrm{GW},p}$. Further, we show that $2^{-1/p} u_{\mathrm{GW},p} \leq u_{\mathrm{GW},p}^{\text{sturm}}$, that $u_{\mathrm{GW},p}^{\text{sturm}}$ and $u_{\mathrm{GW},p}$ are topologically equivalent for $1 \leq p < \infty$ and that $u_{\mathrm{GW},\infty}^{\text{sturm}} = u_{\mathrm{GW},\infty}$.

Computational Aspects: Similar as for the ultrametric Gromov-Hausdorff distance, we derive a polynomial time computable algorithm for the calculation of $u_{\mathrm{GW},\infty}$ $(= u_{\mathrm{GW},\infty}^{\text{sturm}})$ based on the *weighted quotients* of the considered ultrametric measure spaces. Let $\mathcal{X} = (\mathcal{X}, u_{\mathcal{X}}, \mu_{\mathcal{X}})$ be an ultrametric measure space. Then, it is possible to define for any $t > 0$ an equivalence relation $\sim_t$ on $(\mathcal{X}, u_{\mathcal{X}})$ as follows: $x \sim_t x'$ if and only if $u_{\mathcal{X}}(x, x') \leq t$. For $x \in \mathcal{X}$, the equivalence class of x with respect to $\sim_t$ is denoted as $[x]_t^{\mathcal{X}}$ (or simply $[x]_t$ if the corresponding ultrametric measure space is clear from the context). Further, let $\mathcal{X}_t$ denote the collection of equivalence classes of $\mathcal{X}$ under $\sim_t$. There is a canonical way to turn $\mathcal{X}_t$ into an ultrametric measure space: We observe that the subsequent definition yields an ultrametric on $\mathcal{X}_t$

$$u_{\mathcal{X}_t}([x]_t, [x']_t) := \begin{cases} u_{\mathcal{X}}(x, x'), & [x]_t \neq [x']_t \\ 0, & [x]_t = [x']_t. \end{cases}$$

Furthermore, denote by $Q_t : (X, u_X) \to (X_t, u_{X_t})$ the quotient map sending $x \in \mathcal{X}$ to $[x]_t$ and define $\mu_{\mathcal{X}_t} = Q_t \# \mu_{\mathcal{X}}$. Then, the ultrametric measure space $\mathcal{X}_t = (\mathcal{X}_t, u_{\mathcal{X}_t}, \mu_{\mathcal{X}_t})$ is the *weighted quotient* of $\mathcal{X}$ at level t. Figure 3.1 illustrates the weighted quotient in a simple example. Let $\mathcal{X} = (\mathcal{X}, u_{\mathcal{X}}, \mu_{\mathcal{X}})$ and

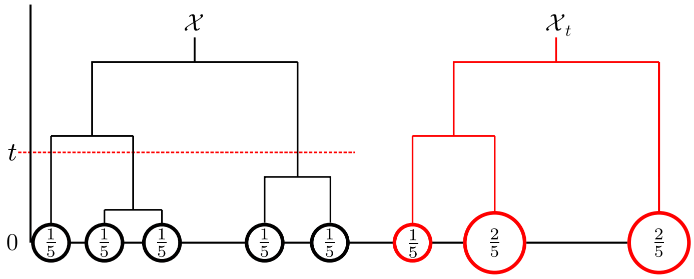

Figure 3.1: **Weighted Quotient:** An ultrametric measure space (black) and its weighted quotient at level t (red).

$\mathcal{Y} = (\mathcal{Y}, u_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ denote two compact ultrametric measure spaces. Then, it follows that

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \min\{t \geq 0 \,|\, \mathcal{X}_t \cong \mathcal{Y}_t\},$$

which can be used to define a polynomial time algorithm for the computation of $u_{\mathrm{GW},\infty}$ on the basis of tree isomorphisms (see Section 5 in Paper B for more details).

Unfortunately, the computation of $u^{\mathrm{sturm}}_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}$ for $1 \leq p < \infty$ requires solving complicated (combinatorial) optimization problems that are closely related to various NP-hard problems. Nevertheless, it is possible to approximate local minima of the optimization problem underlying $u_{\mathrm{GW},p}$ via conditional gradient descent. Furthermore, we derive two lower bounds for $u_{\mathrm{GW},p}$ that are structurally extremely similar to the ones of $d_{\mathrm{GW},p}$. More precisely, we show that for two ultrametric measure spaces $\mathcal{X} = (\mathcal{X}, u_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $\mathcal{Y} = (\mathcal{Y}, u_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ it holds that

$$u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq \mathbf{TLB}^{\mathrm{ult}}_p(\mathcal{X},\mathcal{Y}) \geq \mathbf{SLB}^{\mathrm{ult}}_p(\mathcal{X},\mathcal{Y}),$$

where

$$\mathbf{TLB}^{\mathrm{ult}}_p(\mathcal{X},\mathcal{Y}) = \inf_{\pi \in \mathcal{C}(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \left(\int_{\mathcal{X}\times\mathcal{Y}} \left(d^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}_{\mathrm{W},p} \left(u_X(x,\cdot)_{\#}\mu_{\mathcal{X}}, u_Y(y,\cdot)_{\#}\mu_{\mathcal{Y}} \right) \right)^p \, d\pi(x,y) \right)^{1/p} \tag{3.8}$$

and

$$\mathbf{SLB}^{\mathrm{ult}}_p(\mathcal{X},\mathcal{Y}) = d^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}_{\mathrm{W},p} \left((u_{\mathcal{X}})_{\#}(\mu_{\mathcal{X}} \otimes \mu_{\mathcal{X}}), (u_{\mathcal{Y}})_{\#}(\mu_{\mathcal{Y}} \otimes \mu_{\mathcal{Y}}) \right). \tag{3.9}$$

Comparing the lower bounds $\mathbf{TLB}^{\mathrm{ult}}_p$ and $\mathbf{TLB}_p$ as well as $\mathbf{SLB}^{\mathrm{ult}}_p$ and $\mathbf{SLB}_p$, we realize that for both cases the Wasserstein distance on the metric space $(\mathbb{R}, |\cdot|)$ is replaced by the one on the ultrametric space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$. Since also $d^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}_{\mathrm{W},p}$ has an explicit representation (see Section 2 in Paper B), it is possible to calculate $\mathbf{TLB}^{\mathrm{ult}}_p$ and $\mathbf{SLB}^{\mathrm{ult}}_p$ efficiently.

Application: In a simulation study we empirically showcase that $d_{\mathrm{GW},1}$ and $u_{\mathrm{GW},1}$ react very differently to various disturbances of the considered ultrametric measure space. In particular, we find that the ultrametric Gromov-Wasserstein distance is much more sensitive to large scale disturbances, e.g., changes in the diameter, than $d_{\mathrm{GW},p}$.
Moreover, we asses the lower bounds $\mathbf{SLB}_1$ as well as $\mathbf{SLB}^{\mathrm{ult}}_1$ for the comparison of *phylogenetic*

tree shapes, i.e., the comparison of the trees' connectivity structures without referring to their labels or the length of their branches. It is well known that these shapes carry important information about macroevolutionary processes (Mooers and Heard, 1997; Blum and François, 2006; Dayarian and Shraiman, 2014; Wu and Choi, 2016) and can be used to study and compare the evolution of different pathogens. More precisely, we repeat a comparison from Colijn and Plazzotta (2018) and compare two sets of different phylogenetic tree shapes based on HA protein sequences from human influenza A (H3N2) from different regions. The results show that $\mathbf{SLB}_1^{\mathrm{ult}}$ as well as $\mathbf{SLB}_p$ are well suited for this kind of task.

3.2 Discussion and Related Work

By construction $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ are closely related to $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$ as well as the ultrametric Gromov-Hausdorff distance u_{GH}. Sturm's Gromov-Wasserstein distance is studied in Sturm (2006, 2012) and the Gromov-Wasserstein distance is investigated by Mémoli (2007, 2011); Chowdhury and Mémoli (2019). Furthermore, the ultrametric Gromov-Hausdorff distance u_{GH} is introduced in Zarichnyi (2005), its theoretical properties are studied in Qiu (2009) and a polynomial time algorithm for its computation is devised in Mémoli et al. (2021b). Additionally, Mémoli and Wan (2019) study a variant of the Gromov-Hausdorff metric on the collection of all p-metric spaces, $1 \leq p \leq \infty$, that coincides for $p = \infty$ with u_{GH}.

Similar in spirit to our work are Evans (2007), who describes some variants of the Gromov-Hausdorff distance between metric trees, and Greven et al. (2009), who introduces metric measure space representations of trees and a certain Gromov-Prokhorov type of metric on the collection of these representations.

In Kloeckner (2015), the author derives an explicit formulation of the Wasserstein distance on a given ultrametric space. This has been another motivation for studying Sturm's ultrametric Gromov-Wasserstein distance. Furthermore, it allows us to obtain an explicit formulation of $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ (it is easy to check that the space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ is ultrametric) which we use to reduce the computational complexity of the lower bounds $\mathbf{SLB}_p^{\mathrm{ult}}$ and $\mathbf{TLB}_p^{\mathrm{ult}}$.

From a practical point of view our work is related to Poon et al. (2013); Lewitus and Morlon (2016); Colijn and Plazzotta (2018); Liu et al. (2020); Kim et al. (2020), where the authors propose and study different metrics and dissimilarities on the collections of phylogenetic tree shapes. Recall that we have empirically compared our approach with the one proposed by Colijn and Plazzotta (2018).

Conclusion: In this work, we introduce and study Sturm's ultrametric Gromov-Wasserstein distance $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ as well as the ultrametric Gromov-Wasserstein distance $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. In particular, we showcase that these metrics induce equivalent typologies on $\mathcal{U}^w$ that differ from the one induced by $d_{\mathrm{GW},p}$ (which is equivalent to the one induced by $d_{\mathrm{GW},p}^{\mathrm{sturm}}$). Moreover, we demonstrate that it is possible to calculate $u_{\mathrm{GW},\infty}$ ($= u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$) in polynomial time and explore the potential of $\mathbf{SLB}_p$ and $\mathbf{SLB}_p^{\mathrm{ult}}$ for the comparison of phylogenetic tree shapes. Due to its computational complexity, the search for classes of metric measure spaces, where (a variant of) the Gromov-Wasserstein distance admits a polynomial time algorithm or an explicit solution remains an interesting question for research.

CHAPTER 4

Distance-to-Measure Density based Geometric Analysis of Complex Data

In this last chapter of the main part, we illustrate the approach pursued in Paper C and outline the main results.

In numerous applications it is of great interest to analyze and classify noisy point clouds on the basis of their small scale characteristics without taking large scale information (such as the over all shape of the point cloud) into account (Vosselman et al., 2004; Hayakawa and Oguchi, 2016; Libeskind et al., 2017). This is particularly true for the analysis of human chromatin loops based on 3D single molecule localization microscopy (SMLM) data, where the goal is to track the degrading of naturally appearing chromatin loops (Hao et al., 2021). We propose a signature based on (a version of) the Distance-to-Measure signature (see Section 2.2) to tackle this type of problems. More precisely, we suggest to consider the point clouds in $\mathbb{R}^d$ as a random sample from a *Euclidean metric measure space* $\mathcal{X} = (\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$, i.e., a metric measure space where $\mathcal{X} \subset \mathbb{R}^d$ and $\|\cdot\|$ stands for the Euclidean distance. Then, it is possible to map each point cloud considered to an estimate of the density of the corresponding DTM-signature (in the following referred to as *DTM-density*) which then can be used for many analysis tasks (e.g., classification).

For the sake of completeness, we introduce the DTM-function of an Euclidean metric measure space next. Let $\mathcal{X}$ be an Euclidean metric measure space. We define the *Distance-to-Measure (DTM) function* with *mass parameter* $\kappa \in (0, 1]$ corresponding to $\mathcal{X}$ for $x \in \mathbb{R}^d$ as

$$\mathrm{d}_{\mathcal{X},\kappa}(x) = \frac{1}{\kappa} \int_0^{\kappa} F_x^{-1}(u)\, du, \tag{4.1}$$

where $F_x(t) = P\left(\|X - x\|^2 \leq t\right)$, $X \sim \mu_{\mathcal{X}}$, and F_x^{-1} stands for the quantile function of F_x. It is important to note that the above definition differs slightly from (2.12), as we consider squared Euclidean distances (i.e., we use the definition of the DTM-function considered in Chazal et al. (2017)). We recall that the DTM-signature is built upon the DTM-function, which has to be estimated from the data in practice. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$, let

$$\widehat{F}_{x,n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\|x-X_i\|^2 \leq t\}}$$

and denote by $\widehat{F}_{x,n}^{-1}$ the corresponding quantile function. Then, it is possible to define (similar as in Section 2.2) a plug-in estimator for $\mathrm{d}_{\mathcal{X},\kappa}(x)$ by replacing F_x^{-1} with $\widehat{F}_{x,n}^{-1}$. This yields

$$\widehat{d}_{\mathcal{X},\kappa}(x) = \frac{1}{\kappa}\int_0^{\kappa} \widehat{F}_{x,n}^{-1}(u)\, du. \tag{4.2}$$

For $\kappa = \frac{k}{n}$, we can rewrite $\widehat{d}_{\mathcal{X},\kappa}$ as a nearest neighbor mean as follows

$$\widehat{d}_{\mathcal{X},\kappa}(x) = \frac{1}{k}\sum_{X_i \in N_k(x)} \|X_i - x\|^2, \tag{4.3}$$

where the set $N_k(x)$ consists of the k nearest neighbors of x among the data points $X_1, \ldots, X_n$. Recall that the DTM-signature is given as $\mathrm{d}_{\mathcal{X},\kappa}(X)$, where $X \sim \mu_{\mathcal{X}}$. Throughout the following, we will assume that $\mathrm{d}_{\mathcal{X},\kappa}(X)$ admits a Lebesgue density $f_{\mathrm{d}_{\mathcal{X},\kappa}}$. In Section 2.2 we have seen that κ can be viewed as a scale parameter: Choosing κ small corresponds to considering only local neighborhoods, whereas $\kappa = 1$ also incorporates global information (this is also highlighted by (4.3)). As previously mentioned, we propose to apply the density of $\mathrm{d}_{\mathcal{X},\kappa}(X)$ as a signature for data analysis. If $\mathrm{d}_{\mathcal{X},\kappa}$ was known, we could estimate this density via the kernel density estimator

$$\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{d}_{\mathcal{X},\kappa}(X_i) - y}{h}\right). \tag{4.4}$$

However, we have already argued that also $\mathrm{d}_{\mathcal{X},\kappa}$ has to be estimated from the data, as it is usually unknown. Replacing $\mathrm{d}_{\mathcal{X},\kappa}$ with $\widehat{d}_{\mathcal{X},\kappa}$ in (4.4) yields the following estimator for $f_{\mathrm{d}_{\mathcal{X},\kappa}}$

$$\widehat{f}_{\widehat{d}_{\mathcal{X},\kappa}}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\widehat{d}_{\mathcal{X},\kappa}(X_i) - y}{h}\right). \tag{4.5}$$

An important observation to make is the fact that the random variables $\widehat{d}_{\mathcal{X},\kappa}(X_i)$ and $\widehat{d}_{\mathcal{X},\kappa}(X_j)$ are stochastically dependent for each $i \neq j$, i.e., $\widehat{f}_{\widehat{d}_{\mathcal{X},\kappa}}$ is, different from $\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}$, based on dependent data.

4.1 Main Results

Since the kernel density estimator $\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}$ is based on independent data, it is well known that (given some standard assumptions) for $n \to \infty$ and $h = o\left(n^{-1/5}\right)$ with $nh \to \infty$

$$\sqrt{nh}\left(\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}(y) - f_{\mathrm{d}_{\mathcal{X},\kappa}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}_{\mathcal{X},\kappa}}(y)\int_{\mathbb{R}} K^2(u)\, du\right),$$

where "⇒" stands for the usual weak convergence. In Paper C, we derive an analogous statement for $\widehat{f}_{d_{\mathcal{X},\kappa}}$. Furthermore, we illustrate that (estimates of) DTM-densities can be used as stable signatures for pose invariant object discrimination and that they are particularly useful for the analysis and comparison of small scale characteristics.

Assumptions: Before we illustrate the pointwise limit theorem for the kernel density estimator $\widehat{f}_{d_{\mathcal{X},\kappa}}$, we briefly discuss the corresponding assumptions. It is noteworthy that the assumption that the DTM-signature $\mathrm{d}_{\mathcal{X},\kappa}(X)$, $X \sim \mu_{\mathcal{X}}$, admits a Lebesgue density is for $\kappa < 1$ slightly restrictive. If the considered metric measure space $\mathcal{X}$ has little to no local structure, it is possible that the measure corresponding to $\mathrm{d}_{\mathcal{X},\kappa}(X)$, $X \sim \mu_{\mathcal{X}}$, admits a pure point component if κ is too small. For instance, if $\mathcal{X} = (\mathcal{X}, \|\cdot\|, \mu_{\mathcal{X}})$, where $\mathcal{X}$ denotes the unit disc in $\mathbb{R}^2$ and $\mu_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$, then only $\mathrm{d}_{\mathcal{X},1}(X)$ is absolutely continuous with respect to the Lebesgue measure. Exemplarily, we highlight in Paper C that these cases are rather pathological and of little practical relevance.
Furthermore, it is obvious that the sole existence of $f_{\mathrm{d}_{\mathcal{X},\kappa}}$ is not sufficient to derive the asymptotics of $\sqrt{nh}\left(\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}(y) - f_{\mathrm{d}_{\mathcal{X},\kappa}}(y)\right)$. Additional to some standard assumptions on $f_{\mathrm{d}_{\mathcal{X},\kappa}}$ as well as the corresponding kernel, we require some regularity of the level sets of $\mathrm{d}_{\mathcal{X},\kappa}$ as well as the underlying set $\mathcal{X}$. Before we state these assumptions, we recall some properties of level sets and introduce the *Hausdorff measure*.

Let $U \subset \mathbb{R}^d$ be open and $g : U \to \mathbb{R}$ be k-times continuously differentiable. Suppose that $c \in \mathbb{R}$ such that $g^{-1}(\{c\}) \neq \emptyset$ and $\nabla g \neq 0$ on $g^{-1}(\{c\})$. Then, $g^{-1}(\{c\})$ is a C^k-manifold of dimension $d-1$ (Rudolph and Schmidt, 2012, Thm. 1.2.1). The intrinsic volume of $g^{-1}(\{c\})$ can be determined via the $(d-1)$-dimensional Hausdorff measure. This outer measure can be defined for all subsets of $\mathbb{R}^d$ and is fundamental to various areas of geometric measure theory (for more information on the subject see Federer (1969); Morgan (2016)). Given a set $A \subset \mathbb{R}^d$, the *k-dimensional Hausdorff measure* $\mathscr{H}^k(A)$ is given by

$$\mathscr{H}^k(A) := \lim_{\delta\to 0} \inf_{\substack{A\subseteq\bigcup S_i,\\ \mathrm{diam}(S_i)\leq\delta}} \sum \alpha_k \left(\frac{\mathrm{diam}\,(S_i)}{2}\right)^k,$$

where the infimum is taken over all countable coverings S_i of A with $\mathrm{diam}\,(S_i) < \delta$ and α_k stands for the volume of the unit ball in $\mathbb{R}^k$. We note that the restriction of $\mathscr{H}^k$ to the Borel sets of $\mathbb{R}^d$ yields a Borel measure (Morgan, 2016, Sec.2).
The final fact about level sets stated here concerns the relation of $g^{-1}(\{y\})$ and $g^{-1}(\{y+v\})$ for v small. Suppose that there exists an open set U and $h_0' > 0$ such that $g^{-1}([y-h_0', y+h_0']) \subset U$ and such that the function

$$\varphi(u) : U \subset \mathbb{R}^d \to \mathbb{R}^d,\ u \mapsto \frac{\nabla g(u)}{\|\nabla g(u)\|^2}$$

is continuously differentiable. By Cauchy-Lipschitz's theory (Hirsch and Smale, 1974; Amann,

2011) there exists $0 < h_0 \leq h_0'$ such that one can construct a flow $\Phi : [-h_0, h_0] \times W \to \mathbb{R}^d$ with

$$\begin{cases} \frac{\partial}{\partial t}\Phi(t,x) = \frac{\nabla g(\Phi(t,x))}{\|\nabla g(\Phi(t,x))\|^2} \\ \Phi(0,x) = x, \end{cases}$$

where $W \subset \mathbb{R}^d$ is an open set that contains $g^{-1}([y-h_0, y+h_0])$. Differentiating the function $t \mapsto g(\Phi(t,x))$ immediately shows that $g(\Phi(t,x)) = g(x)+t$. This implies that $\Phi(t, g^{-1}(\{y\})) = g^{-1}(\{y+t\})$. In particular, $\{\Phi(v,\cdot)\}_{v\in[-h_0,h_0]}$ constitutes a one parameter family of C^1-diffeomorphisms between $g^{-1}(\{y\})$ and $g^{-1}(\{y+v\})$, $v \in [-h_0, h_0]$. This family is in the following referred to as *canonical level set flow of* $g^{-1}(\{y\})$.

Condition 4.1. *Let* $f_{\mathrm{d}_{\mathcal{X},\kappa}}$ *be supported on* $[D_1, D_2]$ *and let* $y \in [D_1, D_2]$. *Assume that there exists* $\epsilon > 0$ *such that* $f_{\mathrm{d}_{\mathcal{X},\kappa}}$ *is twice continuously differentiable on* $(y-\epsilon, y+\epsilon)$. *Further, suppose that the function* $\mathrm{d}_{\mathcal{X},\kappa} : \mathbb{R}^d \to \mathbb{R}$ *is* $C^{2,1}$ *on an open neighborhood of the level set*

$$\Gamma_y := \mathrm{d}_{\mathcal{X},\kappa}{}^{-1}(\{y\}) = \{x \in \mathbb{R}^d : \mathrm{d}_{\mathcal{X},\kappa}(x) = y\},$$

that $\nabla \mathrm{d}_{\mathcal{X},\kappa} \neq 0$ *on* Γ_y *and that there exists* $h_0 > 0$ *such that for all* $-h_0 < v < h_0$,

$$\mathcal{I}_{\mathcal{X}}(y;v) := \int_{\Gamma_y} \left| \mathbb{1}_{\{x\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| \, d\mathcal{H}^{d-1}(x) \leq C_y|v|, \tag{4.6}$$

where $\{\Phi(v,\cdot)\}_{v\in[-h_0,h_0]}$ *denotes the canonical level set flow of* Γ_y *and* C_y *denotes a finite constant that depends on* y *and* $\mathrm{d}_{\mathcal{X},\kappa}$. *Suppose that the kernel* $K : \mathbb{R} \to \mathbb{R}_+$, *is an even, twice continuously differentiable function with* $\mathrm{supp}\,(K) = [-1,1]$. *If* $\kappa < 1$, *we assume additionally that there are constants* $C_\kappa > 0$ *and* $1 \leq b < 5$ *such that for* $u \in (0,1)$ *it holds*

$$\omega_{\mathcal{X}}(u) := \sup_{x\in\mathcal{X}} \sup_{t,t'\in(0,1)^2, |t-t'|<u} \left| F_x^{-1}(t) - F_x^{-1}(t') \right| \leq C_\kappa u^{1/b}. \tag{4.7}$$

Limit Distribution: Given the assumptions discussed in the previous paragraph, we prove for $n \to \infty$, $h = o\left(n^{-1/5}\right)$ and $nh \to \infty$ that

$$\sqrt{nh}\left(\widehat{f_{\widehat{d}_{\mathcal{X},\kappa}}}(y) - f_{\mathrm{d}_{\mathcal{X},\kappa}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}_{\mathcal{X},\kappa}}(y) \int_{-1}^{1} K^2(u)\,du\right). \tag{4.8}$$

Consequently, the kernel density estimator $\widehat{f_{\widehat{d}_{\mathcal{X},\kappa}}}$ behaves pointwise asymptotically just as the kernel density estimator $\widehat{f}_{\mathrm{d}_{\mathcal{X},\kappa}}$ that is based on independent data.

Application: We employ the DTM-densities estimates with $\kappa = 1$ to distinguish between different metric measure spaces in order to illustrate the potential of this signature for pose invariant object discrimination. Furthermore, we apply the proposed signature with small κ for chromatin loop analysis. Chromatin fibers form chromosomes that are essential parts of cell nuclei and carry important genetic information. The investigation of "topologically associating domains" (TADs), which are self-interacting genomic regions and closely related with loops in the chromatin fibers, are of particular interest (Nuebler et al., 2018). With the help of synthetic, noisy SMLM-data we illustrate that DTM-densities can be used to classify the data according to their inherent loop frequencies (a small scale characteristic) without considering the global shape of the corresponding point clouds.

4.2 Discussion and Related Work

As already discussed in Section 2.2, the DTM-signature is introduced in Brécheteau (2019) (based on a slightly different DTM-function, compare (2.12) and (4.1)) and we refer to said section for a discussion of the results of this paper. However, we stress that we, different from Brécheteau (2019), do not need to subsample to derive the asymptotics of the signature proposed.
Additionally, we have already illustrated in Section 2.2, the relation of the DTM-signature to the Gromov-Wasserstein distance and particularly the relation to the lower bound $\mathbf{FLB}_1$. For a discussion of the usage of lower bounds (and the corresponding distance based signatures) in practice, we refer to Section 1.3 and Section 2.2.

Clearly, the DTM-function defined in (4.1) (or alternatively defined in (2.12)) is fundamental to the definition of the DTM-signature. This function is proposed for geometric inference in Chazal et al. (2011) and has been investigated in the context of topological data analysis and support estimation (Chazal et al., 2013; Buchet et al., 2014). However, not only $\mathrm{d}_{\mathcal{X},\kappa}$ but also $\widehat{d_{\mathcal{X},\kappa}}$ is thoroughly studied in the literature. In particular, many consistency properties of $\widehat{d_{\mathcal{X},\kappa}}$, that are central to our derivation of (4.8), are proven by Chazal et al. (2016, 2017).

The alternative formulation of $\widehat{d_{\mathcal{X},\kappa}}$ in (4.3) showcases the connection of $\mathrm{d}_{\mathcal{X},\kappa}$ to nearest neighbor distributions. Data analysis based on nearest neighbor distributions is quite common in various applications in biology (Zou and Wu, 1995; Meng et al., 2020) and physics (Torquato et al., 1990; Bhattacharjee, 2003; Hsiao et al., 2020). In a certain sense, the DTM-signature is an adaptation of a nearest neighbor distribution to a setting, where the samples stem from a continuous probability distribution, and it is noteworthy that taking the mean over a percentage of the nearest neighbors seems to robustify the method making it applicable in settings with relative high levels of noise.

We point out that the kernel density estimator $f_{\mathrm{d}_{\mathcal{X},\kappa}}$ defined in (4.5) is based on dependent data. Indeed, kernel density estimation based on dependent data is investigated in a variety of settings in the literature. It is considered for various mixing and linear processes connected to weakly dependent time series (Castellana and Leadbetter, 1986; Robinson, 1983; Liebscher, 1996; Lu, 2001; Wu and Mielniczuk, 2002). Further, the behavior of kernel density estimators of symmetric functions of the data as well as undirected dyadic data are studied (Frees, 1994; Graham et al., 2019). For these settings, a dependency structure that is reminiscent of U-statistics has to be treated carefully. However, it is easy to verify that none of these dependency frameworks applies to our setting.

Conclusion: We verify that, given technical but practically mild assumptions, the DTM-density estimate $\widehat{f_{d_{\mathcal{X},\kappa}}}$ behaves asymptotically pointwise just as a usual kernel density estimator based on independent random variables. Furthermore, we illustrate empirically that (estimates of) the DTM-density are a natural, powerful signature for the analysis of small scale differences in chromatin loop analysis. Potential further directions for research are the study of regularity of the DTM-density given a set $\mathcal{X}$ as well as the construction of a statistical test for comparing different DTM-densities.

Bibliography

Alvarez-Esteban, P. C., Del Barrio, E., Cuesta-Albertos, J. A., and Matran, C. (2008). Trimmed comparison of distributions. *Journal of the American Statistical Association*, 103(482):697–704.

Alvarez-Melis, D. and Jaakkola, T. (2018). Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890.

Amann, H. (2011). *Ordinary Differential Equations: An Introduction to Nonlinear Analysis*, volume 13. Walter de Gruyter.

Arcones, M. A. and Giné, E. (1994). U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Processes and their Applications*, 52(1):17–38.

Bartal, Y. (1996). Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193. IEEE.

Berrendero, J. R., Cuevas, A., and Pateiro-López, B. (2016). Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146(C):237–247.

Bhattacharjee, B. (2003). Nth-nearest-neighbor distribution functions of an interacting fluid from the pair correlation function: A hierarchical approach. *Physical Review E*, 67(4):041208.

Blum, M. G. and François, O. (2006). Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, 55(4):685–691.

Bobkov, S. and Ledoux, M. (2019). *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, volume 261. American Mathematical Society.

Boutin, M. and Kemper, G. (2004). On reconstructing n-point configurations from the distribution of distances or areas. *Advances in Applied Mathematics*, 32(4):709–735.

Brécheteau, C. (2019). A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature. *Electronic Journal of Statistics*, 13(1):795–849.

Brinkman, D. and Olver, P. J. (2012). Invariant histograms. *The American Mathematical Monthly*, 119(1):4–24.

Bronstein, A. M., Bronstein, M. M., Bruckstein, A. M., and Kimmel, R. (2009a). Partial similarity of objects, or how to compare a centaur to a horse. *International Journal of Computer Vision*, 84(2):163.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2006a). Efficient computation of isometry-invariant distances between surfaces. *SIAM Journal on Scientific Computing*, 28(5):1812–1836.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2006b). Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2009b). Topology-invariant similarity of nonrigid shapes. *International Journal of Computer Vision*, 81(3):281.

Bronstein, A. M., Bronstein, M. M., Kimmel, R., Mahmoudi, M., and Sapiro, G. (2010). A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *International Journal of Computer Vision*, 89(2-3):266–286.

Brown, P., Pullan, W., Yang, Y., and Zhou, Y. (2016). Fast and accurate non-sequential protein structure alignment using a new asymmetric linear sum assignment heuristic. *Bioinformatics*, 32(3):370–377.

Buchet, M., Chazal, F., Dey, T. K., Fan, F., Oudot, S. Y., and Wang, Y. (2014). Topological analysis of scalar fields with outliers. *preprint arXiv:1412.1680*.

Bunne, C., rez Melis, D., Krause, A., and Jegelka, S. (2019). Learning generative models across incomparable spaces. In *36th International Conference on Machine Learning*, pages 851–861. Proceedings of Machine Learning Research.

Burago, D., Burago, Y., and Ivanov, S. (2001). *A Course in Metric Geometry*, volume 33. American Mathematical Society.

Bustos, B., Keim, D. A., Saupe, D., Schreck, T., and Vranić, D. V. (2005). Feature-based similarity search in 3D object databases. *ACM Computing Surveys (CSUR)*, 37(4):345–387.

Carlsson, G. and Mémoli, F. (2010). Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11(4):1425–1470.

Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Processes and their Applications*, 21(2):179–193.

Chapel, L., Alaya, M. Z., and Gasso, G. (2020). Partial optimal tranport with applications on positive-unlabeled learning. *Advances in Neural Information Processing Systems*, 33:2903–2913.

Chazal, F., Cohen-Steiner, D., Guibas, L. J., Mémoli, F., and Oudot, S. Y. (2009). Gromov-Hausdorff stable signatures for shapes using persistence. In *Proceedings of the Symposium on Geometry Processing*, pages 1393–1403.

Chazal, F., Cohen-Steiner, D., and Mérigot, Q. (2011). Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751.

Chazal, F., Fasy, B., Lecci, F., Michel, B., Rinaldo, A., Rinaldo, A., and Wasserman, L. (2017). Robust topological inference: Distance to a measure and kernel distance. *The Journal of Machine Learning Research*, 18(1):5845–5884.

Chazal, F., Guibas, L. J., Oudot, S. Y., and Skraba, P. (2013). Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):1–38.

Chazal, F., Massart, P., Michel, B., et al. (2016). Rates of convergence for robust geometric inference. *Electronic Journal of Statistics*, 10(2):2243–2286.

Chen, J. and Safro, I. (2011). Algebraic distance on graphs. *SIAM Journal on Scientific Computing*, 33(6):3468–3490.

Chowdhury, S. and Mémoli, F. (2019). The Gromov-Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787.

Chowdhury, S. and Needham, T. (2021). Generalized spectral clustering via Gromov-Wasserstein learning. In *24 thInternational Conference on Artificial Intelligence and Statistics*, pages 712–720. Proceedings of Machine Learning Research.

Colijn, C. and Plazzotta, G. (2018). A metric on phylogenetic tree shapes. *Systematic biology*, 67(1):113–126.

Czado, C. and Munk, A. (1998). Assessing the similarity of distributions-finite sample performance of the empirical mallows distance. *Journal of Statistical Computation and Simulation*, 60(4):319–346.

Dayarian, A. and Shraiman, B. I. (2014). How to infer relative fitness from a sample of genomic sequences. *Genetics*, 197(3):913–923.

Dede, S. (2009). An empirical central limit theorem in L_1 for stationary sequences. *Stochastic Processes and their Applications*, 119(10):3494–3515.

Dedecker, J. and Merlevede, F. (2017). Behavior of the wasserstein distance between the empirical and the marginal distributions of stationary α-dependent sequences. *Bernoulli*, 23(3):2083–2127.

del Barrio, E., Cuesta-Albertos, J. A., Matrán, C., and Rodríguez-Rodríguez, J. M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *The Annals of Statistics*, 27(4):1230–1239.

del Barrio, E., Giné, E., and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1):131–189.

Demetci, P., Santorella, R., Sandstede, B., Noble, W. S., and Singh, R. (2020). Gromov-Wasserstein optimal transport to align single-cell multi-omics data. *BioRxiv*.

Dong, Y. and Sawin, W. (2020). COPT: Coordinated optimal transport on graphs. *preprint arXiv:2003.03892*.

d'Amico, M., Frosini, P., and Landi, C. (2010). Natural pseudo-distance and optimal matching between reduced size functions. *Acta Applicandae Mathematicae*, 109(2):527–554.

Evans, S. N. (2007). *Probability and Real Trees: École D'Été de Probabilités de Saint-Flour XXXV-2005*. Springer.

Fakcharoenphol, J., Rao, S., and Talwar, K. (2004). A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485–497.

Federer, H. (1969). *Geometric Measure Theory*. Springer.

Frees, E. W. (1994). Estimating densities of functions of observations. *Journal of the American Statistical Association*, 89(426):517–525.

Gellert, M., Hossain, M. F., Berens, F. J. F., Bruhn, L. W., Urbainsky, C., Liebscher, V., and Lillig, C. H. (2019). Substrate specificity of thioredoxins and glutaredoxins–towards a functional classification. *Heliyon*, 5(12):e02943.

Graham, B. S., Niu, F., and Powell, J. L. (2019). Kernel density estimation for undirected dyadic data. *preprint arXiv:1907.13630*.

Greven, A., Pfaffelhuber, P., and Winter, A. (2009). Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322.

Gromov, M., Katz, M., Pansu, P., and Semmes, S. (1999). *Metric Structures for Riemannian and Non-Riemannian Spaces*, volume 152. Springer.

Hamza, A. B. and Krim, H. (2003). Geodesic object representation and recognition. In *International conference on discrete geometry for computer imagery*, pages 378–387. Springer.

Hao, X., Parmar, J. J., Lelandais, B., Aristov, A., Ouyang, W., Weber, C., and Zimmer, C. (2021). Super-resolution visualization and modeling of human chromosomal regions reveals cohesin-dependent loop structures. *Genome Biology*, 22(1):1–20.

Hayakawa, Y. S. and Oguchi, T. (2016). Applications of terrestrial laser scanning in geomorphology. *Journal of Geography (Chigaku Zasshi)*, 125(3):299–324.

Hirsch, M. and Smale, S. (1974). *Differential Equations, Dynamical Systems, and Linear Algebra (Pure and Applied Mathematics, Vol. 60)*. Springer.

Holm, L. and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology*, 233(1):123–138.

Hsiao, S.-S., Chen, K.-T., and Ite, A. Y. (2020). Mean field theory of weakly-interacting rydberg polaritons in the eit system based on the nearest-neighbor distribution. *Optics Express*, 28(19):28414–28429.

Hundrieser, S., Klatt, M., Staudt, T., and Munk, A. (2022). A unifying approach to distributional limits for empirical optimal transport. *preprint arXiv:2202.12790*.

Jain, A. K. and Dorai, C. (2000). 3D object recognition: Representation and matching. *Statistics and Computing*, 10(2):167–182.

Jardine, N. and Sibson, R. (1971). *Mathematical Taxonomy*. John Wiley & Sons.

Kantorovich, L. V. and Rubinstein, G. S. (1958). On a space of completely additive functions. *Vestnik Leningrad University*, 13(7):52–59.

Kim, J., Rosenberg, N. A., and Palacios, J. A. (2020). Distance metrics for ranked evolutionary trees. *Proceedings of the National Academy of Sciences*, 117(46):28876–28886.

Kloeckner, B. R. (2015). A geometric study of Wasserstein spaces: Ultrametrics. *Mathematika*, 61(1):162–178.

Kolodny, R., Koehl, P., and Levitt, M. (2005). Comprehensive evaluation of protein structure alignment methods: Scoring by geometric measures. *Journal of Molecular Biology*, 346(4):1173–1188.

Kufareva, I. and Abagyan, R. (2011). Methods of protein structure comparison. In *Homology Modeling*, pages 231–257. Springer.

Kuo, H.-Y., Su, H.-R., Lai, S.-H., and Wu, C.-C. (2014). 3D object detection and pose estimation from depth image for robotic bin picking. In *2014 IEEE international conference on automation science and engineering (CASE)*, pages 1264–1269. IEEE.

Lewitus, E. and Morlon, H. (2016). Characterizing and comparing phylogenies from their Laplacian spectrum. *Systematic biology*, 65(3):495–507.

Libeskind, N. I., van de Weygaert, R., Cautun, M., Falck, B., Tempel, E., Abel, T., Alpaslan, M., Aragón-Calvo, M. A., Forero-Romero, J. E., Gonzalez, R., Gottlöber, S., Hahn, O., Hellwing, W. A., Hoffman, Y., Jones, B. J. T., Kitaura, F., Knebe, A., Manti, S., Neyrinck, M., Nuza, S. E., Padilla, N., Platen, E., Ramachandra, N., Robotham, A., Saar, E., Shandarin, S., Steinmetz, M., Stoica, R. S., Sousbie, T., and Yepes, G. (2017). Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications*, 65(1):69–80.

Liu, P., Biller, P., Gould, M., and Colijn, C. (2020). Polynomial phylogenetic analysis of tree shapes. *bioRxiv*.

Lowe, D. G. (2001). Local feature view clustering for 3D object recognition. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE.

Lu, Z. (2001). Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics*, 53(3):447–468.

Mallows, C. L. (1972). A note on asymptotic joint normality. *The Annals of Mathematical Statistics*, 43(2):508–515.

Mason, D. M. (1984). Weak convergence of the weighted empirical quantile process in $L^2(0, 1)$. *The Annals of Probability*, 12(1):243–255.

Mémoli, F. (2007). On the use of Gromov-Hausdorff distances for shape comparison. In *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association.

Mémoli, F. (2011). Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487.

Mémoli, F., Munk, A., Wan, Z., and Weitkamp, C. (2021a). The ultrametric Gromov-Wasserstein distance. *preprint arXiv:2101.05756*.

Mémoli, F. and Needham, T. (2021). Distance distributions and inverse problems for metric measure spaces. *preprint arXiv:1810.09646v3*.

Mémoli, F. and Sapiro, G. (2004). Comparing point clouds. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 32–40.

Mémoli, F. and Sapiro, G. (2005). A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5(3):313–347.

Mémoli, F., Smith, Z., and Wan, Z. (2021b). The Gromov-Hausdorff distance between ultrametric spaces: Its structure and computation. *preprint arXiv:2110.03136*.

Mémoli, F. and Wan, Z. (2019). Gromov-Hausdorff distances on p-metric spaces and ultrametric spaces. *preprint arXiv:1912.00564*.

Meng, H., Gao, Y., Yang, X., Wang, K., and Tian, J. (2020). K-nearest neighbor based locally connected network for fast morphological reconstruction in fluorescence molecular tomography. *IEEE transactions on medical imaging*, 39(10):3019–3028.

Mézard, M., Parisi, G., and Virasoro, M. A. (1987). *Spin Glass Theory and Beyond: An Introduction to the Replica Method and its Applications*, volume 9. World Scientific Publishing Company.

Mooers, A. O. and Heard, S. B. (1997). Inferring evolutionary process from phylogenetic tree shape. *The quarterly review of Biology*, 72(1):31–54.

Morgan, F. (2016). *Geometric Measure Theory: A Beginner's Guide*. Academic press.

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):223–241.

Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. *Nature*, 576(7785):132–137.

Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *The Annals of Probability*, pages 1291–1298.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706.

Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2002). Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4).

Papazov, C., Haddadin, S., Parusel, S., Krieger, K., and Burschka, D. (2012). Rigid 3D geometry matching for grasping of known objects in cluttered scenes. *The International Journal of Robotics Research*, 31(4):538–553.

Pardalos, P. M. and Vavasis, S. A. (1991). Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22.

Peyré, G., Cuturi, M., and Solomon, J. (2016). Gromov-Wasserstein averaging of kernel and distance matrices. In *33rd International Conference on Machine Learning*, pages 2664–2672. Proceedings of Machine Learning Research.

Poon, A. F., Walker, L. W., Murray, H., McCloskey, R. M., Harrigan, P. R., and Liang, R. H. (2013). Mapping the shapes of phylogenetic trees from human and zoonotic RNA viruses. *PLoS one*, 8(11):e78122.

Qiu, D. (2009). Geometry of non-archimedean Gromov-Hausdorff distance. *P-Adic Numbers, Ultrametric Analysis, and Applications*, 1(4):317.

Rammal, R., Toulouse, G., and Virasoro, M. A. (1986). Ultrametricity for physicists. *Reviews of Modern Physics*, 58:765–788.

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207.

Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121.

Rudolph, G. and Schmidt, M. (2012). *Differential Geometry and Mathematical Physics*. Springer.

Ruggeri, M. R. and Saupe, D. (2008). Isometry-invariant matching of point set surfaces. In *Eurographics 2008 Workshop on 3D Object Retrieval*. The Eurographics Association.

Scetbon, M., Peyré, G., and Cuturi, M. (2021). Linear-time Gromov-Wasserstein distances using low rank couplings and costs. *preprint arXiv:2106.01128*.

Séjourné, T., Vialard, F.-X., and Peyré, G. (2021). The unbalanced Gromov-Wasserstein distance: Conic formulation and relaxation. *Advances in Neural Information Processing Systems*, 34.

Semple, C., Steel, M., et al. (2003). *Phylogenetics*, volume 24. Oxford University Press on Demand.

Shi, Y., Thompson, P. M., de Zubicaray, G. I., Rose, S. E., Tu, Z., Dinov, I., and Toga, A. W. (2007). Direct mapping of hippocampal surfaces with intrinsic shape context. *NeuroImage*, 37(3):792–807.

Srivastava, S., Lal, S. B., Mishra, D. C., Angadi, U., Chaturvedi, K., Rai, S. N., and Rai, A. (2016). An efficient algorithm for protein structure comparison using elastic shape analysis. *Algorithms for Molecular Biology*, 11(1):1–11.

Sturm, K.-T. (2006). On the geometry of metric measure spaces. *Acta Mathematica*, 196(1):65–131.

Sturm, K.-T. (2012). The space of spaces: Curvature bounds and gradient flows on the space of metric measure spaces. *preprint arXiv:1208.0434*.

Torquato, S., Lu, B., and Rubinstein, J. (1990). Nearest-neighbor distribution functions in many-body systems. *Physical Review A*, 41(4):2059.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, volume 1. Springer New York.

Vaserstein, L. N. (1969). Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72.

Vayer, T., Courty, N., Tavenard, R., Laetitia, C., and Flamary, R. (2019). Optimal transport for structured data with application on graphs. In *36th International Conference on Machine Learning*, volume 97, pages 6275–6284. Proceedings of Machine Learning Research.

Veltkamp, R. C. and Latecki, L. J. (2006). Properties and performance of shape similarity measures. In *Data Science and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg.

Villani, C. (2008). *Optimal Transport: Old and New*. Springer Science & Business Media.

Vosselman, G., Gorte, B., and Sithole, G. andRabbani, T. (2004). Recognizing structure in laser scanner point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46(8):33–38.

Weitkamp, C. A., Proksch, K., Staudt, T., Lelandais, B., and Zimmer, C. (2022). From small scales to large scales: Distance-to-Measure density based geometric analysis of complex data. *In preparation*.

Weitkamp, C. A., Proksch, K., Tameling, C., and Munk, A. (2020). Gromov-Wasserstein distance based object matching: Asymptotic inference. *preprint arXiv:2006.12287*.

Wendler, M. (2012). U-processes, U-quantile processes and generalized linear statistics of dependent data. *Stochastic Processes and their Applications*, 122(3):787–807.

Wills, P. and Meyer, F. G. (2020). Metrics for graph comparison: A practitioner's guide. *Plos one*, 15(2):e0228728.

Wilson, R. C. and Zhu, P. (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9):2833–2841.

Wu, T. and Choi, K. P. (2016). On joint subtree distributions under two evolutionary models. *Theoretical Population Biology*, 108(C):13–23.

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *The Annals of Statistics*, 30(5):1441 – 1459.

Zarichnyi, I. (2005). Gromov-Hausdorff ultrametric. *preprint arXiv:math/0511437*.

Zou, G. and Wu, H.-I. (1995). Nearest-neighbor distribution of interacting biological entities. *Journal of Theoretical Biology*, 172(4):347–353.

Addenda

The following addenda contains the three articles A, B and C that form the basis of this thesis. An introductory summary lists each article's reference and abstract.

Distribution of Distances based Object Matching: Asymptotic Inference
Christoph Alexander Weitkamp, Katharina Proksch, Carla Tameling and Axel Munk
Preprint available, arXiv:2006.12287 (2020)

Abstract In this paper, we aim to provide a statistical theory for object matching based on a lower bound of the Gromov-Wasserstein distance related to the distribution of (pairwise) distances of the considered objects. To this end, we model general objects as metric measure spaces. Based on this, we propose a simple and efficiently computable asymptotic statistical test for pose invariant object discrimination. This is based on a (β-trimmed) empirical version of the afore-mentioned lower bound. We derive for the trimmed and untrimmed case the distributional limits of this test statistic. For this purpose, we introduce a novel U-type process indexed in β and show its weak convergence. The theory developed is investigated in Monte Carlo simulations and applied to structural protein comparisons.

The ultrametric Gromov-Wasserstein distance
Facundo Mémoli, Axel Munk, Zhengchao Wan and Christoph Alexander Weitkamp
Preprint available, arXiv:2101.05756 (2021)

Abstract In this paper, we investigate compact ultrametric measure spaces which form a subset $\mathcal{U}^w$ of the collection of all metric measure spaces $\mathcal{M}^w$. In analogy with the notion of the ultrametric Gromov-Hausdorff distance on the collection of ultrametric spaces $\mathcal{U}$, we define ultrametric versions of two metrics on $\mathcal{U}^w$, namely of Sturm's Gromov-Wasserstein distance of order p and of the Gromov-Wasserstein distance of order p. We study the basic topological and geometric properties of these distances as well as their relation and derive for $p = \infty$ a polynomial time algorithm for their calculation. Further, several lower bounds for both distances are derived and some of our results are generalized to the case of finite ultra-dissimilarity spaces. Finally, we study the relation between the Gromov-Wasserstein distance and its ultrametric version (as well as the relation between the

corresponding lower bounds) in simulations and apply our findings for phylogenetic tree shape comparisons.

From Small Scales to Large Scales: Distance-to-Measure Density based Geometric Analysis of Complex Data

Katharina Proksch, Christoph Alexander Weitkamp, Thomas Staudt, Christophe Zimmer and Benoît Lelandais

Preprint available, arXiv:2205.07689 (2022)

Abstract How can we tell complex point clouds with different small scale characteristics apart, while disregarding global features? Can we find a suitable transformation of such data in a way that allows to discriminate between differences in this sense?

In this paper, we consider the analysis and classification of complex point clouds as they are obtained, e.g., via single molecule localization microscopy. We focus on the task of identifying differences between noisy point clouds based on small scale characteristics, while disregarding large scale information such as overall size. We propose an approach based on a transformation of the data via the so-called Distance-to-Measure (DTM) function, a transformation which is based on the average of nearest neighbor distances. For each data set, we estimate the probability density of average local distances of all data points and use the estimated densities for classification. While the practical performance of the proposed methodology is very good, the theoretical study of the density estimators is quite challenging, as they are based on *non-i.i.d.* observations that have been obtained via a complicated transformation. In fact, the transformed data are stochastically dependent in a non-local way that is not captured by typical dependence measures. Nonetheless, we show that the asymptotic behaviour of the density estimator is driven by a kernel density estimator of certain i.i.d. random variables by using theoretical properties of U-statistics, which allows to handle the dependencies via a Hoeffding decomposition. We show via a numerical study and in an application to simulated single molecule localization microscopy data of chromatin fibers that unsupervised classification tasks based on estimated DTM-densities achieve very good separation results.

CHAPTER A

Distribution of Distances based Object Matching: Asymptotic Inference

Distribution of Distances based Object Matching: Asymptotic Inference

Christoph Alexander Weitkamp[*] Katharina Proksch [†] Carla Tameling [*]
Axel Munk [*‡]

March 31, 2022

Abstract

In this paper, we aim to provide a statistical theory for object matching based on a lower bound of the Gromov-Wasserstein distance related to the distribution of (pairwise) distances of the considered objects. To this end, we model general objects as metric measure spaces. Based on this, we propose a simple and efficiently computable asymptotic statistical test for pose invariant object discrimination. This is based on a (β-trimmed) empirical version of the afore-mentioned lower bound. We derive for the trimmed and untrimmed case the distributional limits of this test statistic. For this purpose, we introduce a novel U-type process indexed in β and show its weak convergence. The theory developed is investigated in Monte Carlo simulations and applied to structural protein comparisons.

Keywords Gromov-Wasserstein distance, metric measures spaces, U-processes, distributional limits, protein matching

MSC 2010 subject classification Primary: 62E20, 62G20, 65C60 Secondary: 60E05

1 Introduction

Over the last decades, the acquisition of geometrically complex data in various fields of application has increased drastically. For the digital organization and analysis of such data it is important to have meaningful notions of *similarity* between datasets as well as between shapes. This most certainly holds true for the area of 3-D object matching,

[*]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen
[†]Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Hallenweg 19, 7522NH Enschede
[‡]Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen

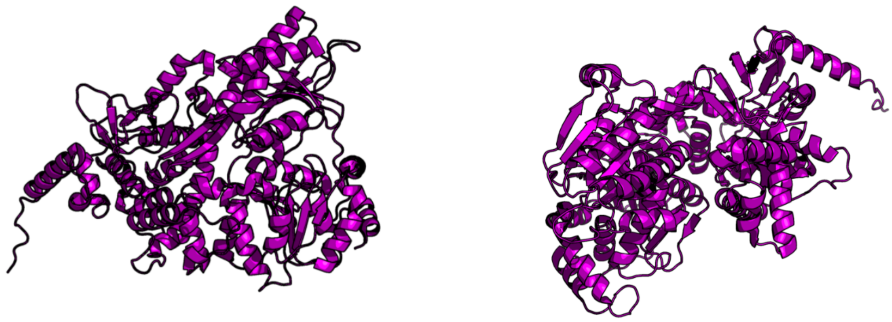

Fig. 1: Illustration of the proteins to be compared: Cartoon representation of the DEAH-box RNA-helicase Prp43 from chaetomium thermophilum bound to ADP (PDB ID: 5D0U [65]) in two different poses. The DEAH-box helicase Prp43 unwinds double stranded RNA and rearranges RNA/protein complexes. It has essential roles in pre-mRNA splicing and ribosome biogenesis [4, 40].

which has many relevant applications, for example in computer vision [67, 70], mechanical engineering [5, 30] or molecular biology [39, 50]. In most of these applications, an important challenge is to distinguish between shapes while regarding identical objects in different spatial orientations as equal. A prominent example is the comparison of 3-D protein structures, which is important for understanding structural, functional and evolutionary relationships among proteins [38, 62]. Most known protein structures are published as coordinate files, where for every atom a 3-D coordinate is estimated based on an indirect observation of the protein's electron density (see Rhodes [53] for further details), and stored e.g. in the protein database **PDB** [8]. These coordinate files lack any kind of orientation and any meaningful comparison has to take this into account. Figure 1 (created with PyMOL [58]) shows two cartoon representations of the backbone of the protein structure 5D0U in two different poses. These two representations obtained from the same coordinate file highlight the difficulty to identify them from noisy measurements.

Consequently, many approaches to pose invariant shape matching, classification and recognition have been suggested and studied in the literature. The majority of these methods computes and compares certain invariants or signatures in order to decide whether the considered objects are *equal* up to a previously defined notion of invariance. In the literature, these methods are often called *feature* (or *signature*) based methods, see Cárdenas et al. [18] for a comprehensive survey. Some examples for features are the *shape distributions* [51], that are connected to the distributions of lengths, areas and volumes of an object, the *shape contexts* [7], that rely in a sense on a local distribution of inter-point distances of the considered object, and *reduced size functions* [29], which count the connected components of certain lower level sets.
As noted by Mémoli [44, 45], several signatures describe different aspects of a metric between objects. In these and subsequent papers, the author develops a unifying view point

by representing an object as *metric measure space* $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$, where $(\mathcal{X}, d_{\mathcal{X}})$ is a compact metric space and $\mu_{\mathcal{X}}$ denotes a Borel probability measure on $\mathcal{X}$. The additional probability measure, whose support is assumed to be $\mathcal{X}$, can be thought of as signaling the importance of different regions of the modeled object. Based on the original work of Gromov [33], Mémoli [45] introduced the *Gromov-Wasserstein distance* of order $p \in [1, \infty)$ between two (compact) metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ which will be fundamental to this paper. It is defined as

$$\mathcal{GW}_p(\mathcal{X}, \mathcal{Y}) = \inf_{\pi \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})} J_p(\pi), \tag{1}$$

where

$$J_p(\pi) := \frac{1}{2} \left(\int_{\mathcal{X}\times\mathcal{Y}} \int_{\mathcal{X}\times\mathcal{Y}} \left| d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \right|^p \, \pi(dx \times dy) \, \pi(dx' \times dy') \right)^{\frac{1}{p}}.$$

Here, $\mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ stands for the set of all couplings of $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$, i.e., the set of all measures π on the product space $\mathcal{X} \times \mathcal{Y}$ such that $\pi(A \times \mathcal{Y}) = \mu_{\mathcal{X}}(A)$ and $\pi(\mathcal{X} \times B) = \mu_{\mathcal{Y}}(B)$ for all measurable sets $A \subset \mathcal{X}$ and $B \subset \mathcal{Y}$. In Section 5 of Mémoli [45] it is ensured that the Gromov-Wasserstein distance $\mathcal{GW}_p$ is suitable for pose invariant object matching by proving that it is a metric on the collection of all isomorphism classes of metric measure spaces.[1] Hence, objects are considered to be the same if they can be transformed into each other without changing the distances between their points and such that the corresponding measures are preserved. For example, if the distance is Euclidean, this leads to identifying objects up to translations, rotations and reflections [41]. This makes the Gromov-Wasserstein distance theoretically well suited for a variety of shape matching tasks, including protein structure comparisons. However, the practical usage of the Gromov-Wasserstein approach is severely hindered by its computational complexity: Already for two finite metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ with metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ and probability measures $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$, respectively, the computation of $\mathcal{GW}_p(\mathcal{X}, \mathcal{Y})$ boils down to solving a (non-convex) quadratic program [45, Sec. 7]. This is in general NP-hard [52]. To circumvent the precise determination of the Gromov-Wasserstein distance in practice, it can be approximated by conditional gradient descent [44, 45]. The result of this numerical scheme, however, is difficult to interpret, as it does not come with any guarantee how close it is to the minimum in (1). Nevertheless, this approach has been used in various applications and led to several extensions of the Gromov-Wasserstein distance [19, 20, 60], especially in the area of machine learning [2, 17, 66, 72]. Gellert et al. [32] pursued a different route to approximating the Gromov-Wasserstein distance by applying certain lower bounds of the Gromov-Wasserstein distance derived in [44, 45] for the comparison of the isosurfaces of

[1] Two metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ are isomorphic (denoted as $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}}) \cong (\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$) if and only if there exists an isometry $\phi : \mathcal{X} \to \mathcal{Y}$ such that $\phi \# \mu_{\mathcal{X}} = \mu_{\mathcal{Y}}$. Here, $\phi \# \mu_{\mathcal{X}}$ denotes the pushforward measure.

various proteins. Among other things, the authors used that

$$\mathcal{GW}_p(\mathcal{X},\mathcal{Y})\geq\frac{1}{2}(DoD_p\,(\mathcal{X},\mathcal{Y}))^{\frac{1}{p}}:=\frac{1}{2}\left(\inf_{\pi\in\widetilde{\mathcal{M}}}\int_{\mathcal{X}^2\times\mathcal{Y}^2}\left|d_{\mathcal{X}}(x,x')-d_{\mathcal{Y}}(y,y')\right|^p d\pi(x,x'y,y')\right)^{\frac{1}{p}},$$

where $\widetilde{\mathcal{M}} := \mathcal{M}(\mu_{\mathcal{X}}\otimes\mu_{\mathcal{X}},\mu_{\mathcal{Y}}\otimes\mu_{\mathcal{Y}})$ and that DoD_p can be reformulated in terms of the *distribution of (pairwise) distances*. Let μ^U be the probability measure of the random variable $d_{\mathcal{X}}(X,X')$, where $X,X'\overset{i.i.d.}{\sim}\mu_{\mathcal{X}}$, and let μ^V be the one of $d_{\mathcal{Y}}(Y,Y')$, with $Y,Y'\overset{i.i.d.}{\sim}\mu_{\mathcal{Y}}$. Then, we call μ^U and μ^V the distribution of the (pairwise) distances of $(\mathcal{X},d_{\mathcal{X}},\mu_{\mathcal{X}})$ and $(\mathcal{Y},d_{\mathcal{Y}},\mu_{\mathcal{Y}})$, respectively. It is shown in Chowdhury and Mémoli [19, Thm. 24] that

$$DoD_p\,(\mathcal{X},\mathcal{Y})=\int_0^1\left|U^{-1}(t)-V^{-1}(t)\right|^p dt, \tag{2}$$

where U^{-1} and V^{-1} are the quantile functions of μ^U and μ^V, respectively. Thus, this bound quantifies the differences between the distributions of pairwise distances of the metric measure spaces $(\mathcal{X},d_{\mathcal{X}},\mu_{\mathcal{X}})$ and $(\mathcal{Y},d_{\mathcal{Y}},\mu_{\mathcal{Y}})$ in terms of the *Kantorovich distance* (see e.g. Villani [69]).

In this paper, we investigate the statistical properties of the sample counterpart of DoD_p in (2), which is on the one hand extremely simple to compute in a quadratic number of elementary operations (see Section 1.1) and on the other hand statistically accessible and useful for inference tasks such as object discrimination when the data are randomly sampled or the data set is massive and subsampling becomes necessary. Generally, DoD_p is a simple and natural measure to compare distance matrices. Such distance matrices underlie many methods of data analysis, e.g. various multidimensional scaling techniques (see Dokmanic et al. [28]). Hence, we believe that our analysis is of quite general statistical interest beyond the described scenario.

1.1 The Proposed Approach

Given two metric measure spaces, denoted as $(\mathcal{X},d_{\mathcal{X}},\mu_{\mathcal{X}})$ and $(\mathcal{Y},d_{\mathcal{Y}},\mu_{\mathcal{Y}})$, we aim to construct an (asymptotic) test based on $DoD_p(\mathcal{X},\mathcal{Y})$ for the hypothesis testing problem

$$H_0:\ DoD_p(\mathcal{X},\mathcal{Y})=0\ \text{ vs }\ H_1:\ DoD_p(\mathcal{X},\mathcal{Y})>0. \tag{3}$$

As DoD_p is a lower bound of the Gromov-Wasserstein distance, if $DoD_p(\mathcal{X},\mathcal{Y})$ is positive, the same holds for $\mathcal{GW}_p(\mathcal{X},\mathcal{Y})$. Thus, we infer that $(\mathcal{X},d_{\mathcal{X}},\mu_{\mathcal{X}})$ is not isomorphic to $(\mathcal{Y},d_{\mathcal{Y}},\mu_{\mathcal{Y}})$, if $DoD_p(\mathcal{X},\mathcal{Y})>0$. Hence, it is possible (as done in the following) to employ an (asymptotic) level α test for H_0 for pose invariant object discrimination, i.e., to test $H_0^*:\mathcal{X}\cong\mathcal{Y}$, against the alternative $H_1^*:\mathcal{X}\not\cong\mathcal{Y}$ at significance level $1-\alpha$. It is well known that the distribution of distances does not uniquely characterize a metric measure space [45], i.e., there are metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $DoD_p(\mathcal{X},\mathcal{Y}) = 0$, although $\mathcal{GW}_p(\mathcal{X},\mathcal{Y}) > 0$. In consequence, a test for H_0 applied to test for H_0^* cannot develop power for every alternative in H_1^*. However, this seems to be a minor issue for many practical applications. Indeed, the distribution of distances was proposed as a feature itself for feature based object matching and was shown to work well in various examples [9, 16, 32, 51]. Furthermore, the discriminative abilities of the distribution of distances are well studied theoretically [14, 46], see also Section 2 and Section 4.

To set up our statistical framework, let $X_1,\dots,X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and $Y_1,\dots,Y_m \overset{i.i.d.}{\sim} \mu_{\mathcal{Y}}$ be two independent samples and let $\mathcal{X}_n = \{X_1,\dots,X_n\}$ and $\mathcal{Y}_m = \{Y_1,\dots,Y_m\}$. The sample analog to (2) is to be defined with respect to the empirical measures and we obtain the *DoD-statistic* as

$$\widehat{DoD}_p = \widehat{DoD}_p\left(\mathcal{X}_n,\mathcal{Y}_m\right) := \int_0^1 \left|U_n^{-1}(t) - V_m^{-1}(t)\right|^p dt, \tag{4}$$

where, for $t \in \mathbb{R}$, U_n and V_m are defined as the empirical c.d.f.'s of all pairwise distances of the samples $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively, i.e.,

$$U_n(t) := \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\le t\}} \text{ and } V_m(t) := \frac{2}{m(m-1)} \sum_{1\le k<l\le m} \mathbb{1}_{\{d_{\mathcal{Y}}(Y_k,Y_l)\le t\}}. \tag{5}$$

Besides, U_n^{-1} and V_m^{-1} denote the corresponding empirical quantile functions. We stress that the evaluation of $\widehat{DoD}_p$ boils down to the calculation of a sum and no formal integration is required. Let $d_{(i)}^{\mathcal{X}}$ denote the i-th order statistic of the sample $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\le i<j\le n}$ and let $d_{(i)}^{\mathcal{Y}}$ be defined analogously. Let $N := n(n-1)/2$ and $M := m(m-1)/2$. Then,

$$\widehat{DoD}_p = \sum_{i=1}^{N}\sum_{j=1}^{M} \lambda_{ij} \left|d_{(i)}^{\mathcal{X}} - d_{(j)}^{\mathcal{Y}}\right|^p,$$

where $\lambda_{ij} = \left(\frac{i}{N} \wedge \frac{j}{M} - \frac{i-1}{N} \vee \frac{j-1}{M}\right) \mathbb{1}_{\{iM\wedge jN > (i-1)M \vee (j-1)N\}}$. Here, and in the following, $a \wedge b$ denotes the minimum and $a \vee b$ the maximum of two real numbers a and b. Hence, the representation (2) admits an empirical version which is computable in $O\left((m\vee n)^2\right)$ elementary operations, if the computation of one distance is considered as $O(1)$.

1.2 Main Results

The main contributions of the paper are various upper bounds and distributional limits for the statistic defined in (4) (as well as trimmed variants). Based on these, we design an asymptotic test for the hypothesis H_0 defined in (3). Other statistical applications such as confidence intervals for DoD_p or multi-sample extensions are straight forward and omitted for simplicity. We focus, for ease of notation, on the case $p = 2$ (see Section 2.3 for $p \in [1, \infty)$), i.e., we derive for $\beta \in [0, 1/2)$ the limit behavior of the statistic

$$\widehat{DoD}_{(\beta)} := \int_{\beta}^{1-\beta} \left(U_n^{-1}(t) - V_m^{-1}(t)\right)^2 \, dt \tag{6}$$

under the hypothesis as well as under the alternative in (3). While in many applications $\beta = 0$ in (6) is a natural choice, the introduced trimming parameter β can be used to robustify the proposed method [1, 21]. Furthermore, it gives the possibility to focus the comparison on specific areas of the considered distributions of distances when additional information about their shapes is available. In Section 4, we illustrate the influence of this parameter empirically. Next, we briefly summarize the setting in which we are working and introduce the conditions required.

Setting 1.1. *Let $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ be two metric measure spaces and let μ^U and μ^V denote the distributions of (pairwise) distances of the spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$, respectively. For U the c.d.f. of μ^U, assume that U is differentiable with derivative u and let U^{-1} be the quantile function of U. Let V, V^{-1} and v be defined analogously. Further, let the samples $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} \mu_{\mathcal{Y}}$ be independent of each other and let U_n^{-1} and V_m^{-1} denote the empirical quantile functions of U_n and V_m in (5).*

Since the statistic $\widehat{DoD}_{(\beta)}$ in (6) is based on empirical quantile functions, or more precisely empirical U-quantile functions, we have to ensure that the corresponding U-distribution functions are well-behaved. In the course of this, we distinguish the cases $\beta \in (0, 1/2)$ and $\beta = 0$. The subsequent condition guarantees that the inversion functional $\phi_{inv} : F \mapsto F^{-1}$ is Hadamard differentiable as a map from the set of restricted distribution functions into the space of all bounded functions on $[\beta, 1-\beta]$, in the following denoted as $\ell^\infty[\beta, 1-\beta]$.

Condition 1.2. *Let $\beta \in (0, 1/2)$ and let U be continuously differentiable on an interval $[C_1, C_2] = [U^{-1}(\beta) - \epsilon, U^{-1}(1-\beta) + \epsilon]$ for some $\epsilon > 0$ with strictly positive derivative u and let the analogous assumption hold for V and its derivative v.*

When the densities of μ^U and μ^V vanish at the boundaries of their support, which commonly happens (see Example 2.1), we can no longer rely on Hadamard differentiability to derive the limit distribution of $\widehat{DoD}_{(\beta)}$ under H_0 for $\beta = 0$. In order to deal with this case we require stronger assumptions. The following ones resemble those of Mason [42].

Condition 1.3. *Let U be continuously differentiable on its support. Further, assume there exist constants $-1 < \gamma_1, \gamma_2 < \infty$ and $c_U > 0$ such that $\left|(U^{-1})'(t)\right| \leq c_U t^{\gamma_1}(1-t)^{\gamma_2}$ for $t \in (0, 1)$ and let the analogous assumptions hold for V and $(V^{-1})'$.*

Both, Condition 1.2 and Condition 1.3 are comprehensively discussed in Section 2.1 and various illustrative examples are given there.

Limit distribution under H_0: Here, we have that the distributions of distances of the considered metric measure spaces, μ^U and μ^V, are equal, i.e., $U(t) = V(t)$ for $t \in \mathbb{R}$. Given Condition 1.2 we find that for $\beta \in (0, 1/2)$ (resp. given Condition 1.3 for $\beta = 0$) and $n, m \to \infty$

$$\frac{nm}{n+m}\widehat{DoD}_{(\beta)} \rightsquigarrow \Xi = \Xi(\beta) := \int_{\beta}^{1-\beta} (\mathbb{G}(t))^2 \, dt, \tag{7}$$

where $\mathbb{G}$ is a centered Gaussian process with covariance depending on U (under H_0 we have $U = V$) in an explicit but complicated way, see Theorem 2.6. Further, "$\rightsquigarrow$" denotes weak convergence in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner [68, Part 1]). Additionally, we establish in Section 2 a simple concentration bound for $\widehat{DoD}_{(\beta)}$ and demonstrate that for $\beta \in (0, 1/2)$ and $\alpha \in (0, 1)$ the corresponding α-quantile of Ξ, which is required for testing, can be obtained by a bootstrap scheme, see Section 3.

Limit distribution under H_1: Under the additional assumption (which is only required for $\beta > 0$) that

$$DoD_{(\beta)} := \int_{\beta}^{1-\beta} \left(U^{-1}(t) - V^{-1}(t)\right)^2 \, dt > 0,$$

we can prove (cf. Theorem 2.7) that given Condition 1.2 it holds for $n, m \to \infty$ and $\beta \in (0, 1/2)$ (resp. given Condition 1.3 for $\beta = 0$) that

$$\sqrt{\frac{nm}{n+m}}\left(\widehat{DoD}_{(\beta)} - DoD_{(\beta)}\right) \rightsquigarrow N(0, \sigma^2_{U,V,\lambda}), \tag{8}$$

where $N(0, \sigma^2_{U,V,\lambda})$ denotes a normal distribution with mean 0 and variance $\sigma^2_{U,V,\lambda}$ depending in an explicit way on U, V, β and $\lambda = \lim_{n,m\to\infty} n/(m + n)$.

1.3 Applications

From our theory it follows that for any $\beta \in [0, 1/2)$ a (robust) asymptotic level α test for H_0 against H_1 is given by rejecting H_0 in (3) if

$$\frac{nm}{n+m}\widehat{DoD}_{(\beta)} > \xi_{1-\alpha}, \tag{9}$$

where $\xi_{1-\alpha}$ denotes the $(1 - \alpha)$-quantile of Ξ. This has many possible applications. Exemplarily, in Section 5, we model proteins as metric measure spaces by assuming that the coordinate files are samples from (unknown) distributions (see Rhodes [53]) and apply

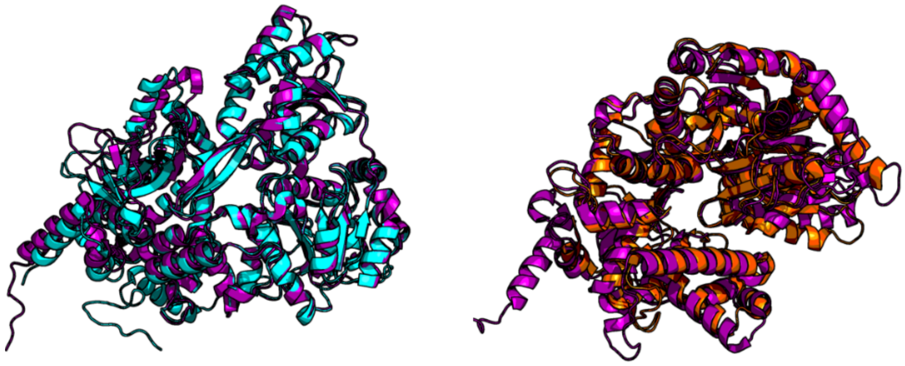

Fig. 2: Illustration of the proteins to be compared: Cartoon representation of the DEAH-box RNA-helicase Prp43 from chaetomium thermophilum bound to ADP (purple, PDB ID: 5D0U [65]) in alignment with Prp43 from saccharomyces cerevisiae in complex with CDP (cyan, PDB ID: 5JPT [54], left) and in alignment with the DEAH-box RNA helicase Prp2 in complex with ADP (orange, PDB ID: 6FAA [57], right). Prp2 is closely related to Prp43 and is necessary for the catalytic activation of the spliceosome in pre-mRNA splicing [37].

the theory developed to compare the protein structures depicted in Figure 2. Our major findings can be summarized as follows:

5D0U vs 5JPT: 5D0U and 5JPT are two structures of the same protein from different organisms. Consequently, their secondary structure elements can almost be aligned perfectly (see Figure 2, left). Only small parts of the structures are slightly shifted and do not overlap in the alignment. Applying (9) for this comparison generally yields no discrimination between these two protein structures, as $DoD_{(\beta)}$ is robust with respect to these kinds of differences. This robustness indeed makes the proposed method particularly suitable for protein structure comparison.

5D0U vs 6FAA: 5D0U and 6FAA are structures from closely related proteins and thus they are rather similar. Their alignment (Figure 2, right) shows minor differences in the orientation of some secondary structure elements and that 5D0U contains an α-helix that is not present in 6FAA. We find that $DoD_{(\beta)}$ is highly sensitive to such a deviation from H_0, as the proposed procedure discriminates very well between both structures already for small sample sizes.

Besides of testing H_0, we mention that our theory immediately allows to perform tests for relevant differences, i.e., to test $H : DoD_{(\beta)} \leq \epsilon$ vs $K : DoD_{(\beta)} > \epsilon$ for some specified $\epsilon > 0$ (see e.g. Munk and Czado [48] or Dette et al. [27] for a discussion). Further, k-sample comparisons and asymptotic confidence intervals for $DoD_{(\beta)}$ can be obtained analogously. Our results also justify subsampling (possibly in combination with bootstrapping) as an effective scheme to reduce the computational costs of $O\left((m \vee n)^2\right)$ further to evaluate

$\widehat{DoD}_{(\beta)}$ for large scale applications.

1.4 Related Work

First, we note that U_n and V_m can be viewed as empirical c.d.f.'s of the $N := n(n-1)/2$ and $M := m(m-1)/2$ random variables $d_{\mathcal{X}}(X_i, X_j)$, $1 \leq i < j \leq n$, and $d_{\mathcal{Y}}(Y_k, Y_l)$, $1 \leq k < l \leq m$, respectively. Hence, (4) can be viewed as the one dimensional empirical Kantorovich distance with N and M (dependent) data, respectively. There is a long standing interest in distributional limits for the one dimensional empirical Kantorovich distance [13, 24, 25, 48, 61, 64] as well as for empirical Kantorovich type distances with more general cost functions [11, 12]. Apparently, the major difficulty in our setting arises from the dependency of the random variables $\{d_{\mathcal{X}}(X_i, X_j)\}$ and the random variables $\{d_{\mathcal{Y}}(Y_k, Y_l)\}$, respectively. Compared to the techniques available for stationary and α-dependent sequences [22, 23], the statistic $\widehat{DoD}_{(\beta)}$ admits an intrinsic structure related to U- and U-quantile processes [3, 49, 71]. Note that for $\beta > 0$ we could have used the results of Wendler [71] to derive the asymptotics of $\widehat{DoD}_{(\beta)}$ as well, as they provide almost sure approximations of the empirical U-quantile processes $\mathbb{U}_n^{-1} := \sqrt{n}\left(U_n^{-1} - U^{-1}\right)$ and $\mathbb{V}_m^{-1} := \sqrt{m}\left(V_m^{-1} - V^{-1}\right)$ in $\ell^\infty[\beta, 1-\beta]$, however at the expense of slightly stronger smoothness requirements on U and V. In contrast, the more interesting case $\beta = 0$ is much more involved as the processes $\mathbb{U}_n^{-1}$ and $\mathbb{V}_m^{-1}$ do in general not converge in $\ell^\infty(0,1)$ under Condition 1.3 and the technique in Wendler [71] fails. Under the null hypothesis, we circumvent this difficulty by targeting our statistic for $\beta = 0$ directly, viewed as a process indexed in β. Under the alternative, we show the Hadamard differentiability of the inversion functional ϕ_{inv} onto the space of $\mathbb{R}$-valued, integrable functions on $(0,1)$ (denoted as $\ell^1(0,1)$) and verify that this is sufficient to derive (8).

Notice that tests based on distance matrices appear naturally in several applications, see, e.g., the recent works by Baringhaus and Franz [6], Montero-Manso and Vilar [47], Sejdinovic et al. [59], where the two sample homogeneity problem, i.e., testing whether two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are equal, is considered for high dimensions. Most similar in spirit to our work is Brécheteau [15] who also considers an asymptotic statistical test for a different lower bound of the Gromov-Wasserstein distance. This is based on a nearest neighbor-type approach and subsampling. However, the subsampling scheme is such that asymptotically all distances considered are independent, while we explicitly deal with the dependency structures present in the entire sample of the $n(n-1)/2$ distances. In Section 4.2 and Section 5.1 we empirically demonstrate that this leads to an increase of power and compare our test with the one proposed by Brécheteau [15] in more detail. Closely related from a practical point of view is also the work of Gellert et al. [32], who used and empirically compared several lower bounds of the Gromov-Wasserstein distance for clustering of various redoxins including our lower bound in (2). In fact, to reduce the computational complexity they heuristically employed a bootstrap scheme related to the one investigated in this paper and reported empirically good results. Finally, we mention that permutation based testing for U-statistics (see e.g. Berrett et al. [10]) is an interesting alternative to our bootstrap test and worth to be investigated further in our context.

1.5 Organization of the Paper

Section 2 states the main results and is concerned with the derivation of a simple finite sample bound for the expectation of $\widehat{DoD}_{(\beta)}$ as well as the proofs of (7) and (8). In Section 3 we propose for $\beta \in (0, 1/2)$ a bootstrapping scheme to approximate the quantiles of Ξ defined in (7). Afterwards in Section 4 we investigate the speed of convergence of $\widehat{DoD}_{(\beta)}$ to its limit distribution under H_0 in a Monte Carlo study. In this section we further study the introduced bootstrap approximation and investigate what kind of differences are detectable employing $\widehat{DoD}_{(\beta)}$ for H_0 by means of various examples. We apply the proposed test for the discrimination of 3-D protein structures in Section 5 and compare our results to the ones obtained by the method of Brécheteau [15]. Our simulations and data analysis of the example introduced previously (see Figure 2) suggest that the proposed $\widehat{DoD}_{(\beta)}$ based test outperforms the one proposed by Brécheteau [15] for protein structure comparisons.
In part Part I of the supplement, we provide additional details for the examples considered and give the full, technical proofs of our main results. Furthermore, we include in Supplement I.C a more general consideration of distributions of Euclidean distances of a certain class of metric measure spaces and Supplement I.E contains additional material on simulation results and examples. Part II of the supplement contains several technical auxiliary results that seem to be folklore, but have not been written down explicitly in the literature, to the best of our knowledge. An R package, that implements the test proposed in Section 3 (see (15)), is available at `https://anonymous.4open.science/r/Distribution-of-Distances-67DC/`.

2 Limit Distributions

In this section, we investigate the limit behavior of the proposed test statistic under the hypothesis H_0 in (3), where it holds $\mu^U = \mu^V$ (see Theorem 2.6), and under the alternative H_1 in (3), where we have $\mu^U \neq \mu^V$ (see Theorem 2.7).

2.1 Conditions on the distributions of distances

Before we come to the limit distributions of the test statistic $\widehat{DoD}_{(\beta)}$ under H_0 and H_1, we discuss Condition 1.2 and Condition 1.3. We ensure that these conditions comprise reasonable assumptions on metric measure spaces that are indeed met in some standard examples.

Example 2.1. Let $\mathcal{X}_1$ be the unit square in $\mathbb{R}^2$, $d_{\mathcal{X}_1}(x,y) = ||x-y||_\infty$ for $x,y \in \mathbb{R}^2$ and let $\mu_{\mathcal{X}_1}$ the uniform distribution on $\mathcal{X}_1$. Let $X, X' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}_1}$. Then, a straightforward calculation shows that the density u_1 of $d_{\mathcal{X}_1}(X,X')$ is given as $u_1(s) = 4s^3 - 12s^2 + 8s$, if $0 \leq s \leq 1$, and zero else. For an illustration of u_1 see Figure 3. Obviously, u_1 is strictly positive and continuous on $(0,1)$ and thus Condition 1.2 is fulfilled for any $\beta \in (0,1/2)$ in the present setting. Furthermore, we find that for $t \in (0,1)$ the quantile function of u_1 is given as $U_1^{-1}(t) = -\sqrt{1-\sqrt{t}} + 1$. Since

$$\left|(U_1^{-1})'(t)\right| = \frac{1}{4\sqrt{1-\sqrt{t}}\sqrt{t}} \leq t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}}$$

for $t \in (0,1)$, the requirements of Condition 1.3 are satisfied.

Given two random point clouds in $\mathbb{R}^d$, it is often natural to assume that they are uniform samples from some compact set and to compare them based on their Euclidean distances, which are easily computable. Even if both samples stem from a curve or a hypersurface, this approach might be reasonable, since the corresponding (possibly more meaningful) intrinsic distances are unknown in general. Hence, the distribution of distances of *standard Euclidean metric measure spaces* (i.e. metric measure spaces $(\mathcal{X}, d_\mathcal{X}, \mu_\mathcal{X})$, where $d_\mathcal{X}$ denotes the Euclidean distance and $\mu_\mathcal{X}$ the uniform distribution on $\mathcal{X}$) deserve special attention. In the following, let $\mathbb{S}^{d-1}$ be the unit sphere in $\mathbb{R}^d$.

Example 2.2. Let $\mathcal{X}_2$ denote a disc in $\mathbb{R}^2$ with diameter 1 and let $\mathcal{X}_3$ denote a square in $\mathbb{R}^2$ with diameter 1. Let $\mathcal{X}_4$ be the the sphere $\mathbb{S}^1$, $\mathcal{X}_5$ the sphere $\mathbb{S}^2$ and $\mathcal{X}_6$ the sphere $\mathbb{S}^4$. Furthermore, let $\mathcal{X}_7 = [0,1]^2 \cup ([0,1] \times [4,5])$. Now, consider the standard Euclidean metric measure spaces induced by the sets $\mathcal{X}_i$ and denote by u_i the density of the distribution of distances of the respective space, $2 \leq i \leq 7$. All densities are illustrated in Figure 3. In Section B.1 of the supplement, we carefully check which of the just defined metric measure spaces meet the requirements of Condition 1.2 for all $\beta > 0$ and which meet the requirements of Condition 1.3. Our findings show that only the distribution of distances of $\mathcal{X}_4$ and $\mathcal{X}_7$ fail to meet the conditions of Condition 1.3 and that only $\mathcal{X}_7$ fails to meet the requirements of Condition 1.2 for all $\beta > 0$ (indeed it does not meet the requirements of Condition 1.2 for any $\beta > 0$).

Due to the importance of standard Euclidean metric measure spaces, we will provide simpler conditions than Condition 1.2 and Condition 1.3 for these spaces in the following. However, Example 2.2 already suggests that the distributions of distances of standard Euclidean metric measure spaces based on curves or surfaces might be less regular and more complex ($\mathcal{X}_3$ does not meet the requirements of Condition 1.3, the density u_3 is unbounded) than those of spaces based on sets with non-empty interior. Therefore, we concentrate in a first

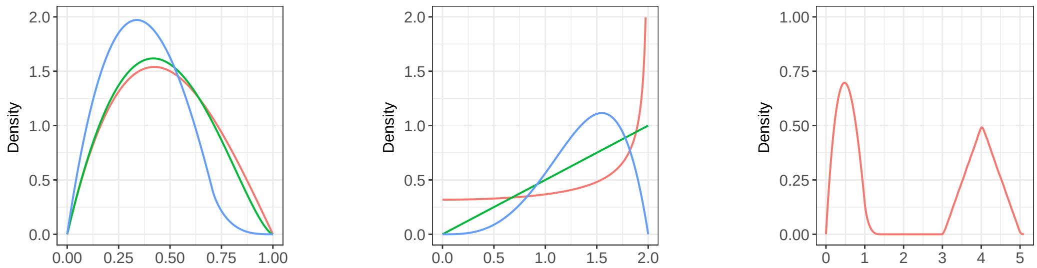

Fig. 3: Distribution of distances II: Representation of the densities u_1 (left, red), u_2 (left, green), u_3 (left, blue), u_4 (middle, red), u_5 (middle, green), u_6 (middle, blue) and u_7 (right) calculated in Example 2.1 and Example 2.2.

step on spaces of the latter kind. We require some notation. Let λ_d denote the Lebesgue measure in $\mathbb{R}^d$ and let $A \subset \mathbb{R}^d$, $d \geq 2$, be a bounded Borel set with $\lambda_d(A) > 0$. Recall that $y \in \mathbb{R}^d$ is determined by its polar coordinates (t, v), where $t = \|y\|_2$ and $v \in \mathbb{S}^{d-1}$ is the unit length vector y/t. Thus, we define the *covariance function* [63, Sec. 3.1] for $y = tv \in \mathbb{R}^d$ as $K_A(t,v) = K_A(y) = \lambda_d\left(A \cap (A - y)\right)$, where $A - y = \{a - y : a \in A\}$, and introduce the *isotropized set covariance function* [63, Sec. 3.1]

$$k_A(t) = \frac{1}{(\lambda_d(A))^2} \int_{\mathbb{S}^{d-1}} K_A(t,v)\, dv.$$

Furthermore, we define the diameter of a given metric space $(\mathcal{X}, d_{\mathcal{X}})$ as $\operatorname{diam}(\mathcal{X}) = \sup\{d_{\mathcal{X}}(x_1, x_2) : x_1, x_2 \in \mathcal{X}\}$.

Lemma 2.3. *Let $\mathcal{X} \subset \mathbb{R}^d$, $d \geq 2$, be a compact Borel set with $\lambda_d(\mathcal{X}) > 0$, $d_{\mathcal{X}}$ the Euclidean metric and $\mu_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$. Let $\operatorname{diam}(\mathcal{X}) = D$.*

(i) If $k_{\mathcal{X}}$ is strictly positive on $[0, D)$, then the induced metric measure space $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ meets the requirements of Condition 1.2 for any $\beta \in (0, 1/2)$.

(ii) If additionally there exists $\epsilon > 0$ and $\eta > 0$ such that

1. *the function $k_{\mathcal{X}}$ is monotonically decreasing on $(D - \epsilon, D)$;*
2. *we have $k_{\mathcal{X}}(t) \geq c_{\mathcal{X}}(D - t)^\eta$ for $t \in (D - \epsilon, D)$, for some $0 < c_{\mathcal{X}} < \infty$,*

then $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ also fulfills the requirements of Condition 1.3.

Remark 2.4. Let $\mathcal{X} \subset \mathbb{R}^{d_1}$ with $\lambda_{d_1}(\mathcal{X}) > 0$ and $\mathcal{Y} \subset \mathbb{R}^{d_2}$ with $\lambda_{d_2}(\mathcal{Y}) = 0$, $d_1 \leq d_2$, and consider the standard Euclidean metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ induced by $\mathcal{X}$ and $\mathcal{Y}$, respectively. Suppose that $\phi : \mathcal{X} \to \mathcal{Y}$ is a measure preserving isometry (i.e. $\phi\#\mu_{\mathcal{X}} = \mu_{\mathcal{Y}}$) and let $X, X' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ independent of $Y, Y' \overset{i.i.d.}{\sim} \mu_{\mathcal{Y}}$. Then, it clearly holds that the distributions of distances of both spaces agree, i.e., $d_{\mathcal{X}}(X,X') \stackrel{D}{=} d_{\mathcal{Y}}(Y,Y')$. Hence, Lemma 2.3 can be applied to $\mathcal{Y}$ as well.

The full proof of the above lemma is deferred to Section B.2 of the supplement. To conclude this subsection, we remark that we investigate the distributions of distances of standard Euclidean metric measure spaces based on various curves and hypersurfaces in Section C of the supplement. There, we derive, under several technical assumptions, an analogue of Lemma 2.3.

2.2 The Hypothesis

Throughout this subsection we assume that the distributions of distances of the two considered metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ are equal, i.e., that $\mu^U = \mu^V$. Further, assume that $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and $Y_1, \ldots, Y_m \stackrel{i.i.d.}{\sim} \mu_{\mathcal{Y}}$ are two independent samples. In order to study the finite sample bias of the statistic $\widehat{DoD}_{(\beta)}$, the following bound is helpful (for its proof see Section B.3 of the supplementary material).

Theorem 2.5. *Let $\beta \in [0, 1/2)$, let $\mu^U = \mu^V$ and let Setting 1.1 be met. Further, let*

$$J_2\left(\mu^U\right) = \int_{-\infty}^{\infty} \frac{U(t)(1-U(t))}{u(t)}\, dt < \infty. \tag{10}$$

Then it holds for $m, n \geq 3$ that

$$\mathbb{E}\left[\widehat{DoD}_{(\beta)}\right] \leq \left(\frac{8}{n+1} + \frac{8}{m+1}\right) J_2(\mu^U).$$

The next theorem states that $\widehat{DoD}_{(\beta)}$, based on $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and $Y_1, \ldots, Y_m \stackrel{i.i.d.}{\sim} \mu_{\mathcal{Y}}$, converges, appropriately scaled, in distribution to the integral of a squared Gaussian process. This will allow us to construct an asymptotic level α test using (estimates of) the theoretical $1-\alpha$ quantiles of Ξ, denoted as $\xi_{1-\alpha}$, in (9). The case $\beta \in (0, 1/2)$ is considered in part (i), whereas the case $\beta = 0$ is considered in part (ii).

Theorem 2.6. *Assume Setting 1.1 and suppose that $\mu^U = \mu^V$.*

(i) Let Condition 1.2 be met and let $m, n \to \infty$ such that $n/(n+m) \to \lambda \in (0,1)$. Then,

$$\frac{nm}{n+m} \int_{\beta}^{1-\beta} \left(U_n^{-1}(t) - V_m^{-1}(t)\right)^2 dt \rightsquigarrow \Xi := \int_{\beta}^{1-\beta} \mathbb{G}^2(t)\, dt,$$

where $\mathbb{G}$ is a centered Gaussian process with covariance

$$\operatorname{Cov}\left(\mathbb{G}(t), \mathbb{G}(t')\right) = \frac{4}{(u \circ U^{-1}(t))(u \circ U^{-1}(t'))} \Gamma_{d_{\mathcal{X}}}(U^{-1}(t), U^{-1}(t')). \tag{11}$$

Here,

$$\Gamma_{d_{\mathcal{X}}}(t,t') = \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t\}}\, d\mu_{\mathcal{X}}(y) \int \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t'\}}\, d\mu_{\mathcal{X}}(y)\, d\mu_{\mathcal{X}}(x)$$
$$- \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t\}}\, d\mu_{\mathcal{X}}(y)\, d\mu_{\mathcal{X}}(x) \int\int \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t'\}}\, d\mu_{\mathcal{X}}(y)\, d\mu_{\mathcal{X}}(x).$$

(ii) If we assume Condition 1.3 instead of Condition 1.2, then the analogous statement holds for the untrimmed version, i.e., for $\beta = 0$.

The full proof of Theorem 2.6 can be found in Section B.4 of the supplementary material.

2.3 The Alternative

In this subsection, we are concerned with the behavior of $\widehat{DoD}_{(\beta)}$ given that the distributions of distances of the metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ do not coincide. We distinguish the cases $\beta \in (0, 1/2)$ and $\beta = 0$.

Theorem 2.7. *Assume Setting 1.1.*

(i) Assume that Condition 1.2 holds, let $m, n \to \infty$ such that $\frac{n}{n+m} \to \lambda \in (0,1)$ and let $DoD_{(\beta)} > 0$. Then, it follows that

$$\sqrt{\frac{nm}{n+m}}\left(\widehat{DoD}_{(\beta)} - DoD_{(\beta)}\right)$$

converges in distribution to a normal distribution with mean 0 and variance

$$16\lambda \int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)} \int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)} (x - V^{-1}(U(x)))(y - V^{-1}(U(y)))\Gamma_{d_{\mathcal{X}}}(x,y)\, dxdy$$
$$+16(1-\lambda) \int_{V^{-1}(\beta)}^{V^{-1}(1-\beta)} \int_{V^{-1}(\beta)}^{V^{-1}(1-\beta)} (U^{-1}(V(x)) - x)(U^{-1}(V(y)) - y)\Gamma_{d_{\mathcal{Y}}}(x,y)\, dxdy.$$

Here, $\Gamma_{d_{\mathcal{X}}}(x,y)$ is as defined in Theorem 2.6 and $\Gamma_{d_{\mathcal{Y}}}(x,y)$ is defined analogously.

(ii) If we assume Condition 1.3 instead of Condition 1.2, then the analogous statement holds for the untrimmed version, i.e., for $\beta = 0$.

The proof of Theorem 2.7 is given in Section B.5 of the supplementary material.

Remark 2.8. The assumptions of Theorem 2.7 (i) include that β is chosen such that $DoD_{(\beta)} > 0$. Suppose on the other hand that $\mu^U \neq \mu^V$, but $DoD_{(\beta)} = 0$, i.e., their quantile functions agree Lebesgue a.e. on the interval $[\beta, 1-\beta]$. Then, the limits found in Theorem 2.7 are degenerate and it is easy to verify along the lines of the proof of Theorem 2.6 that $\widehat{DoD}_{(\beta)}$ exhibits the same distributional limit as in the case $DoD_{(\beta)} = 0$.

Remark 2.9. As noted by a referee, it is possible to slightly relax the assumptions of Theorem 2.7 (ii). It is sufficient to assume that U and V admit continuous densities that are strictly positive on the interior of their respective support and that $J_2(\mu^U)$ (see (10) for a definition) as well as $J_2(\mu^V)$ are finite (see Section B.5.2 of the supplement for more information). These relaxed assumptions are related to those of Proposition 2.3 of Del Barrio et al. [26].

Remark 2.10. An immediate application for Theorem 2.7 is testing for relevant differences (or equivalence testing), i.e., testing $H : DoD_{(\beta)} \leq \epsilon$ vs $K : DoD_{(\beta)} > \epsilon$ (or K vs H) for some specified $\epsilon > 0$ (see [27, 48]). In both cases, the quantiles required for testing are quantiles of the limiting normal distribution. Hence, a consistent estimator for the limiting variance (e.g. a plug-in estimator) yields consistent estimates for the quantiles required.

Remark 2.11. So far we have restricted ourselves to the case $p = 2$. However, most of our findings transfer to results for the statistic $\widehat{DoD}_p$, $p \in [1, \infty)$, in (4). Using the same ideas one can directly derive Theorem 2.5 and Theorem 2.6 for (a trimmed version of) $\widehat{DoD}_p$ (see Sections B.3 and B.4 of the supplement) under slightly different assumptions. Only the proof of Theorem 2.7 requires more care (see Section B.4 in the supplement).

3 Bootstrapping the Quantiles

The quantiles of the limit distribution of $\widehat{DoD}_{(\beta)}$ under H_0 depend on the unknown distribution U and are therefore in general not accessible. One possible approach, which is quite cumbersome, is to estimate the covariance matrix of the Gaussian limit process G from the data and use this to approximate the quantiles required. Alternatively, we suggest to directly bootstrap the quantiles of the limit distribution of $\widehat{DoD}_{(\beta)}$ under H_0. To this end, we define and investigate the bootstrap versions of U_n, U_n^{-1} and $\mathbb{U}_n^{-1} := \sqrt{n}\left(U_n^{-1} - U^{-1}\right)$.

Let μ_n denote the empirical measure based on the sample $X_1, \ldots, X_n$. Given the sample values, let $X_1^*, \ldots, X_{n_B}^*$ be an independent identically distributed sample of size n_B from μ_n. Then, the bootstrap estimator of U_n is defined as

$$U_{n_B}^*(t) := \frac{2}{n_B(n_B-1)} \sum_{1 \leq i < j \leq n_B} \mathbb{1}_{\{d_{\mathcal{X}}(X_i^*, X_j^*) \leq t\}},$$

the bootstrap empirical U-process is for $t \in \mathbb{R}$ given as $\mathbb{U}^*_{n_B}(t) = \sqrt{n_B}(U^*_{n_B}(t) - U_n(t))$ and the corresponding (empirical) bootstrap quantile process for $t \in (0,1)$ as

$$\left(\mathbb{U}^*_{n_B}\right)^{-1}(t) = \sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1}(t) - U_n^{-1}(t)\right).$$

One can easily verify along the lines of the proof of Theorem 2.6 that for $n \to \infty$ it also holds for $\beta \in (0,1/2)$

$$\int_\beta^{1-\beta} \left(\mathbb{U}_n^{-1}(t)\right)^2 dt \rightsquigarrow \Xi = \int_\beta^{1-\beta} \mathbb{G}^2(t)\, dt. \tag{12}$$

This suggests to approximate the quantiles of Ξ by the bootstrapped ones of

$$\Xi^*_{n_B} := \int_\beta^{1-\beta} \left(\left(\mathbb{U}^*_{n_B}\right)^{-1}(t)\right)^2 dt. \tag{13}$$

Let $\beta \in (0,1/2)$, suppose that Condition 1.2 holds, let $\sqrt{n_B} = o(n)$ and let $\xi^{(R)}_{n_B,\alpha}$ denote the empirical bootstrap quantile of R independent bootstrap realizations $\Xi^{*,(1)}_{n_B}, \ldots, \Xi^{*,(R)}_{n_B}$. Under these assumptions, we derive (cf. Section D of the supplement) that for any $\alpha \in (0,1)$,

$$\lim_{n,n_B,R\to\infty} \mathbb{P}\left(\int_\beta^{1-\beta} \left(\mathbb{U}_n^{-1}(t)\right)^2 dt \geq \xi^{(R)}_{n_B,\alpha}\right) = \alpha. \tag{14}$$

Because of (12) the statement (14) guarantees the consistency of $\xi^{(R)}_{n_B,\alpha}$ for $n, n_B, R \to \infty$. Hence, a consistent bootstrap analogue of the test defined by the decision rule (9) is for $\beta \in (0,1/2)$ given by the *bootstrapped Distribution of Distances (DoD)*-test

$$\Phi^*_{DoD}(\mathcal{X}_n, \mathcal{Y}_m) = \begin{cases} 1, & \text{if } \frac{nm}{n+m}\widehat{DoD}_{(\beta)} > \xi^{(R)}_{n_B,1-\alpha} \\ 0, & \text{if } \frac{nm}{n+m}\widehat{DoD}_{(\beta)} \leq \xi^{(R)}_{n_B,1-\alpha}. \end{cases} \tag{15}$$

4 Simulations

We investigate the finite sample behavior of $\widehat{DoD}_{(\beta)}$ in Monte Carlo simulations. To this end, we simulate the speed of convergence of $\widehat{DoD}_{(\beta)}$ under H_0 to its limit distribution (see Theorem 2.6). Moreover, we showcase the accuracy of the approximation by the bootstrap scheme proposed in Section 3 and investigate what kind of differences are detectable in the finite sample setting using the bootstrapped DoD-test Φ^*_{DoD} defined in (15). Based on Theorem 2.7, it is further possible to test $H : DoD_{(\beta)} \leq \epsilon$ vs $K : DoD_{(\beta)} > \epsilon$ for some specified $\epsilon > 0$ (see Remark 2.10). However, only few distributions of distances are known explicitly and hence the choice of ϵ is slightly problematic. Therefore and due to page restrictions, we did not include this application in the paper. All simulations were performed in R (R Core Team [43]). In order to increase the readability of this section, several tables have been postponed to Section E.1 of the supplementary material.

4.1 The Hypothesis

We begin with the simulation of the finite sample distribution under the hypothesis and consider the metric measure space $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ from Example 2.1, where $\mathcal{X}$ denotes the unit square in $\mathbb{R}^2$, $d_{\mathcal{X}}$ the distance induced by the supremum norm and $\mu_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$. We generate for $n = m = 10, 50, 100, 250$ two samples $\mathcal{X}_n$ and $\mathcal{X}'_n$ of $\mu_{\mathcal{X}}$ and calculate for $\beta = 0.01$ the statistic $\frac{n}{2}\widehat{DoD}_{(\beta)}$. For each n, we repeat this process 10,000 times. The finite sample distribution is then compared to a Monte Carlo sample of its theoretical limit distribution (sample size 10,000). Kernel density estimators (Gaussian kernel with bandwidth given by Silverman's rule) and Q-Q-plots are displayed in Figure 4. All plots highlight that the finite sample distribution of $\widehat{DoD}_{(\beta)}$ is already well approximated by its theoretical limit distribution for moderate sample sizes. Moreover, for $n = 10$ the quantiles of the finite sample distribution of $\widehat{DoD}_{(\beta)}$ are in general larger than the ones of the sample of its theoretical limit distribution, which suggests that the DoD-test will be rather conservative for small n. For $n \geq 50$ most quantiles of the finite sample distribution of $\widehat{DoD}_{(\beta)}$ match the ones of its theoretical limit distribution reasonably well.

4.2 The Bootstrap Test

We now investigate the finite sample properties of the bootstrap test Φ^*_{DoD} (defined in (15)). To this end, let $\mu_{\mathcal{W}_1}$ denote the uniform distribution on a 3D-pentagon (inner pentagon side length: 1, Euclidean distance between inner and outer pentagon: 0.4, height: 0.4) and let $\mu_{\mathcal{W}_6}$ denote the uniform distribution on a torus (center radius: 1.169, tube radius: 0.2) with the same center and orientation (see the plots for $t_0 = 0$ and $t_6 = 1$ in Figure 5). To interpolate between these spaces, we consider $\Pi_{\mu_{\mathcal{W}_1}}^{\mu_{\mathcal{W}_6}}(t)$, $t \in [0, 1]$, the 2-Wasserstein geodesic between $\mu_{\mathcal{W}_1}$ and $\mu_{\mathcal{W}_6}$ (see e.g. Santambrogio [56, Sec. 5.4] for a formal definition). Figure 5 displays for $t_i \in \{0, 0.1, 0.2, 0.4, 0.6, 1\}$ the metric measure spaces $(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})$, where $\mathcal{W}_i = \text{supp}\left(\Pi_{\mu_{\mathcal{W}_1}}^{\mu_{\mathcal{W}_6}}(t_i)\right)$, $d_{\mathcal{W}_i}$ denotes the Euclidean distance and $\mu_{\mathcal{W}_i} = \Pi_{\mu_{\mathcal{W}_1}}^{\mu_{\mathcal{W}_6}}(t_i)$, discretely approximated based on 40,000 points with the WSGeometry-package [34].

Before we employ the bootstrap DoD-test with $\beta = 0.01$ to compare $\mathcal{W}_1$ to the spaces $\{\mathcal{W}_i\}_{i=1}^6$, we consider the bootstrap approximation proposed in Section 3 in this setting. Therefore, we generate $n = 100, 250, 500, 1000$ realizations of $\mu_{\mathcal{W}_1}$ and calculate for $n_B = n$

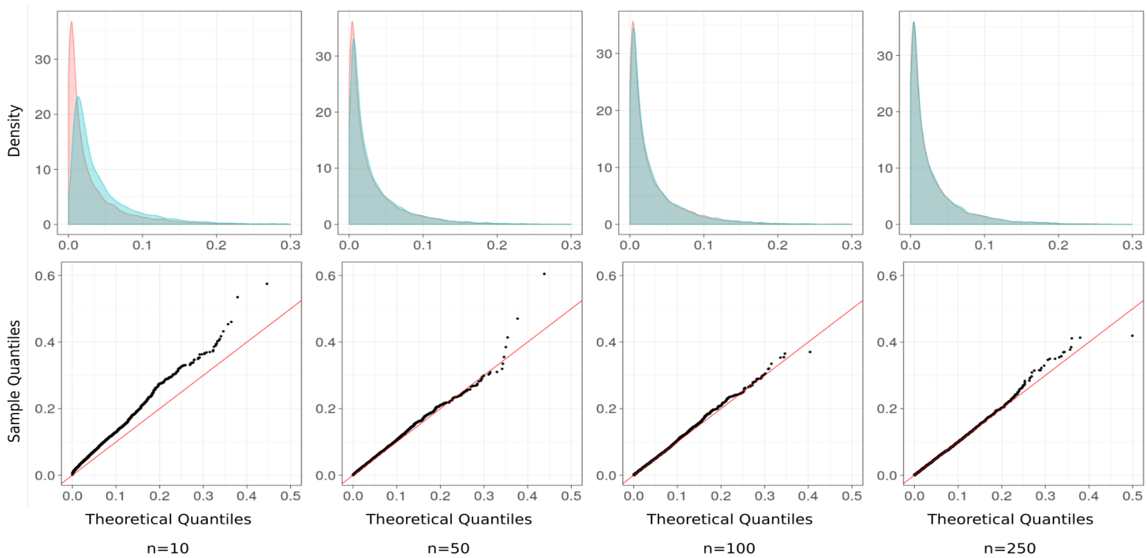

Fig. 4: Finite sample accuracy of the limit law under the hypothesis: Upper row: Kernel density estimators of the sample of $\widehat{DoD}_{(\beta)}$ (in blue) and a Monte Carlo sample of its theoretical limit distribution (in red, sample size 10,000) for $n = 10, 50, 100, 250$ (from left to right). Lower row: The corresponding Q-Q-plots.

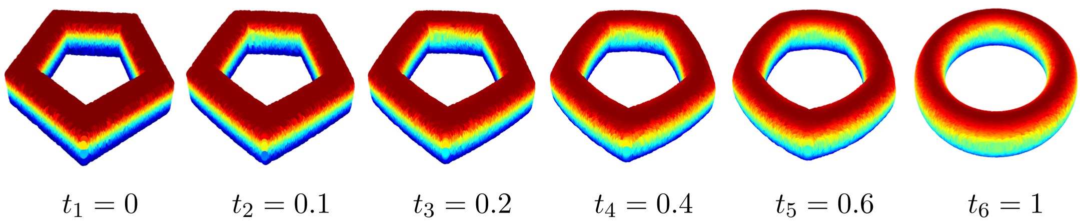

Fig. 5: Different metric measure spaces: A graphical illustration of the metric measure spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^{6}$.

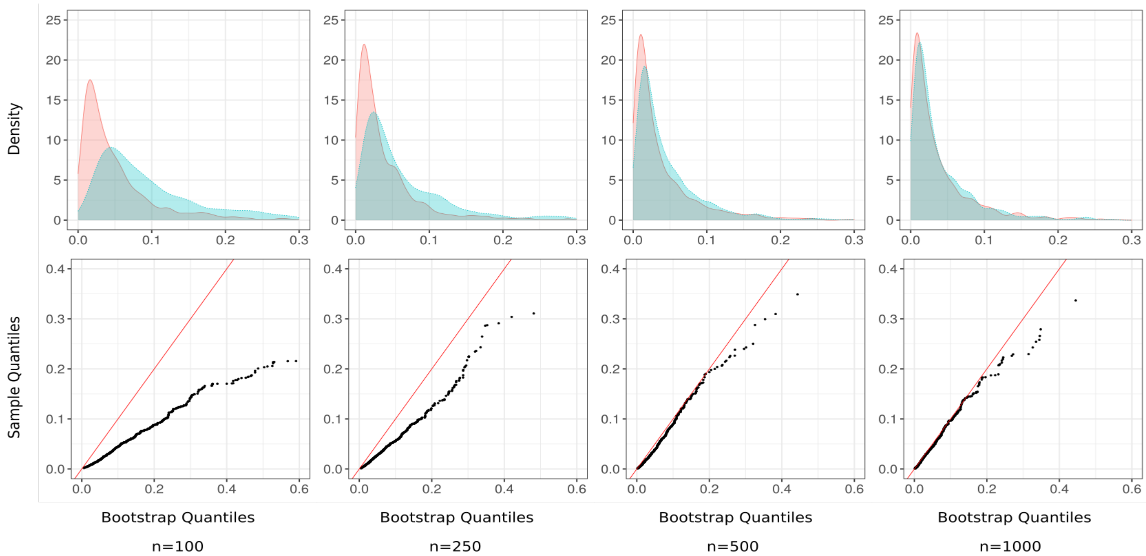

Fig. 6: Bootstrap under the hypothesis: Illustration of the n out of n plug-in bootstrap approximation for the statistic $\widehat{DoD}_{(\beta)}$ based on two samples from $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$. Upper row: Kernel density estimators of 1000 realizations of $\widehat{DoD}_{(\beta)}$ (in red) and its bootstrap approximation (blue, 1000 replications) for $n = 100, 250, 500, 1000$ (from left to right). Lower row: The corresponding Q-Q-plots.

based on these samples 1000 times

$$\Xi^*_{n_B} = \int_{0.01}^{0.99} \left(\left(\mathbb{U}^*_{n_B} \right)^{-1} (t) \right)^2 dt$$

as described in Section 3. We then compare for the different n the obtained finite sample distributions to ones of $\widehat{DoD}_{(\beta)}(\mathcal{W}_{1,n}, \mathcal{W}'_{1,n})$ ($\mathcal{W}_{1,n}$ and $\mathcal{W}'_{1,n}$ denote two independent samples of $\mu_{\mathcal{W}_1}$ of size n). The results are summarized as kernel density estimators (Gaussian kernel with bandwidth given by Silverman's rule) and Q-Q-plots in Figure 6. Both, the kernel density estimators and the Q-Q-plots show that for $n \leq 250$ the bootstrap quantiles are clearly larger than the empirical quantiles leading to a rather conservative procedure for smaller n, an effect that disappears for large n.

Next, we aim to apply Φ^*_{DoD} for $\beta = 0.01$ at 5%-significance level for discriminating between $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ and each of the spaces $(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})$, $i = 1, \ldots, 6$. To this end, we bootstrap the quantile $\xi_{0.95}$ based on samples from $\mu_{\mathcal{W}_1}$ as described in Section 3 ($R = 1000$) and then we apply the test Φ^*_{DoD}, defined in (15), with the bootstrapped quantile $\xi^{(R)}_{n_B,\alpha}$ on 1000 samples of size $n = 100, 250, 500, 1000$ as illustrated in Section 3. We find that the prespecified significance level (see $t_1 = 0$) is never exceeded and the test is rather conservative for smaller n. Concerning the power of the test Φ^*_{DoD}, we observe that it consistently increases with the increasing Wasserstein distance between the measures

$\mu_{\mathcal{W}_i}$, $1 \leq i \leq 6$. For $n \geq 250$ the differences between $\mathcal{W}_1$ and $\mathcal{W}_i$, $4 \leq i \leq 6$ (see Figure 5) are clearly detected. If we choose $n = 1000$, even the spaces $\mathcal{W}_1$ and $\mathcal{W}_3$ (that correspond to t_1 and t_3) are almost always discriminated, although in this case, the (approximated) Wasserstein distance between $\mu_{\mathcal{W}_1}$ and $\mu_{\mathcal{W}_3}$ is smaller than 0.034. The test even develops some power for the comparison of $\mathcal{W}_1$ and $\mathcal{W}_2$ despite their strong similarity (see Figure 5). The detailed results are summarized in Table E.1 in Section E.1 of the supplement.

In order to highlight how much power we gain in the finite sample setting by carefully handling the occurring dependencies we repeat the above comparisons, but calculate $\widehat{DoD}_{(\beta)}$ only based on the independent distances, i.e., on the distances $\{d_{\mathcal{X}}(X_1, X_2), d_{\mathcal{X}}(X_3, X_4), \ldots, d_{\mathcal{X}}(X_{n-1}, X_n)\}$ and $\{d_{\mathcal{Y}}(Y_1, Y_2), d_{\mathcal{Y}}(Y_3, Y_4), \ldots, d_{\mathcal{Y}}(Y_{m-1}, Y_m)\}$, instead of all available distances. From now on this statistic is denoted as $\widehat{D}_{\beta,ind}$. Similarly, we construct an asymptotic level α test $\Phi_{D_{ind}}$ based on $\widehat{D}_{\beta,ind}$. The results for comparing $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ and $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ using $\Phi_{D_{ind}}$ with $\beta = 0.01$ are displayed in Table E.2 in Section E.1 of the supplementary material. Apparently, $\Phi_{D_{ind}}$ keeps its prespecified significance level of $\alpha = 0.05$, but develops significantly less power than Φ^*_{DoD} in the finite sample setting.

Furthermore, we investigate the influence of β on our results. To this end, we repeat the previous comparisons with $n = 500$ and $\beta = 0, 0.01, 0.05, 0.25$. It highlights that the test Φ^*_{DoD} holds its level for all β. While the results are overall comparable, we observe some slight differences for the various values of β. For instance, for $\beta = 0.25$ the test Φ^*_{DoD} develops slightly more power for the comparison of $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ and $(\mathcal{W}_3, d_{\mathcal{W}_3}, \mu_{\mathcal{W}_3})$ than for $\beta = 0$. Apparently, in this case the respective (true) distributions of distances strongly resemble each other for small and large distances and the comparison of $\mathcal{W}_1$ and $\mathcal{W}_3$ becomes to some degree more informative, if we do not consider these distances. All results are summarized in Table E.3 in Section E.1 of the supplement.

To conclude this subsection, we remark that in the above simulations the quantiles required for the applications of Φ^*_{DoD} were always estimated based on samples of $\mu_{\mathcal{Y}}$. Evidently, this slightly affects the results obtained, but we found that this influence is not significant.

5 Structural Protein Comparisons

Next, we apply the DoD-test to compare the protein structures displayed in Figure 2. First, we compare 5D0U with itself, in order to investigate the actual significance level of the proposed test under H_0 in a realistic example. Afterwards, 5D0U is compared with 5JPT and with 6FAA, respectively. However, before we can apply Φ^*_{DoD}, we need to model proteins as metric measure spaces. Thus, we briefly recap some well known facts about proteins to motivate the subsequent approach. A protein is a polypeptide chain made up of amino acid residues linked together in a definite sequence. Tracing the repeated amide, C^{α} and carbonyl atoms of each amino acid residue, a so called *backbone* can be identified. It is well established that the distances between the C^α atoms of the backbone contain most of the information about the protein's structure [35, 36, 55]. In order to verify that the test is able to compare protein structures based on subsamples (which might be important for database queries), we randomly select $n = 10, 50, 100, 250, 500$ from the 650-750 C^α atoms of the respective proteins and assume that the corresponding coordinates are samples of unknown distributions $\{\mu_{\mathcal{X}_i}\}_{i=1}^3$ supported on Borel sets $\mathcal{X}_i \subset \mathbb{R}^3$ with $\lambda_3(\mathcal{X}_i) > 0$ that are equipped with the Euclidean distance. We stress that although the backbone of a protein is usually represented as a curve in $\mathbb{R}^3$ (see e.g. Figure 2), it is important to note that these representations are extracted from indirect, noisy observations of the electron density (see [53]). In consequence, it is more realistic to assume that positions are drawn from a tube-like structure with non-empty interior. We choose $\beta = 0.01$, $\alpha = 0.05$ and determine for each n the bootstrap quantile $\xi^{(R)}_{n_B,0.95}$ based on a sample of size n from 5D0U ($R = 1000$, $n_B = n$) as illustrated in Section 3. This allows us to directly apply the test Φ^*_{DoD} on the drawn samples. The results of our comparisons are summarized in Figure 7. It displays the empirical significance level resp. the empirical power of the proposed method as a function of n.

5D0U vs 5D0U: In accordance with the previous simulation study this comparison (see Figure 7, left) shows that Φ^*_{DoD} is conservative in this application as well.

5D0U vs 5JPT: We have already mentioned in Section 1.3 that 5D0U and 5JPT are structures of the same protein from two different organisms and thus highly similar (their alignment has a root mean deviation of less than 0.59 Å). The empirical power for this comparison (Figure 7, middle) stays for all n below $\alpha = 0.05$. Thus, the test does not discriminate between the two protein structures in accordance with our biological knowledge.

5D0U vs 6FAA: Although the protein structures 5D0U and 6FAA are similar at large parts (their alignment has a root mean square deviation of 0.75 Å), the DoD-test is able to discriminate between them with high statistical power. The empirical power (Figure 7, right) is a strictly monotonically increasing function in n that is greater than 0.63 for $n \geq 100$ and approaches 1 for $n = 500$ (recall that we use random samples of the $650-750$ C^α atoms).

Finally, we remark that throughout this section we have always based the quantiles required for testing on samples of the protein structure 5D0U. By the definition of Φ^*_{DoD} it is evident that this influences the results. If we compared the proteins 6FAA and 5D0U using Φ^*_{DoD} with quantiles obtained by a sample of 6FAA, the results would change slightly, but remain comparable.

5.1 Comparison to the DTM-test

We investigate how the test proposed by Brécheteau [15], which is based on an empirical

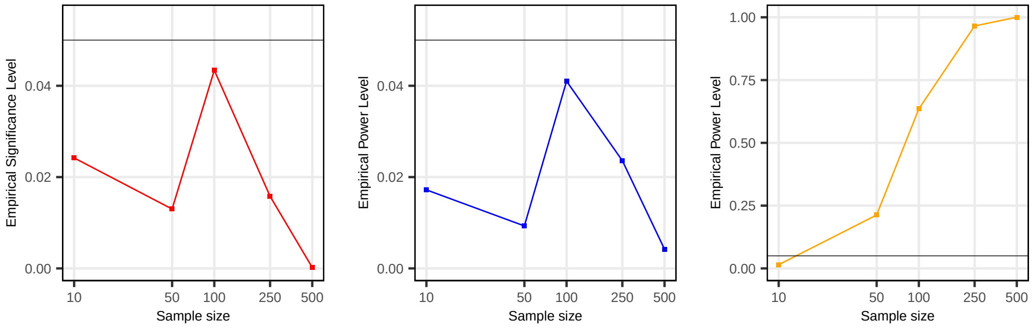

Fig. 7: Protein Structure Comparison: Empirical significance level for comparing 5D0U with itself (left), empirical power for the comparison of 5D0U with 5JPT (middle) as well as the the empirical power for comparing 5D0U with 6FAA (right). 1000 repetitions of the test Φ^*_{DoD} have been simulated for each n.

version of another lower bound for the Gromov-Kantorovich distance, compares to Φ^*_{DoD}. To this end, we first briefly introduce the method proposed in [15], empirically study various toy examples and analyze the differences of both tests for protein structure comparison. We summarize the results of these comparisons here and give the tables with the precise results in Section E.2 of the supplement.

Let $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ denote a metric measure space with $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$. For $n_S \leq n$ the *empirical distance to measure signature* with mass parameter $\kappa = k/n$ is then defined as

$$D_{\mathcal{X}_n,\kappa}(n_S) := \frac{1}{n_S} \sum_{i=1}^{n_S} \left(\frac{1}{k} \sum_{j=1}^{k} d_{\mathcal{X}} \left(X_i, X_i^{(j)} \right) \delta_{\mathcal{X}_n,\kappa}(X_i) \right), \tag{16}$$

where $X_i^{(j)}$ denotes the j'th nearest neighbor of X_i in the sample $\mathcal{X}_n$ (for general κ see Brécheteau [15]). In particular, we observe that $D_{\mathcal{X}_n,\kappa}(n_S)$ denotes a discrete probability distribution on $\mathbb{R}$. Let $D_{\mathcal{Y}_n,\kappa}(n_S)$ be defined analogously to (16). Then, given that $\frac{n_S}{n} = o(1)$, Brécheteau [15] constructs an asymptotic level α test for H_0^* defined in Section 1.1 based on the 1-Kantorovich distance between the respective empirical distance to measure signatures, i.e., on the test statistic

$$T_{n_S,\kappa}(\mathcal{X}_n, \mathcal{Y}_n) := \mathcal{K}_1 \left(D_{\mathcal{X}_n,\kappa}(n_S), D_{\mathcal{Y}_n,\kappa}(n_S) \right). \tag{17}$$

The corresponding test, that rejects if $T_{n_S,\kappa}(\mathcal{X}_n, \mathcal{Y}_n)$ exceeds a bootstrapped critical value q_α^{DTM}, is denoted as Φ_{DTM} in the following. Brécheteau [15] proves that, similar to $\widehat{DoD}_{(\beta)}$, the statistic $T_{n_S,\kappa}$ is a (subsampled) empirical version of a lower bound $T_\kappa(\mathcal{X}, \mathcal{Y})$ of the Gromov-Kantorovich distance (see [15, Sec. 1] for a formal definition). It is important to note that there are metric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $T_\kappa(\mathcal{X}, \mathcal{Y}) = 0$ although

$DoD_{(0)}(\mathcal{X},\mathcal{Y}) > 0$ and vice versa (see Section B.7 of the supplement detailed comparison of $T_\kappa(\mathcal{X},\mathcal{Y})$ and $DoD_{(0)}(\mathcal{X},\mathcal{Y})$).

We now compare both methods in two simulated examples. To this end, we first repeat the comparisons of $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ with the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Section 4.2 for the definitions) with Φ_{DTM}. Secondly, we simulate the empirical power of Φ^*_{DoD} in the setting of Section 4.2 of Brécheteau [15] for the comparison of different spiral types. For both comparisons, we choose a significance level of $\alpha = 0.05$. We remark that the test Φ_{DTM} is not easily applied in the finite sample setting. Although it is an asymptotic test of level α, the parameters n_S and κ have to be chosen carefully for the test to hold its prespecified significance level for finite samples. In particular, choosing n_S and κ large violates the assumption of (asymptotic) independence underlying the results of Brécheteau [15]. In both settings, we found comparable results. While the test Φ_{DTM} (just like Φ^*_{DoD}) approximately holds his α level in both frameworks ($\kappa \leq 0.1$ and $n_S \leq n/15$ for the comparison of the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6/\kappa = 0.05$ and $n_S = 20$ for spiral comparison of [15, Sec. 4.2]), the additional subsampling in the definition of $T_{n_S,\kappa}(\mathcal{X},\mathcal{Y})$ in (17) leads to a notable loss of power. The complete results of these comparisons can be found in Table E.4 and Table E.5 of Section E.2 of the supplement.

Finally, we come to the protein structure comparison. We repeat the previous comparisons of 5D0U, 5JPT and 6FAA for a significance level $\alpha = 0.05$, $n = 100, 250, 500$, $n_S = N/5$ and $\kappa = 0.05, 0.1$. The test Φ_{DTM} approximately holds its significance level and is more sensitive to small local changes such as slight shifts of structural elements for small mass parameters κ compared to Φ^*_{DoD}. However, the evident differences between 5D0U and 6FAA are detected much better by Φ^*_{DoD} (see Figure 7). The complete results of this numerical study are reported in Table E.6 (cf. Section E.2 of the supplement).

5.2 Discussion

We conclude this section with some remarks on the modeling of proteins as metric measure spaces. So far, we have treated all C^α atoms as equally important, although it appears to be reasonable for some applications to put major emphasis on the cores of the proteins. Further, one could have included that the error of measurement is in general higher for some parts of the protein by adjusting the measure on the considered space accordingly. We remark that throughout this section we have considered proteins as rigid objects and shown that this allows us to efficiently discriminate between them. However, it is well known that proteins undergo different conformational states. In such a case the usage of the Euclidean metric as done previously will most likely cause Φ^*_{DoD} to discriminate between the different conformations, as the Euclidean distance is not suited for the matching of flexible objects [31]. Depending on the application one might want to take this into account by adopting a different metric reflecting (estimates of the) corresponding intrinsic distances and to modify the theory developed. Conceptually, this is straightforward but beyond the scope of this illustrative example.

Acknowledgements

We are grateful to F. Mémoli, V. Liebscher and M. Wendler for interesting discussions and helpful comments and to C. Brécheteau for providing R code for the application of the DTM-test. Support by the DFG Research Training Group 2088 Project A1, CRC 1456 and Cluster of Excellence MBExC 2067 is gratefully acknowledged.

References

[1] P. C. Alvarez-Esteban, E. Del Barrio, J. A. Cuesta-Albertos, and C. Matran. Trimmed comparison of distributions. *Journal of the American Statistical Association*, 103(482):697–704, 2008.

[2] D. Alvarez-Melis and T. Jaakkola. Gromov-wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890, 2018.

[3] M. A. Arcones and E. Giné. U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Processes and their Applications*, 52(1):17–38, 1994.

[4] J. E. Arenas and J. N. Abelson. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings of the National Academy of Sciences*, 94(22):11798–11802, 1997.

[5] C. K. Au and M. M. F. Yuen. Feature-based reverse engineering of mannequin for garment design. *Computer-Aided Design*, 31(12), 1999.

[6] L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206, 2004.

[7] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), 2002.

[8] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL http://http://www.rcsb.org.

[9] J. R. Berrendero, A. Cuevas, and B. Pateiro-López. Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146:237–247, 2016.

[10] T. B. Berrett, I. Kontoyiannis, and R. J. Samworth. Optimal rates for independence testing via U-statistic permutation tests. *arXiv preprint arXiv:2001.05513*, 2020.

[11] P. Berthet and J.-C. Fort. Weak convergence of empirical Wasserstein type distances. *arXiv preprint arXiv:1911.02389*, 2019.

[12] P. Berthet, J.-C. Fort, and T. Klein. A central limit theorem for wasserstein type distances between two different real distributions. *Ann. Inst. Henri Poincaré Probab. Stat., hal-01526879*, 2018.

[13] S. Bobkov and M. Ledoux. *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, volume 261. American Mathematical Society, 2019.

[14] M. Boutin and G. Kemper. On reconstructing n-point configurations from the distribution of distances or areas. *Advances in Applied Mathematics*, 32(4), 2004.

[15] C. Brécheteau. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature. *Electronic Journal of Statistics*, 13(1):795–849, 2019.

[16] D. Brinkman and P. J. Olver. Invariant histograms. *The American Mathematical Monthly*, 119(1): 4–24, 2012.
[17] C. Bunne, D. Alvarez-Melis, A. Krause, and S. Jegelka. Learning generative models across incomparable spaces. In *International Conference on Machine Learning*, pages 851–861. PMLR, 2019.
[18] B. Cárdenas, B. Eugenio, D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranić. Feature-based similarity search in 3D object databases. *ACM computing surveys*, 37(4), 2005.
[19] S. Chowdhury and F. Mémoli. The Gromov-Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.
[20] S. Chowdhury and T. Needham. Generalized spectral clustering via Gromov-Wasserstein learning. *arXiv preprint arXiv:2006.04163*, 2020.
[21] C. Czado and A. Munk. Assessing the similarity of distributions-finite sample performance of the empirical mallows distance. *Journal of Statistical Computation and Simulation*, 60(4):319–346, 1998.
[22] S. Dede. An empirical central limit theorem in L_1 for stationary sequences. *Stochastic Processes and their Applications*, 119(10):3494–3515, 2009.
[23] J. Dedecker, F. Merlevède, et al. Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α-dependent sequences. *Bernoulli*, 23(3):2083–2127, 2017.
[24] E. del Barrio, J. A. Cuesta-Albertos, C. Matrán, and J. M. Rodríguez-Rodríguez. Tests of goodness of fit based on the L_2-Wasserstein distance. *The Annals of Statistics*, 27(4), 1999.
[25] E. del Barrio, E. Giné, and F. Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, 11(1), 2005.
[26] E. Del Barrio, P. Gordaliza, and J.-M. Loubes. A central limit theorem for L^p transportation cost on the real line with application to fairness assessment in machine learning. *Information and Inference: A Journal of the IMA*, 8(4):817–849, 2019.
[27] H. Dette, W. Wu, et al. Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *Annals of Statistics*, 47(6):3578–3608, 2019.
[28] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli. Euclidean distance matrices: Essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30, 2015.
[29] M. d'Amico, P. Frosini, and C. Landi. Natural pseudo-distance and optimal matching between reduced size functions. *Acta applicandae mathematicae*, 109(2):527–554, 2010.
[30] M. El-Mehalawi and A. R. Miller. A database system of mechanical components based on geometric and topological similarity. Part II: Indexing, retrieval, matching, and similarity assessment. *Computer-Aided Design*, 35(1), 2003.
[31] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *IEEE Transactions on pattern analysis and machine intelligence*, 25(10):1285–1295, 2003.
[32] M. Gellert, M. F. Hossain, F. J. F. Berens, L. W. Bruhn, C. Urbainsky, V. Liebscher, and C. H. Lillig. Substrate specificity of thioredoxins and glutaredoxins–towards a functional classification. *Heliyon*, 5 (12):e02943, 2019.
[33] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces.* vol. 152 of Progress in Mathematics (Birkhäuser), 1999.
[34] F. Heinemann and N. Bonneel. *WSGeometry: Compute Wasserstein Barycenters, Geodesics, PCA and Distances*, 2021. URL https://CRAN.R-project.org/package=WSGeometry. R package version 1.0.
[35] L. Holm and C. Sander. Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology*, 233(1):123–138, 1993.
[36] T. A. Jones and S. Thirup. Using known substructures in protein model building and crystallography. *The EMBO Journal*, 5(4):819–822, 1986.
[37] S.-H. Kim and R.-J. Lin. Spliceosome activation by Prp2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Molecular and Cellular Biology*, 16(12):6810–6819, 1996.
[38] R. Kolodny, P. Koehl, and M. Levitt. Comprehensive evaluation of protein structure alignment methods: Scoring by geometric measures. *Journal of Molecular Biology*, 346(4):1173–1188, 2005.

[39] E. Krissinel and K. Henrick. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D: Biological Crystallography*, 60(12), 2004.

[40] S. Lebaron, C. Froment, M. Fromont-Racine, J.-C. Rain, B. Monsarrat, M. Caizergues-Ferrer, and Y. Henry. The splicing ATPase Prp43p is a component of multiple preribosomal particles. *Molecular and Cellular Biology*, 25(21):9269–9282, 2005.

[41] J. S. Lomont. *Applications of Finite Groups*. Academic Press, 2014.

[42] D. M. Mason. Weak convergence of the weighted empirical quantile process in $L^2(0,1)$. *The Annals of Probability*, 12(1):243–255, 1984.

[43] R Core Team. R*: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/.

[44] F. Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. In *Proceedings Point Based Graphics*, 2007.

[45] F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.

[46] F. Mémoli and T. Needham. Distance distributions and inverse problems for metric measure spaces. *arXiv preprint arXiv:arXiv:1810.09646v3*, 2021.

[47] P. Montero-Manso and J. A. Vilar. Two-sample homogeneity testing: A procedure based on comparing distributions of interpoint distances. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(3):234–252, 2019.

[48] A. Munk and C. Czado. Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), 1998.

[49] D. Nolan and D. Pollard. Functional limit theorems for U-processes. *The Annals of Probability*, pages 1291–1298, 1988.

[50] R. Nussinov and H. J. Wolfson. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. *Proceedings of the National Academy of Sciences*, 88 (23), 1991.

[51] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4), 2002.

[52] P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22, 1991.

[53] G. Rhodes. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Elsevier, 2010.

[54] J. Robert-Paganin, M. Halladjian, M. Blaud, S. Lebaron, L. Delbos, F. Chardon, R. Capeyrou, O. Humbert, Y. Henry, A. K. Henras, et al. Functional link between DEAH/RHA helicase Prp43 activation and ATP base binding. *Nucleic acids research*, 45(3):1539–1552, 2016.

[55] M. G. Rossman and A. Liljas. Recognition of structural domains in globular proteins. *Journal of Molecular Biology*, 85(1):177–181, 1974.

[56] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 55. Springer, 2015.

[57] A. Schmitt, F. Hamann, P. Neumann, and R. Ficner. Crystal structure of the spliceosomal DEAH-Box ATPase Prp2. *Acta Crystallographica Section D: Structural Biology*, 74(7):643–654, 2018.

[58] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[59] D. Sejdinovic, B. Sriperumbudur, A. Gretton, K. Fukumizu, et al. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291, 2013.

[60] J. Solomon, G. Peyré, V. G. Kim, and S. Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)*, 35(4):72, 2016.

[61] M. Sommerfeld and A. Munk. Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):219–238, 2018.

[62] S. Srivastava, S. B. Lal, D. C. Mishra, U. Angadi, K. Chaturvedi, S. N. Rai, and A. Rai. An efficient algorithm for protein structure comparison using elastic shape analysis. *Algorithms for Molecular*

Biology, 11(1):27, 2016.

[63] D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* John Wiley & Sons, 2008.

[64] C. Tameling, M. Sommerfeld, and A. Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *The Annals of Applied Probability*, 29(5):2744–2781, 2019.

[65] M. J. Tauchert, J.-B. Fourmann, H. Christian, R. Lührmann, and R. Ficner. Structural and functional analysis of the RNA helicase Prp43 from the thermophilic tukaryote chaetomium thermophilum. *Acta Crystallographica Section F: Structural Biology Communications*, 72(2):112–120, 2016.

[66] V. Titouan, N. Courty, R. Tavenard, and R. Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284, 2019.

[67] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *Proceedings Ninth IEEE International Conference on Computer Vision*, 2003.

[68] A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag, 1996.

[69] C. Villani. *Topics in optimal transportation.* American Mathematical Soc., 2003.

[70] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, volume 1, pages I–I. IEEE, 2001.

[71] M. Wendler. U-processes, U-quantile processes and generalized linear statistics of dependent data. *Stochastic Processes and their Applications*, 122(3):787–807, 2012.

[72] H. Xu, D. Luo, R. Henao, S. Shah, and L. Carin. Learning autoencoders with relational regularization. In *International Conference on Machine Learning*, pages 10576–10586. PMLR, 2020.

Distribution of Distances based Object Matching: Supplementary Material

Christoph Alexander Weitkamp* Katharina Proksch† Carla Tameling*
Axel Munk *‡

Contents

I Supplement A: Additional Details **2**

B Proofs of Section 2 and Section 5.1 **3**
B.1 Missing Details from Example 2.2 3
B.2 Proof of Lemma 2.3 6
B.3 Proof of Theorem 2.5 10
B.4 Proof of Theorem 2.6 17
B.4.1 First Part 17
B.4.2 Second Part 21
B.5 Proof of Theorem 2.7 32
B.5.1 First Part 32
B.5.2 Second Part 37
B.6 Proof of Lemma B.12 43
B.7 The DTM-Signature 52
B.7.1 Connections to other Lower Bounds 53

C Distributions of Euclidean Distances on k-Ahlfors Regular Sets **57**
C.1 Preliminary Results 57
C.2 Main Results 65

*Institute for Mathematical Stochastics, University of Göttingen, University of Göttingen
†Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente
‡Max Planck Institute for Biophysical Chemistry, Göttingen

D Bootstrap Approximation **74**

E Detailed Simulation Results **80**
E.1 Details of Section 4.2 80
E.1.1 Empirical Power 80
E.1.2 Distribution of Independent Distances 81
E.1.3 Influence of Beta 81
E.2 Details of Section 5.1 82
E.2.1 Torus-Pentagon Comparison 82
E.2.2 Spiral Comparison 83
E.2.3 Protein Structure Comparison 84

II Supplement B: Auxiliary Results **85**

F U- and U-Quantile Processes **86**
F.1 Distributional Limits for the Empirical U-Process 87
F.2 Distributional Limits for the Empirical U-Quantile Processes 90

G Introduction of the Bootstrap **93**
G.1 The General Bootstrap 93
G.2 The Bootstrap for the Empirical U-Process 95

H Technical Results **103**
H.1 Integrated Difference of Quantile Functions 103
H.2 Differentiation and Integration 105
H.3 Measurability of Function Classes 105
H.4 Further Measurability Issues 108

Part I

Supplement A: Additional Details

In the first part of the supplementary material we reconsider Example 2.2, give the full proofs of the theorems of Section 2 and investigate the DTM-signature introduced in Section 5.1 more closely. Furthermore, we study more general distributions of Euclidean distances, validate the bootstrap scheme suggested in Section 3 for $\beta > 0$ and present our simulation results in more detail.
Throughout this part of the supplementary material $||\cdot||$ denotes the Euclidean norm

Tab. 1: Distribution of distances: A summary of the results of Section B.1 where we check which of the metric measure spaces $\mathcal{X}_i$, $2 \leq i \leq 7$, defined in Example 2.2 meet the requirements of Condition 1.2 for all $\beta > 0$ and which meet the requirements of Condition 1.3.

	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_5$	$\mathcal{X}_6$	$\mathcal{X}_7$
Condition 1.2	✓	✓	✓	✓	✓	✗
Condition 1.3	✓	✓	✗	✓	✓	✗

on $\mathbb{R}^d$ and $\mathscr{B}(\mathbb{R})$ denotes the Borel sets on $\mathbb{R}$. Further, "$\Rightarrow$" denotes the classical weak convergence (see Billingsley [6]), "$\rightsquigarrow$" stands for weak convergence in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner [40, Part 1]) and "$\underset{M}{\overset{P}{\rightsquigarrow}}$" denotes weak convergence for the bootstrap as introduced in Kosorok [23, Sec. 2.2.3] . Moreover, we assume that the random variables $X_1, \dots, X_n, Y_1, \dots, Y_m$ defined in Setting 1.1 live on a common probability space $(\Omega, \mathcal{A})$. Let $T \subset \mathbb{R}^d$ be an arbitrary set. Then, the space $\ell^\infty(T)$ denotes the usual space of all uniformly bounded, $\mathbb{R}$-valued functions on T and $\ell^p(T)$ the one of all p-integrable, $\mathbb{R}$-valued functions on T. Further, $C_b(T)$ stands for the space of all continuous, bounded, $\mathbb{R}$-valued functions on T. Let $(\mathcal{X}, \mathcal{M}, \mu)$ denote a measure space. Similarly to the previous definitions, we denote by $\ell^p(\mu)$ the space of all real valued functions on $\mathcal{X}$ that are p-integrable with respect to μ.

B Proofs of Section 2 and Section 5.1

In the following section, we provide additional details for Example 2.2 and prove the statements from Section 2 as well as Section 5.1.

B.1 Missing Details from Example 2.2

We reconsider the metric measure spaces $\mathcal{X}_i = (\mathcal{X}_i, d_{\mathcal{X}_i}, \mu_{\mathcal{X}_i})$, $2 \leq i \leq 7$, defined in Example 2.2 and verify that their respective distributions of distances (do not) meet the requirements of Condition 1.2 for all $\beta > 0$ and Condition 1.3. The results of our considerations are concisely summarized in Table 1. Throughout this section, we make use of the notation introduced in Section 2.1 of the paper. We briefly recall that for $X^i, \widetilde{X}^i \overset{i.i.d.}{\sim} \mu_{\mathcal{X}_i}$, we define $U_i(t) = \mathbb{P}\left(d_{\mathcal{X}_i}(X^i, \widetilde{X}^i) \leq t\right)$ and $u_i(t) = U_i'(t)$, $2 \leq i \leq 7$. Further, recall that $k_{\mathcal{X}_i}$ denotes the isotropized set covariance function of the set $\mathcal{X}_i$, $i \in \{2, 3\}$ (see e.g. (B.3)) for a formal definition).

The metric measure space $\mathcal{X}_2$: First of all, we consider the Euclidean metric measure space $(\mathcal{X}_2, d_{\mathcal{X}_2}, \mu_{\mathcal{X}_2})$ based on the disc with diameter one. It was shown (see e.g. Moltchanov

[30]) that u_2, is given as

$$u_2(t) = \begin{cases} 8t\left(\frac{2}{\pi}\arccos(t) - \frac{2t}{\pi}\sqrt{1-t^2}\right), & \text{if } 0 \le t \le 1 \\ 0, & \text{else.} \end{cases}$$

It is easy to verify that u_2 is continuous and strictly positive on the interior of its support (i.e. the requirements of Condition 1.2 are met for all $\beta > 0$). Furthermore, we see that u_2 is monotonically decreasing on $[0.9, 1]$ and that for $t \in [0.9, 1]$ we have (using Lemma B.2) that

$$k_{\mathcal{X}_2}(t) = \frac{u_2(t)}{t} \ge \frac{16}{\pi}(2-t)^2.$$

In consequence, it follows by Lemma 2.3 that $\mathcal{X}_2$ also meets the requirements of Condition 1.3.

The metric measure space $\mathcal{X}_3$*:* Next, we investigate the distribution of distances of the Euclidean metric measure space $(\mathcal{X}_3, d_{\mathcal{X}_3}, \mu_{\mathcal{X}_3})$, where $\mathcal{X}_3$ denotes a square with diameter one. It has been shown in Philip [34] that the density u_3 is given as

$$u_3(t) = \begin{cases} 4t\left(\pi + 2t^2 - 4\sqrt{2}t\right), & \text{if } 0 \le t < 1/\sqrt{2} \\ 4t\left(4\arctan\left(\frac{1}{\sqrt{2t^2-1}}\right) - 2t^2 + 4\sqrt{2t^2-1} - 2 - \pi\right), & \text{if } 1/\sqrt{2} \le t \le 1 \\ 0, & \text{else.} \end{cases}$$

Once again, it is easy to verify that u_3 is continuous and strictly positive on the interior of its support. Hence, $\mathcal{X}_3$ meets the restrictions of Condition 1.2 for all $\beta > 0$. Again, we observe that u_2 is monotonically decreasing on $[0.9, 1]$ and that for $t \in [0.9, 1]$ we have that

$$k_{\mathcal{X}_3}(t) = \frac{u_3(t)}{t} \ge (1-t)^4.$$

Thus, an application of Lemma 2.3 shows that $\mathcal{X}_3$ meets the requirements of Condition 1.3.

The metric measure space $\mathcal{X}_4$*:* Recall that $\mathcal{X}_4$ denotes the Euclidean metric measure space based on the sphere $\mathbb{S}^1 \subset \mathbb{R}^2$. In this case it has been shown (see e.g. [24]) that

$$U_4(t) = \begin{cases} 0, & \text{if } t < 0 \\ \frac{2}{\pi}\arcsin(0.5t), & \text{if } 0 \le t \le 2 \\ 1, & \text{if } t > 2 \end{cases} \quad \text{and} \quad u_4(t) = \begin{cases} \frac{1}{\pi}\sqrt{\frac{1}{1-(0.5t)^2}}, & \text{if } 0 \le t \le 2 \\ 0, & \text{else.} \end{cases}$$

It is straight forward to verify that u_4 is strictly positive, but not continuous for $t = 2$ (see e.g. the visualization of u_4 in Figure 3 of the paper). Thus, it is immediately clear that $\mathcal{X}_4$ does not meet the requirements of Condition 1.3, as u_4 is not continuous on $[0, 2]$ (however it meets the requirements of Condition 1.2 for all $\beta > 0$).

The metric measure space $\mathcal{X}_5$*:* Next, we reconsider the Euclidean metric measure space $(\mathcal{X}_5, d_{\mathcal{X}_5}, \mu_{\mathcal{X}_5})$, where $\mathcal{X}_5 = \mathbb{S}^2 \subset \mathbb{R}^3$. Then, it follows (see e.g. [37]) that

$$U_5(t) = \begin{cases} 0, & \text{if } t < 0 \\ \frac{t^2}{4}, & \text{if } 0 \leq t \leq 2 \\ 1, & \text{if } t > 2 \end{cases} \quad \text{and} \quad u_5(t) = \begin{cases} \frac{t}{2}, & \text{if } 0 \leq t \leq 2 \\ 0, & \text{else.} \end{cases}$$

Clearly, u_5 is continuous on $[0,2]$ and strictly positive on the interior of its support. Hence, $\mathcal{X}_5$ meets the requirements of Condition 1.2 for all $\beta > 0$. In order to verify that the same is true for $\mathcal{X}_5$ and Condition 1.3, we need to verify that there exist $\theta_1, \theta_2 > -1$ such that

$$\left|(U_5^{-1})'(t)\right| \leq c_{U_5} t^{\theta_1}(1-t)^{\theta_2}. \tag{B.1}$$

Fortunately, it is easy to verify (see e.g. Section B.2) that (B.1) is equivalent to

$$(U_5(t))^{-\theta_1} (1 - U_5(t))^{-\theta_2} \leq c_{U_5} \left|u_5(t)\right|. \tag{B.2}$$

Since u_5 is strictly positive on the interior of its support and continuous, it is sufficient to verify that there exists $\epsilon > 0$, $\theta_1, \theta_2 > -1$ and a constant $c_{U_5} > 0$ such that $(U_5(t))^{-\theta_1} \leq c_{U_5} u_5(t)$ for all $t \in [0, \epsilon]$ as well as $(1 - U_5(t))^{-\theta_2} \leq c_{U_5} u_5(t)$ for $t \in [1-\epsilon, 1]$ (this is illustrated in the proof of Lemma B.2, see Section B.2). It is possible to verify that $(U_5(t))^{1/2} \leq u_5(t)$ for $t \in [0, 0.1]$. As we have additionally that $u_5(t) > 1/2$ for $t > 1$, it immediately follows that there exists $\theta_2 > -1$ such that $(1 - U_3(t))^{-\theta_2} \leq u_3(t)$ for all $t \in [1.9, 2]$.

The metric measure space $\mathcal{X}_6$*:* Further, we reconsider the Euclidean metric measure space $\mathcal{X}_6$ based on the sphere $\mathbb{S}^4 \subset \mathbb{R}^5$. It has been shown by Sidiropoulos [37] that

$$U_6(t) = \begin{cases} 0, & \text{if } t < 0 \\ \frac{3t^4}{16} - \frac{3t^6}{96}, & \text{if } 0 \leq t \leq 2 \\ 1, & \text{if } t > 2 \end{cases} \quad \text{and} \quad u_4(t) = \begin{cases} \frac{3}{4}t^3\left(1 - t^2/4\right), & \text{if } 0 \leq t \leq 2 \\ 0, & \text{else.} \end{cases}$$

Once again it is easily checked that the assumptions imposed by Condition 1.2 are met in this setting. With the same argumentation as previously, it is sufficient to verify that there exists $\epsilon > 0$, $\theta_1, \theta_2 > -1$ and a constant $c_{U_6} > 0$ such that $(U_6(t))^{-\theta_1} \leq c_{U_6} u_6(t)$ for all $t \in [0, \epsilon]$ as well as $(1 - U_6(t))^{-\theta_2} \leq c_{U_6} u_6(t)$ for $t \in [1-\epsilon, 1]$. In this particular case, we find that $(U_6(t))^{3/4} \leq 2u_6(t)$ for $t \in [0, 0.1]$ as well as $(1 - U_6(t))^{2/3} \leq 2u_6(t)$ for $t \in [1.9, 2]$.

The metric measure space $\mathcal{X}_7$*:* Finally, we come to the Euclidean metric measure space $(\mathcal{X}_7, d_{\mathcal{X}_7}, \mu_{\mathcal{X}_7})$, where $\mathcal{X}_7 = [0,1]^2 \cup ([0,1] \times [4,5])$. Let $X, X' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}_7}$. Since $\mu_{\mathcal{X}}([0,1]^2) = \mu_{\mathcal{X}}([0,1] \times [4,5]) = 0.5$, it follows that $\mathbb{P}(d_{\mathcal{X}_7}(X, X')) \leq \sqrt{2}) = 0.5$, $\mathbb{P}(\sqrt{2} \leq d_{\mathcal{X}_7}(X, X')) \leq 3) = 0$ and $\mathbb{P}(d_{\mathcal{X}_7}(X, X')) \geq 3) = 0.5$. Hence, there exists no $\beta \in (0, 1/2)$ such that u_7 is strictly positive on $[C_1, C_2] = [U_7^{-1}(\beta) - \epsilon, U_7^{-1}(1-\beta) + \epsilon]$. Consequently, Condition 1.2

cannot be satisfied in this setting. The same arguments show that neither does Condition 1.3. In general, we see that a metric measure space $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ cannot meet the requirements of Condition 1.2 for all $\beta > 0$, if the set $\mathcal{X}$ is disconnected in such a way that the diameters of both connected parts are smaller than the gap in between them. In such a case the cumulative distribution function of the corresponding distribution of distances is not strictly increasing and thus Condition 1.2 cannot hold for all $\beta > 0$.

B.2 Proof of Lemma 2.3

We start by showing that at least the smoothness requirements of Condition 1.2 and Condition 1.3 are usually met when equipping compact sets $\mathcal{X} \subset \mathbb{R}^d$ (with $\lambda_d(\mathcal{X}) > 0$) with the Euclidean metric and the uniform distribution. For the special case $d = 2$ this has been proven in Berrendero et al. [5, Prop. 1]. In the following, we briefly recap their arguments and verify that they are also valid for general $d \geq 2$.

We start by recalling the definition of the covariance function (cf. Section 2.1). Let λ_d denote the Lebesgue measure in $\mathbb{R}^d$. The covariance function of a bounded Borel set $A \subset \mathbb{R}^d$ with $\lambda_d(A) > 0$, $d \geq 1$, is defined by

$$K_A(y) \coloneqq \lambda_d \left(A \cap (A - y) \right),$$

where $y \in \mathbb{R}^d$, $A - y = \{a - y : a \in A\}$. Alternatively, K_A can be expressed in terms of a convolution of two indicator functions

$$K_A = \mathbb{1}_A * \mathbb{1}_{-A},$$

where $-A$ denotes the symmetric set $-A = \{-x : x \in A\}$ [25]. This relates K_A to the density function of $X - X'$, where X, X' are independent random variables uniformly distributed on A [5]. The following lemma summarizes some relevant properties of K_A (cf. Cabo and Baddeley [12, Lemma 1.3] and Galerne [21, Prop. 2]).

Lemma B.1. *Let $A \subset \mathbb{R}^d$ be a bounded Borel set with $\lambda_d(A) > 0$ and covariance function K_A.*

(i) For all $y \in \mathbb{R}^d$, $0 \leq K_A(y) \leq K_A(0) = \lambda_d(A)$. Moreover, $K_A(y) = 0$ whenever $||y|| \geq \operatorname{diam}(A)$, $K_A(y) = K_A(-y)$ for all $y \in \mathbb{R}^d$ and K_A is uniformly continuous on $\mathbb{R}^d$.

(ii) For any integrable $f : [0, \infty) \to \mathbb{R}$,

$$\int_A \int_A f(||x - y||)\, dx\, dy = \int_{\mathbb{R}^d} f(||w||) K_A(w)\, dw.$$

This is the so-called "Borel's overlap formula". Two particularly interesting cases are obtained for $f \equiv 1$ and $f(x) = \mathbb{1}_{[0,t]}(x)/\left(\lambda_d(A)\right)^2$, leading respectively to

$$\int_{\mathbb{R}^d} K_A(w)\, dw = (\lambda_d(A))^2$$

and

$$\mathbb{P}\left(\|X - X'\| \leq t\right) = \frac{1}{(\lambda_d(A))^2} \int_{B(0,t)} K_A(w)\, dw, \text{ for } t \geq 0,$$

where X, X' are independent random variables uniformly distributed on A and $B(0,t)$ denotes the Euclidean ball centered at the origin with radius t.

Next, we recall the definition of the isotropized set covariance function (see Section 2.1). Let $\mathbb{S}_{d-1}$ denote the unit sphere in $\mathbb{R}^d$. Then, $y \in \mathbb{R}^d$ is determined by the polar coordinates (t, v), where $t = \|y\|$ is a real number and $v \in \mathbb{S}_{d-1}$ is the unit vector y/t. We get

$$K_A(y) = K_A(t, v) = \lambda_d\left(A \cap (A - tv)\right).$$

The isotropized set covariance function is defined as follows (recall that $\lambda_d(A) > 0$ by assumption)

$$k_A(t) = \frac{1}{(\lambda_d(A))^2} \int_{\mathbb{S}_{d-1}} K_A(t, v)\, dv. \tag{B.3}$$

Furthermore, let $D = \operatorname{diam}(\mathcal{X})$. With Lemma B.1 at our disposal we can now show the following.

Lemma B.2. *Let $\mathcal{X} \subset \mathbb{R}^d$, $d \geq 2$, be a compact Borel set with $\lambda_d(\mathcal{X}) > 0$, $d_{\mathcal{X}}$ the Euclidean metric and $\mu_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$. Let $X, X' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Then, the distribution function U of $d_{\mathcal{X}}(X, X')$ is for $t \in [0, D]$ given as*

$$U(t) = \int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds.$$

Furthermore, the corresponding density

$$u(t) = t^{d-1} k_{\mathcal{X}}(t)$$

is continuous.

Proof of Lemma B.2. By Lemma B.1, we have that

$$\mathbb{P}\left(\|X - X'\| \leq t\right) = \frac{1}{(\lambda_d(\mathcal{X}))^2} \int_{B(0,t)} K_{\mathcal{X}}(w)\, dw.$$

Next, we rewrite the above expression in spherical coordinates, i.e.,

$$\begin{aligned}
w_1 &= r\cos(\phi_1)\\
w_2 &= r\sin(\phi_1)\cos(\phi_2)\\
&\vdots\\
w_{d-1} &= r\sin(\phi_1)\cdots\sin(\phi_{d-2})\cos(\phi_{d-1})\\
w_d &= r\sin(\phi_1)\cdots\sin(\phi_{d-2})\sin(\phi_{d-1}),
\end{aligned}$$

where r ranges over $[0,t]$, ϕ_{d-1} over $[0,2\pi)$ and $\phi_1,\dots\phi_{d-2}$ over $[0,\pi)$. This gives

$$\mathbb{P}\left(\|X-X'\|\leq t\right) = \frac{1}{(\lambda_d(\mathcal{X}))^2}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{t} K_{\mathcal{X}}\left(\vec{\varphi}_r\right) r^{d-1}\sin^{d-2}(\phi_1)\sin^{d-3}(\phi_2)$$
$$\times\sin^{d-4}(\phi_3)\cdots\sin(\phi_{d-2})\,dr\,d\phi_1\cdots\,d\phi_{d-2}\,d\phi_{d-1},$$

where $\vec{\varphi}_r = (r\cos(\phi_1), r\sin(\phi_1)\cos(\phi_2),\dots, r\sin(\phi_1)\cdots\sin(\phi_{d-1}))$. Obviously, the above expression is bounded by one and thus the theorem of Tonelli/Fubini [6, Thm. 18.3] yields that for $t\in[0,D]$

$$U(t) = \int_0^t r^{d-1}k_{\mathcal{X}}(r)\,dr.$$

This shows that $u = U'$ is given by

$$u(t) = t^{d-1}k_{\mathcal{X}}(t) = \frac{1}{(\lambda_d(\mathcal{X}))^2}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} K_{\mathcal{X}}\left(\vec{\varphi}_t\right) t^{d-1}\sin^{d-2}(\phi_1)\sin^{d-3}(\phi_2)$$
$$\times\sin^{d-4}(\phi_3)\cdots\sin(\phi_{d-2})\,d\phi_1\cdots\,d\phi_{d-2}\,d\phi_{d-1}.$$

The continuity of $K_{\mathcal{X}}$ on $\mathbb{R}^d$ (see Lemma B.1 (a)) induces the continuity of u and the proof is complete. □

Based on the above results we demonstrate Lemma 2.3.

Proof of Lemma 2.3. First Part: By Lemma B.2 we have that in the present framework the density u exists and is continuous. Hence, the smoothness requirements of Condition 1.2 are met. It remains to show that for all $\beta\in(0,1/2)$ there exists some $\epsilon>0$ such that u is positive on $[U^{-1}(\beta)-\epsilon, U^{-1}(1-\beta)+\epsilon]$, i.e., that u is strictly positive on $(0,D)$.

By Lemma B.2 we have that

$$u(t) = \frac{1}{(\lambda_d(\mathcal{X}))^2}\int_{\mathbb{S}_{d-1}} t^{d-1}K_{\mathcal{X}}(t,v)\,dv = t^{d-1}k_{\mathcal{X}}(t).$$

As $t^{d-1} > 0$ for $t > 0$ and $k_{\mathcal{X}}$ is strictly positive on $[0, D)$ by assumption, we conclude that u is strictly positive on $(0, D)$. This yields the claim.

Second Part: By Lemma B.2 the smoothness requirements of Condition 1.3 are met. Let $\theta_1 = -\frac{d-1}{d}$ and $\theta_2 = -\frac{\eta}{\eta+1}$. Clearly, we have that $\theta_1 > -1$ and $\theta_2 > -1$ and hence the claim follows once we have shown that

$$\left|(U^{-1})'(t)\right| \leq c_U t^{\theta_1}(1-t)^{\theta_2} \tag{B.4}$$

for all $t \in (0, 1)$. Here, $c_U > 0$ denotes a finite constant.

Since $(U^{-1})' = 1/\left(u \circ U^{-1}\right)$, (B.4) is equivalent to

$$(U(t))^{-\theta_1}\left(1 - U(t)\right)^{-\theta_2} \leq c_U \left|u(t)\right|$$

for $t \in (0, D)$. This in combination with the representation of U and u given by Lemma B.2 yields that we have to verify that under the assumptions made

$$\left(\int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_1} \left(1 - \int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_2} \leq c_U \left|t^{d-1} k_{\mathcal{X}}(t)\right|. \tag{B.5}$$

for a constant $c_U > 0$ and $t \in (0, D)$. Reconsidering the above expression, we realize that the left hand side is bounded for $t \in [\epsilon, D-\epsilon]$ and the right hand side is bounded away from zero for $t \in [\epsilon, D-\epsilon]$ by assumption, i.e., we have

$$\frac{\sup_{t\in[\epsilon,D-\epsilon]} \left(\int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_1} \left(1 - \int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_2}}{\inf_{t\in[\epsilon,D-\epsilon]} \left|t^{d-1} k_{\mathcal{X}}(t)\right|} < \infty.$$

In consequence, it suffices to verify (B.5) for $t \in (0, \epsilon)$ as well as for $t \in (D-\epsilon, D)$. From now on, $c_{U,1}$ and $c_{U,2}$ denote finite constants that may vary from line to line.

First, we verify (B.5) for $t \in (0, \epsilon)$. As $\theta_1 = -\frac{d-1}{d} \in (-1, 0)$ and $\theta_2 = -\frac{\eta}{\eta+1} \in (-1, 0)$, we have that

$$\begin{aligned}\left(\int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_1} \left(1 - \int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_2} &\leq c_{U,1} \left(\int_0^t s^{d-1} k_{\mathcal{X}}(s)\, ds\right)^{-\theta_1} \\ &\leq c_{U,1} \left(k_{\mathcal{X}}(0)\right)^{-\theta_1} \left(\int_0^t s^{d-1}\, ds\right)^{-\theta_1} = c_{U,1} \left(\frac{t^d}{d}\right)^{-\theta_1}.\end{aligned}$$

Plugging in the definition of $\theta_1 = -\frac{d-1}{d}$ yields that

$$c_{U,1} \left(\frac{t^d}{d}\right)^{-\theta_1} = c_{U,1} t^{d-1} \leq c_{U,1} t^{d-1} k_{\mathcal{X}}(t),$$

where the last inequality holds as $\inf_{t\in(0,\epsilon)} k_{\mathcal{X}}(t) > 0$ by assumption.

Next, we demonstrate (B.5) for $t \in (D-\epsilon, D)$. Since $\theta_1 = -\frac{d-1}{d} \in (-1,0)$ and $\theta_2 = -\frac{\eta}{\eta+1} \in (-1,0)$, it holds

$$\left(\int_0^t s^{d-1}k_{\mathcal{X}}(s)\,ds\right)^{-\theta_1}\left(1-\int_0^t s^{d-1}k_{\mathcal{X}}(s)\,ds\right)^{-\theta_2} \le c_{U,2}\left(1-\int_0^t s^{d-1}k_{\mathcal{X}}(s)\,ds\right)^{-\theta_2}$$
$$= c_{U,2}\left(\int_0^D s^{d-1}k_{\mathcal{X}}(s)\,ds - \int_0^t s^{d-1}k_{\mathcal{X}}(s)\,ds\right)^{-\theta_2} = c_{U,2}\left((s^*)^{d-1}k_{\mathcal{X}}(s^*)(D-t)\right)^{-\theta_2}$$

for some $s^* \in (t,D)$ by the Mean Value Theorem. Since $k_{\mathcal{X}}$ is monotonically decreasing on $(D-\epsilon, D)$ by assumption, we find that

$$c_{U,2}\left((s^*)^{d-1}k_{\mathcal{X}}(s^*)(D-t)\right)^{-\theta_2} \le c_{U,2}\left(D^{d-1}k_{\mathcal{X}}(t)(D-t)\right)^{-\theta_2}.$$

Plugging in that $\theta_2 = -\frac{\eta}{\eta+1}$ and using that $c_{\mathcal{X}}(D-t)^\eta \le k_{\mathcal{X}}(t)$ for $t \in (D-\epsilon, D)$, we obtain that

$$c_{U,2}\left(D^{d-1}k_{\mathcal{X}}(t)(D-t)\right)^{-\theta_2} \le c_{U,2}\left(k_{\mathcal{X}}(t)\right)^{\frac{\eta}{\eta+1}}\left(k_{\mathcal{X}}(t)\right)^{\frac{1}{\eta+1}} \le c_{U,2}t^{d-1}k_{\mathcal{X}}(t),$$

where the last inequality follows as $\sup_{t\in(D-\epsilon,D)}\frac{c_{U,2}}{t^{d-1}} < \infty$. As previously argued, this yields the claim. □

B.3 Proof of Theorem 2.5

In this section, we will prove Theorem 2.5 as well as several auxiliary results required. The key idea of the proof of Theorem 2.5 is to exploit the dependency structure of the samples $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\le i<j\le n}$ and $\{d_{\mathcal{Y}}(Y_k,Y_l)\}_{1\le k<l\le m}$. Let us consider $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\le i<j\le n}$. Since the random variables $X_1,\ldots,X_n$ are independent, it follows that $d_{\mathcal{X}}(X_i,X_j)$ is independent of $d_{\mathcal{X}}(X_{i'},X_{j'})$, whenever i,j,i',j' are pairwise different. This allows us to divide the sample into relatively large groups of independent random variables. This idea is represented in Figure 1. It highlights a possibility to divide the set $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\le i<j\le 6}$ into five sets of three independent distances. Generally, one can prove the following.

Lemma B.3. *Let $n\ge 3$ and let $X_1,\ldots,X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. If n is even, there exists a partition $\{\Pi_k^n\}_{1\le k\le n-1}$ of $\{(i,j)\}_{1\le i<j\le n}$ such that $|\Pi_k^n| = n/2$ for each k and such that the random variables in the set $\{d_{\mathcal{X}}(X_i,X_j)\}_{(i,j)\in\Pi_k^n}$ are independent, $1\le k\le n-1$. If n is odd, there exists a partition $\{\Pi_k^n\}_{1\le k\le n}$ of $\{(i,j)\}_{1\le i<j\le n}$ such that $|\Pi_k^n| = (n-1)/2$ for each k and such that the random variables in the set $\{d_{\mathcal{X}}(X_i,X_j)\}_{(i,j)\in\Pi_k^n}$ are independent, $1\le k\le n$.*

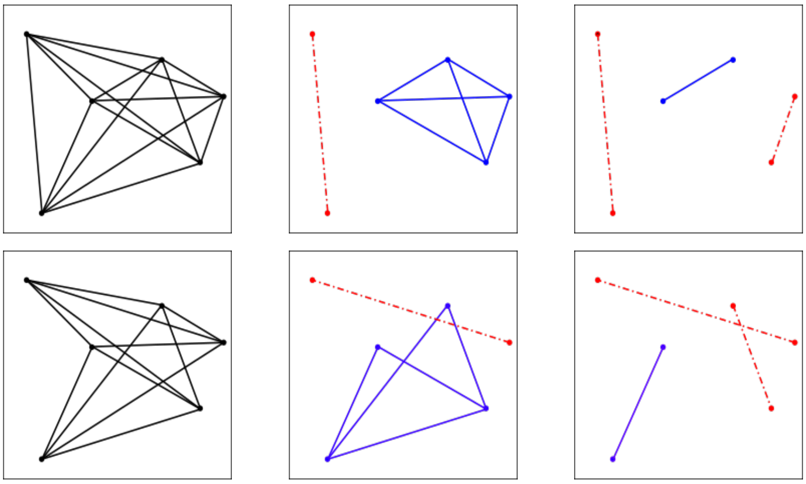

Fig. 1: Partitioning the Distances: Illustration how to partition the set $\{d_{\mathcal{X}}(X_i, X_j)\}_{1\leq i<j\leq 6}$ successively into five set of independent distances of size three. Top row: Left: All pairwise distances between six points. Middle: All distances (blue) that are independent of one chosen distance (red). Right: All distances that are independent of two chosen, independent distances (same color code, right) are shown. Bottom row: The same selection process for the set, where the independent distances displayed in the top right plot were removed.

Proof of Lemma B.3. Let throughout this proof $n \geq 3$. We have already realized that the random variables $g(X_i, X_j)$ and $g(X_k, X_l)$ are independent if and only if $i \neq j \neq k \neq l$. In consequence the problem of finding the independent groups of $\{g(X_i, X_j)\}_{1\leq i<j\leq n}$ is a combinatorial one that is not depending on the considered function g.
In the following, we call $A \subseteq \{(i,j)\}_{1\leq i<j\leq n}$ *independent*, if for all $(i,j),(k,l) \in A$ it holds that $i \neq j \neq k \neq l$. Thus, we obtain the subsequent equivalent reformulation of the statement.
If n is even, there exists a partition $\{\Pi_k^n\}_{1\leq k\leq n-1}$ of $\{(i,j)\}_{1\leq i<j\leq n}$ such that $|\Pi_k^n| = n/2$ and Π_k^n is independent, $1 \leq k \leq n-1$. If n is odd, there exists a partition $\{\Pi_k^n\}_{1\leq k\leq n}$ of $\{(i,j)\}_{1\leq i<j\leq n}$ such that $|\Pi_k^n| = (n-1)/2$ and Π_k^n is independent, $1 \leq k \leq n$.

We will prove this claim by induction. Let $n = 3$. Then, $\{(i,j)\}_{1\leq i<j\leq 3}$ can be represented as

$$\begin{array}{cc} (1,2) & (1,3) \\ & (2,3) \end{array} \tag{B.6}$$

and possible choices for Π_1^3, Π_2^3 and Π_3^3 are

$$\Pi_1^3 = \{(1,2)\}, \ \Pi_2^3 = \{(1,3)\}, \ \Pi_3^3 = \{(2,3)\}. \tag{B.7}$$

Suppose now, there exists an odd $n \geq 3$ such that the statement is true, i.e., we have a partition of $\{(i,j)\}_{1\leq i<j\leq n}$ consisting of n independent sets of size $(n-1)/2$. Since each Π_k^n, $1 \leq k \leq n$, is independent and has size $(n-1)/2$, its tuples contain $(n-1)$ different

numbers. In consequence for each Π_k^n there exist a number $m_k \in \{1,\dots,n\}$ such that m_k is not contained in a tuple in Π_k^n, $1 \le k \le n$. From the symmetry of the considered setting it is clear that

$$\bigcup_{1\le k\le n} m_k = \{1,\dots,n\}\,.$$

In order to obtain a partition of $\{(i,j)\}_{1\le i<j\le n+1}$ with n independent sets of size $(n+1)/2$ ($n+1$ is even) we can just add the element $(m_k, n+1)$ to the set Π_k^n, $1 \le k \le n$. Thus,

$$\Pi_k^{n+1} = \Pi_k^n \cup \{(m_k, n+1)\}\,, \quad 1 \le k \le n.$$

To make this step more illustrative, we present it in the small example $n = 4$, i.e.,

$$\begin{matrix} (1,2) & (1,3) & (1,4) \\ & (2,3) & (2,4) \\ & & (3,4). \end{matrix}$$

We recall that for $n = 3$ we have found the following partition

$$\Pi_1^3 = \{(1,2)\}\,,\ \Pi_2^3 = \{(1,3)\},\ \Pi_3^3 = \{(2,3)\}\,. \tag{B.8}$$

Thus, $m_1 = 3$, $m_2 = 2$ and $m_3 = 1$. In consequence, we obtain the new partition as

$$\Pi_1^4 = \{(1,2),(3,4)\}\,,\ \Pi_2^4 = \{(1,3),(2,4)\},\ \Pi_3^4 = \{(2,3),(1,4)\}\,.$$

Assume now, there exists an even $n \ge 4$ such that the statement is true, i.e., we have a partition $\Pi_1^n,\dots,\Pi_{n-1}^n$ of $\{(i,j)\}_{1\le i<j\le n}$ consisting of $n-1$ independent sets of size $n/2$. At this point, we realize that the map

$$\Phi : \mathbb{N}\times\mathbb{N} \to \mathbb{N}\times\mathbb{N},\ (i,j) \mapsto (i, j-1)$$

maps the set $M_1 = \{(i,j)\}_{i+1<j,1\le i<j\le n+1}$ bijectively onto the set $M_2 = \{(i,j)\}_{1\le i<j\le n}$ in such a way that its dependency structure is maintained. Consequently, we can divide M_1 into $n-1$ independent sets $\Pi_1^{n+1},\dots,\Pi_{n-1}^{n+1}$ of size $n/2$ by the induction hypothesis. It remains to handle the set $\{(1,2),(2,3),(3,4),\dots,(n,n+1)\}$. However, it is easy to see that the sets $\Pi_n^{n+1} = \{(1,2),(3,4),\dots,(n-1,n)\}$ and $\Pi_{n+1}^{n+1} = \{(2,3),(4,5),\dots,(n,n+1)\}$ are independent sets of size $n/2$. In conclusion, we have found under the induction hypothesis that the set $\{(i,j)\}_{1\le i<j\le n+1}$ can be split into $n+1$ independent sets of size $n/2$.
Once again, we demonstrate this step by a short example. Let $n = 5$, then $\{(i,j)\}_{1\le i<j\le 5}$ is given as

$$\begin{matrix} (1,2) & (1,3) & (1,4) & (1,5) \\ & (2,3) & (2,4) & (2,5) \\ & & (3,4) & (3,5) \\ & & & (4,5). \end{matrix}$$

It is easy to see that the set

$$\begin{matrix} (1,3) & (1,4) & (1,5) \\ & (2,4) & (2,5) \\ & & (3,5) \end{matrix}$$

has the same dependence structure as $\{(i,j)\}_{1\le i<j\le 4}$. Thus, we can partition the above set into three sets Π_1^5, Π_2^5 and Π_3^5 of size two. Furthermore, the set $\{(1,2),(2,3),(3,4),(4,5)\}$ can be split into $\Pi_4^5 = \{(1,2),(3,4)\}$ and $\Pi_5^5 = \{(2,3),(4,5)\}$. □

The proof strategy for Theorem 2.5 is now to rewrite the problem at hand as a certain assignment problem and then to restrict the assignments to assignments between groups of independent distances. Before we come to this, we have to introduce some notation and derive another auxiliary result. Let $\mathfrak{S}(B)$ denote the set of all permutations of the finite set B. In the special case $B = \{1,\dots,n\}$, we write $\mathfrak{S}_n$ instead of $\mathfrak{S}(\{1,\dots,n\})$. Let μ and ν denote two Borel probability measures on $\mathbb{R}^d$. Recall that the *Kantorovich (transport) distance of order p* (also known as *Wasserstein distance*, see [42, Def. 6.1]) between μ and ν is defined as

$$\mathcal{K}_p^p(\mu,\nu) := \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^p \, d\pi(x,y),$$

where $\mathcal{M}(\mu,\nu)$ denotes the set of couplings between μ and ν. For the proof of Lemma B.3 it is important to note that the Kantorovich distance between two empirical measures on $\mathbb{R}$ can be bounded as follows.

Lemma B.4. *Given two collections of real numbers $x_1,\dots,x_n$ and $y_1,\dots,y_n$, let μ_n and ν_n denote the corresponding empirical measures. Further, let $\{\mathfrak{P}_k^n\}_{1\le k\le K}$ denote a partition of $\{1,\dots,n\}$, i.e.,*

$$\bigcup_{1\le k\le K} \mathfrak{P}_k^n = \{1,\dots,n\} \quad \text{and} \quad \mathfrak{P}_i^n \cap \mathfrak{P}_j^n = \emptyset \text{ for } i\neq j.$$

Then,

$$\mathcal{K}_p^p(\mu_n,\nu_n) \le \frac{1}{n}\left(\inf_{\sigma\in\mathfrak{S}(\mathfrak{P}_1^n)} \sum_{i\in\mathfrak{P}_1^n} \left|x_i - y_{\sigma(i)}\right|^p + \dots + \inf_{\sigma\in\mathfrak{S}(\mathfrak{P}_K^n)} \sum_{i\in\mathfrak{P}_K^n} \left|x_i - y_{\sigma(i)}\right|^p \right).$$

Further, let $\left\{x_{(i)}^{\mathfrak{P}_k^n}\right\}_{1\le i\le|\mathfrak{P}_k^n|}$ and $\left\{y_{(i)}^{\mathfrak{P}_k^n}\right\}_{1\le i\le|\mathfrak{P}_k^n|}$ denote the ordered samples of $\{x_i\}_{i\in\mathfrak{P}_k^n}$ and $\{y_i\}_{i\in\mathfrak{P}_k^n}$, $1\le k\le K$, respectively. Then it holds

$$\mathcal{K}_p^p(\mu_n,\nu_n) \le \frac{1}{n}\left(\sum_{1\le i\le|\mathfrak{P}_1^n|} \left|x_{(i)}^{\mathfrak{P}_1^n} - y_{(i)}^{\mathfrak{P}_1^n}\right|^p + \dots + \sum_{1\le i\le|\mathfrak{P}_K^n|} \left|x_{(i)}^{\mathfrak{P}_K^n} - y_{(i)}^{\mathfrak{P}_K^n}\right|^p \right).$$

Proof of Lemma B.4. In the proof of Lemma 4.2 of Bobkov and Ledoux [8] it is shown that

$$\mathcal{K}_p^p(\mu_n,\nu_n) = \frac{1}{n}\inf_{\sigma\in\mathfrak{S}_n}\sum_{1\le i\le n}\left|x_i - y_{\sigma(i)}\right|^p.$$

Since we have $\mathfrak{P}_i^n \cap \mathfrak{P}_j^n = \emptyset$ for $i \neq j$, it clearly follows that

$$\frac{1}{n}\inf_{\sigma\in\mathfrak{S}_n}\sum_{1\le i\le n}\left|x_i - y_{\sigma(i)}\right|^p = \frac{1}{n}\inf_{\sigma\in\mathfrak{S}_n}\left(\sum_{i\in\mathfrak{P}_1^n}\left|x_i - y_{\sigma(i)}\right|^p + \dots + \sum_{i\in\mathfrak{P}_K^n}\left|x_i - y_{\sigma(i)}\right|^p\right)$$
$$\le \frac{1}{n}\left(\inf_{\sigma\in\mathfrak{S}(\mathfrak{P}_1^n)}\sum_{i\in\mathfrak{P}_1^n}\left|x_i - y_{\sigma(i)}\right|^p + \dots + \inf_{\sigma\in\mathfrak{S}(\mathfrak{P}_K^n)}\sum_{i\in\mathfrak{P}_K^n}\left|x_i - y_{\sigma(i)}\right|^p\right),$$

which yields the first part of the claim.

The second part follows by an application of Lemma 4.1 of Bobkov and Ledoux [8] with $V(x) = |x|^p$, which yields for $1 \le k \le K$ that

$$\inf_{\sigma\in\mathfrak{S}(\mathfrak{P}_k^n)}\sum_{i\in\mathfrak{P}_k^n}\left|x_i - y_{\sigma(i)}\right|^p = \sum_{1\le i\le|\mathfrak{P}_k^n|}\left|x_{(i)}^{\mathfrak{P}_k^n} - y_{(i)}^{\mathfrak{P}_k^n}\right|^p.$$

□

Let $(S,\mathcal{S})$ be a measurable space, let $x_1,\dots,x_n,y_1,\dots,y_n \in S$ and let $g,h : S\times S \mapsto \mathbb{R}$ be symmetric and measurable functions. Let for any $A \in \mathscr{B}(\mathbb{R})$

$$\mu_n^g(A) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\mathbb{1}_{\{g(x_i,x_j)\in A\}}$$

and

$$\nu_n^h(A) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\mathbb{1}_{\{h(y_i,y_j)\in A\}}.$$

As it is notationally convenient, we define $g_{(i,j)} = g(x_i,x_j)$ and analogously $h_{(i,j)} = h(y_i,y_j)$, $1 \le i < j \le n$. Clearly, Lemma B.4 also holds in the setting just described. The corresponding reformulation of said lemma is stated next.

Corollary B.5. *Let $\Pi_1^n,\dots,\Pi_K^n$ be a partition of $\{(i,j)\}_{1\le i<j\le n}$. Then*

$$\mathcal{K}_p^p\left(\mu_n^g,\nu_n^h\right) \le \frac{2}{n(n-1)}\left(\sum_{k=1}^K \inf_{\sigma\in\mathfrak{S}(\Pi_k^n)}\sum_{(i,j)\in\Pi_k^n}\left|g_{(i,j)} - h_{\sigma((i,j))}\right|^p\right).$$

Further, let $\left\{g_{(l)}^{\Pi_k^n}\right\}_{1\leq l\leq |\Pi_k^n|}$ *and* $\left\{h_{(l)}^{\Pi_k^n}\right\}_{1\leq l\leq |\Pi_k^n|}$ *denote the ordered samples of* $\left\{g_{(i,j)}\right\}_{(i,j)\in\Pi_k^n}$ *and* $\left\{h_{(i,j)}\right\}_{(i,j)\in\Pi_k^n}$, $1\leq k\leq K$, *respectively. Then, it holds*

$$\mathcal{K}_p^p\left(\mu_n^g,\nu_n^h\right)\leq\frac{2}{n(n-1)}\left(\sum_{k=1}^K\sum_{1\leq l\leq|\Pi_k^n|}\left|g_{(l)}^{\Pi_k^n}-h_{(l)}^{\Pi_1^n}\right|^p\right).$$

Now that we have derived all auxiliary results required, we come to the proof of Theorem 2.5.

Proof of Theorem 2.5. We observe that for $\beta\in[0,1/2)$

$$\mathbb{E}\left[\widehat{DoD}_{(\beta)}\right]\leq\mathbb{E}\left[\int_0^1|U_n^{-1}(t)-V_m^{-1}(t)|^2\,dt\right]=\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\mu_m^V\right)\right],$$

where μ_n^U and μ_m^V are the empirical measures corresponding to U_n and V_m, i.e., for $A\in\mathscr{B}(\mathbb{R})$

$$\mu_n^U(A)=\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\in A\}}$$

and

$$\mu_m^V(A)=\frac{2}{m(m-1)}\sum_{1\leq k<l\leq m}\mathbb{1}_{\{d_{\mathcal{Y}}(Y_k,Y_l)\in A\}}.$$

Since it holds $\mu^U=\mu^V$ by assumption, we obtain by the triangle inequality

$$\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\mu_m^V\right)\right]\leq 2\left(\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\mu^U\right)\right]+\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_m^V,\mu^V\right)\right]\right).$$

Thus, it remains to show

$$\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\mu^U\right)\right]\leq\frac{4}{n+1}J_2\left(\mu^U\right). \tag{B.9}$$

In order to demonstrate (B.9), we will use Lemma B.3. Hence, it is notationally convenient to distinguish the cases n even and n odd, although the proof is essentially the same in both settings. In the following, we will therefore restrict ourselves to n odd.

The first step to prove (B.9) is to realize (cf. Bobkov and Ledoux [8, Sec. 4]), that

$$\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\mu^U\right)\right]\leq\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\nu_n^U\right)\right], \tag{B.10}$$

where ν_n^U denotes an independent copy of μ_n^U, i.e.,

$$\nu_n^U(A)=\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\mathbb{1}_{\{d_{\mathcal{X}}(X_i',X_j')\in A\}},$$

for $X_1',\dots,X_n' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. By Lemma B.3, there exists a partition $\Pi_1^n,\dots,\Pi_n^n$ of the set $\{(i,j)\}_{1\leq i<j\leq n}$ with $|\Pi_k^n| = (n-1)/2$, $1\leq k\leq n$, such that the random variables in the sets $\{d_{\mathcal{X}}(X_i,X_j)\}_{(i,j)\in\Pi_k^n}$ and the ones in the sets $\left\{d_{\mathcal{X}}(X_i',X_j')\right\}_{(i,j)\in\Pi_k^n}$ are independent. Let $\left\{d_{(i)}^{\Pi_k^n,X}\right\}_{1\leq i\leq (n-1)/2}$ stand for the ordered sample of $\{d_{\mathcal{X}}(X_i,X_j)\}_{(i,j)\in\Pi_k^n}$ and the let $\left\{d_{(i)}^{\Pi_k^n,X'}\right\}_{1\leq i\leq (n-1)/2}$ stand for the one of $\left\{d_{\mathcal{X}}\left(X_i',X_j'\right)\right\}_{(i,j)\in\Pi_k^n}$, $1\leq k\leq n$. The application of Corollary B.5 with this partition yields that

$$\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U,\nu_n^U\right)\right] \leq \frac{2}{n(n-1)}\mathbb{E}\left[\sum_{k=1}^{n}\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_k^n,X}-d_{(i)}^{\Pi_k^n,X'}\right|^2\right].$$

Furthermore, we realize that, as $X_1,\dots,X_n,X_1',\dots,X_n'$ are independent, identically distributed, it holds for $1\leq k,l\leq n$ that

$$\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_k^n,X}-d_{(i)}^{\Pi_k^n,X'}\right|^2 \overset{D}{=} \sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_l^n,X}-d_{(i)}^{\Pi_l^n,X'}\right|^2.$$

Consequently, we have

$$\mathbb{E}\left[\sum_{k=1}^{n}\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_k^n,X}-d_{(i)}^{\Pi_k^n,X'}\right|^2\right] = n\mathbb{E}\left[\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_1^n,X}-d_{(i)}^{\Pi_1^n,X'}\right|^2\right].$$

We come to the final step of this proof. Let for any $A\in\mathscr{B}(\mathbb{R})$

$$\mu_n^*(A) = \frac{2}{n-1}\sum_{i=1}^{(n-1)/2}\mathbb{1}_{\left\{d_{(i)}^{\Pi_1^n,X}\in A\right\}}$$

and let $\nu_n^*(A)$ be defined analogously. Then, Theorem 4.3 of Bobkov and Ledoux [8] implies that

$$\mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^*,\nu_n^*\right)\right] = \frac{2}{n-1}\mathbb{E}\left[\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_1^n,X}-d_{(i)}^{\Pi_1^n,X'}\right|^2\right].$$

By construction, the samples $\left\{d_i^{\Pi_1^n,X}\right\}_{1\leq i\leq (n-1)/2}$ and $\left\{d_i^{\Pi_1^n,X'}\right\}_{1\leq i\leq (n-1)/2}$ consist of independent random variables and are independent of each other. Furthermore, we have $\mathbb{E}\left[\mu_n^*\right]=\mathbb{E}\left[\nu_n^*\right]=\mu^U$. Since $J_2\left(\mu^U\right)<\infty$ by assumption, it follows by Theorem 5.1 of Bobkov and Ledoux [8] that

$$\frac{2}{n-1}\mathbb{E}\left[\sum_{i=1}^{(n-1)/2}\left|d_{(i)}^{\Pi_1^n,X}-d_{(i)}^{\Pi_1^n,X'}\right|^2\right] = \mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^*,\nu_n^*\right)\right] \leq \frac{4}{n+1}J_2\left(\mu^U\right).$$

This yields (B.9) and thus concludes the proof. □ □

Remark B.6. To conclude this section, we point out how Theorem 2.5 can be generalized to obtain a finite sample bound for

$$\mathbb{E}\left[\widehat{DoD}_{p,(\beta)}\left(\mathcal{X}_n, \mathcal{Y}_m\right)\right] = \mathbb{E}\left[\int_{\beta}^{1-\beta} \left|U_n^{-1}(t) - V_m^{-1}(t)\right|^p \, dt\right], \tag{B.11}$$

where $p \in [1, \infty)$ and $\beta \in [0, 1/2)$. To this end, let

$$J_p\left(\mu^U\right) := \int_{-\infty}^{\infty} \frac{[U(t)(1-U(t))]^{p/2}}{(u(t))^{p-1}} \, dt < \infty.$$

Then, applying Theorem 5.3 of Bobkov and Ledoux [8] instead of Theorem 5.1 in the proof above yields the analogous finite sample bound for the expectation defined in (B.11).

B.4 Proof of Theorem 2.6

In the first part of this subsection, we focus on the proof of Theorem 2.6 (i) and complete the one of Theorem 2.6 (ii) afterwards.

B.4.1 First Part

An essential tool for the verification of Theorem 2.6 (i) are the distributional limits of the empirical U-quantile processes $\mathbb{U}_n^{-1} := \sqrt{n}\left(U_n^{-1} - U^{-1}\right)$ and $\mathbb{V}_m^{-1} := \sqrt{m}\left(V_n^{-1} - V^{-1}\right)$, which we derive next. To this end, we ensure that Theorem F.10 on the weak convergence of the empirical U-quantile process is applicable.

Lemma B.7. *Assume Setting 1.1 and let Condition 1.2 be fulfilled. Then, as $n, m \to \infty$ it holds that*

$$\mathbb{U}_n^{-1} = \sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G}_1 \quad \textit{and} \quad \mathbb{V}_m^{-1} = \sqrt{m}\left(V_n^{-1} - V^{-1}\right) \rightsquigarrow \mathbb{G}_2$$

in $\ell^\infty[\beta, 1-\beta]$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ denote centered Gaussian processes with covariances

$$\operatorname{Cov}\left(\mathbb{G}_1(t), \mathbb{G}_1(t')\right) = \frac{4}{(u \circ U^{-1}(t))(u \circ U^{-1}(t'))} \Gamma_{d_\mathcal{X}}\left(U^{-1}(t), U^{-1}(t')\right)$$

and

$$\operatorname{Cov}\left(\mathbb{G}_2(t), \mathbb{G}_2(t')\right) = \frac{4}{(v \circ V^{-1}(t))(v \circ V^{-1}(t'))} \Gamma_{d_\mathcal{Y}}\left(V^{-1}(t), V^{-1}(t')\right).$$

Here,

$$\Gamma_{d_\mathcal{X}}(t,t') := \int\int \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t\}}\, d\mu_\mathcal{X}(y) \int \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t'\}}\, d\mu_\mathcal{X}(y)\, d\mu_\mathcal{X}(x)$$
$$- \int\int \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t\}}\, d\mu_\mathcal{X}(y)\, d\mu_\mathcal{X}(x) \int\int \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t'\}}\, d\mu_\mathcal{X}(y)\, d\mu_\mathcal{X}(x)$$

and $\Gamma_{d_\mathcal{Y}}$ *is defined analogously.*

Proof of Lemma B.7. We demonstrate the claim by applying Theorem F.10 to $\mathbb{U}_n^{-1}$ and $\mathbb{V}_m^{-1}$, respectively. Therefore, we have to ensure that both processes fulfill the requirements of Theorem F.10. In the following, we concentrate on $\mathbb{U}_n^{-1}$ and remark that $\mathbb{V}_m^{-1}$ can be handled analogously.

First, we observe that U_n and U are (empirical) U-distribution functions (see Definition F.1) with kernel function

$$h(x,y,t) := \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t\}}.$$

Clearly, $h(x,y,t)$ has the correct form for applying Theorem F.10 and we can show that

$$\mathscr{F} = \left\{(x,y) \mapsto \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t\}} : t \in [C_1,C_2]\right\},$$

is a permissible class of functions. To this end, we discern that $\mathscr{F} = \left\{\tilde{f}(\cdot,t) : t\in T\right\}$ with $T = [C_1,C_2]$ and

$$\tilde{f}(x,y,t) = \mathbb{1}_{\{d_\mathcal{X}(x,y)\le t\}}.$$

In order to check that $\tilde{f} : \mathcal{X}\times\mathcal{X}\times[C_1,C_2] \mapsto \mathbb{R}$ is $\mathscr{B}(\mathcal{X}\times\mathcal{X})\otimes\mathscr{B}([C_1,C_2])$-measurable we utilize Theorem H.7, i.e., we verify that $\tilde{f}(x,y,\cdot)$ is right continuous for all $(x,y)\in\mathcal{X}\times\mathcal{X}$ (which is obvious) and that $\tilde{f}(\cdot,\cdot,t)$ is $\mathscr{B}(\mathcal{X}\times\mathcal{X})$-measurable for all $t\in[C_1,C_2]$. Let in the following $\tilde{f}_t = \tilde{f}(\cdot,\cdot,t)$, $t\in[C_1,C_2]$. Then, for any $t\in[C_1,C_2]$ we know that $\tilde{f}_t$ is $\mathscr{B}(\mathcal{X}\times\mathcal{X})$-measurable if and only if $\{d_\mathcal{X}(x,y)\le t\}\in\mathscr{B}(\mathcal{X}\times\mathcal{X})$ [6, Sec. 13]. By definition we have that the distance $d_\mathcal{X}$ is a continuous map from $\mathcal{X}\times\mathcal{X}$ to $\mathbb{R}$, i.e., in particular $\mathscr{B}(\mathcal{X}\times\mathcal{X})$-measurable. Clearly, it holds $[0,t]\in\mathscr{B}(\mathbb{R})$ for any $t\in[C_1,C_2]$, which implies that

$$\{d_\mathcal{X}(x,y)\le t\} = d_\mathcal{X}^{-1}([0,t]) \in \mathscr{B}(\mathcal{X}\times\mathcal{X})$$

for $t\in[C_1,C_2]$. Thus, we conclude that $\tilde{f}_t$ is $\mathscr{B}(\mathcal{X}\times\mathcal{X})$-measurable for $t\in[C_1,C_2]$. Furthermore, $T=[C_1,C_2]$ is a compact, separable metric space and hence $\mathscr{F}$ is a permissible class of functions.

Next, we show that $\mathscr{F}$ is a VC-subgraph class, i.e., we have to confirm that the following class of sets

$$\mathscr{V} = \{\text{subgraph}(f_t) : f_t \in \mathscr{F}\},$$

where

$$\text{subgraph}(f_t) := \{(x, y, s) \in \mathcal{X} \times \mathcal{X} \times \mathbb{R} : 0 < s < f_t(x, y) \text{ or } 0 > s > f_t(x, y)\},$$

is a VC-class (cf. Definition F.5). Therefore, we show, that $\mathscr{V}$ does not shatter any set of two points $\{V, W\}$ with $V, W \in \mathcal{X} \times \mathcal{X} \times \mathbb{R}$. Let $V = (v_1, v_2, r_v)$ and $W = (w_1, w_2, r_w)$ with $v_1, v_2, w_1, w_2 \in \mathcal{X}$ and $r_v, r_w \in \mathbb{R}$. Since all $f_t \in \mathscr{F}$ are indicator functions, the values r_v and r_w have no influence on the intersection of subgraph(f_t) and W or of subgraph(f_t) and V, respectively, as long as they are in $(0, 1)$. Therefore we can without loss of generality assume that $r_v = r_w = 0.5$, i.e., the problem's dimension can be reduced by one. Hence, we denote by $V' = (v_1, v_2)$ the natural projection from V onto $\mathcal{X} \times \mathcal{X}$ and analogously by $W' = (w_1, w_2)$ the projection of W onto $\mathcal{X} \times \mathcal{X}$. Furthermore, we define

$$A = \{(v, w) \in \mathcal{X} \times \mathcal{X} : d_{\mathcal{X}}(v, w) = 0\} = \{(v, w) \in \mathcal{X} \times \mathcal{X} : v = w\}.$$

It is notable that A is a closed and convex subset of $\mathcal{X} \times \mathcal{X}$. Thus, we can define the distance between any V' (respectively W') and A uniquely, as follows

$$d_{\mathcal{X}}^{A}(V') = \inf_{a \in A} d_{\mathcal{X}}\left(V', a\right).$$

Let $f_t(x, y) = \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \leq t\}} \in \mathscr{F}$ and $V = (v_1, v_2, r_v)$. Then, it holds by construction that

$$\text{subgraph}(f_t) \cap \{V\} \neq \emptyset \quad \text{and} \quad d_{\mathcal{X}}^{A}(V') > t \tag{B.12}$$

are equivalent. This means that we have to consider 3 cases in order to verify that $\mathscr{V}$ is a VC-class:

1. $d_{\mathcal{X}}^{A}(V') = d_{\mathcal{X}}^{A}(W')$:
   (a) $d_{\mathcal{X}}^{A}(V') = d_{\mathcal{X}}^{A}(W') = 0$:
   Thanks to (B.12) we know, that in this case
   $$\text{subgraph}(f_t) \cap \{V, W\} = \{V, W\},$$
   for all $f_t \in \mathscr{F}$ (w.o.l.g. $r_v = r_w = 0.5$). Thus, $\mathscr{V}$ cannot shatter sets of this form.
   (b) $d_{\mathcal{X}}^{A}(V') = d_{\mathcal{X}}^{A}(W') \neq 0$:
   In this case we can only pick out the sets $\emptyset$ and $\{V, W\}$. This is clear, as for f_t with $t < d_{\mathcal{X}}^{A}(V')$, we have
   $$\text{subgraph}(f_t) \cap \{V, W\} = \emptyset,$$
   and f_t with for $t \geq d_{\mathcal{X}}^{A}(V')$
   $$\text{subgraph}(f_t) \cap \{V, W\} = \{V, W\}.$$

2. $d^A_{\mathcal{X}}(V') \neq d^A_{\mathcal{X}}(W')$:
   Suppose without loss of generality that $d^A_{\mathcal{X}}(V') < d^A_{\mathcal{X}}(W')$. Then, $\mathscr{V}$ cannot pick out the subset $\{W\}$, because for that to happen, we must choose f_t such that

   $$\text{subgraph}(f_t) \cap \{W\} = \{W\} \text{ and } \text{subgraph}(f_t) \cap \{V\} = \emptyset,$$

   which is equivalent to finding a t such that

   $$d^A_{\mathcal{X}}(W') \leq t \quad \text{and} \quad d^A_{\mathcal{X}}(V') > t.$$

   This is not possible, as by assumption $d^A_{\mathcal{X}}(V') < d^A_{\mathcal{X}}(W')$.

The above considerations prove that there does not exists a set of size two that is shattered by $\mathscr{V}$. Hence, $\mathscr{V}$ is a VC-class.

Finally, we realize that U is differentiable with strictly positive density u on $[C_1, C_2]$ by assumption. Thus, an application of Theorem F.10 yields that $\mathbb{U}_n^{-1} \rightsquigarrow \mathbb{G}_1$ in $\ell^\infty[\beta, 1-\beta]$.

□

With Lemma B.7 available the proof of Theorem 2.6 (i) is straightforward.

Proof of Theorem 2.6 (i). We start the proof by recalling that under the assumptions made we have $V = U$, i.e., $U^{-1} = V^{-1}$. It follows that

$$\begin{aligned}
&\frac{nm}{n+m}\int_\beta^{1-\beta} \left(U_n^{-1}(t) - V_m^{-1}(t)\right)^2\, dt \\
=&\int_\beta^{1-\beta}\left(\sqrt{\frac{nm}{n+m}}\left(U_n^{-1}(t) - U^{-1}(t)\right) - \sqrt{\frac{nm}{n+m}}\left(V_m^{-1}(t) - V^{-1}(t)\right)\right)^2 dt \\
=&\varphi\left(\sqrt{\frac{nm}{n+m}}\left(U_n^{-1}(t) - U^{-1}(t)\right), \sqrt{\frac{nm}{n+m}}\left(V_m^{-1}(t) - V^{-1}(t)\right)\right),
\end{aligned}$$

where $\varphi : \ell^\infty[\beta, 1-\beta] \times \ell^\infty[\beta, 1-\beta] \to \mathbb{R}$ is defined as

$$\varphi(f, g) = \int_\beta^{1-\beta} (f(x) - g(x))^2\, dx.$$

It is easily verified that φ is a continuous map.

Next, we realize that requirements for applying Lemma B.7 are assumed. Consequently, we have that $\sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G}_1$ and $\sqrt{m}\left(V_n^{-1} - V^{-1}\right) \rightsquigarrow \mathbb{G}_2$ in $\ell^\infty[\beta, 1-\beta]$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are centered Gaussian processes with covariances as defined in Lemma B.7.

As $U = V$, it follows that $\mathbb{G}_1 \overset{\mathcal{D}}{=} \mathbb{G}_2 \overset{\mathcal{D}}{=} \mathbb{G}$. Since the processes $\sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - U^{-1}\right)$ and $\sqrt{\frac{nm}{n+m}}\left(V_m^{-1} - V^{-1}\right)$ are independent and $n/(n+m) \to \lambda$, it follows that

$$\left(\sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - U^{-1}\right), \sqrt{\frac{nm}{n+m}}\left(V_m^{-1} - V^{-1}\right)\right) \rightsquigarrow \left(\sqrt{1-\lambda}\mathbb{G}_1, \sqrt{\lambda}\mathbb{G}_2\right)$$

in $\ell^\infty[\beta, 1-\beta] \times \ell^\infty[\beta, 1-\beta]$. By applying the continuous mapping theorem [40, Thm. 1.3.6], we obtain

$$\varphi\left(\sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - U^{-1}\right), \sqrt{\frac{nm}{n+m}}\left(V_m^{-1} - V^{-1}\right)\right) \rightsquigarrow \varphi\left(\sqrt{1-\lambda}\mathbb{G}_1, \sqrt{\lambda}\mathbb{G}_2\right),$$

which equals by definition

$$\int_\beta^{1-\beta}\left(\sqrt{\frac{nm}{n+m}}\left(U_n^{-1}(t) - U^{-1}(t)\right) - \sqrt{\frac{nm}{n+m}}\left(V_m^{-1}(t) - V^{-1}(t)\right)\right)^2 dt$$
$$\rightsquigarrow \int_\beta^{1-\beta}\left(\sqrt{1-\lambda}\mathbb{G}_1(t) - \sqrt{\lambda}\mathbb{G}_2(t)\right)^2 dt.$$

Finally, as $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent and identically distributed, it holds that

$$\mathbb{G} \overset{\mathcal{D}}{=} \sqrt{1-\lambda}\mathbb{G}_1 - \sqrt{\lambda}\mathbb{G}_2,$$

which gives the claim. □

Remark B.8. Since also the maps

$$\varphi_p : \ell^\infty[\beta, 1-\beta] \times \ell^\infty[\beta, 1-\beta] \to \mathbb{R}, \ (f, g) \mapsto \int_\beta^{1-\beta} |f(x) - g(x)|^p \, dx$$

are continuous for $p \in [1, \infty)$, we can employ the same strategy of proof to derive that given Condition 1.2

$$\left(\frac{nm}{n+m}\right)^{p/2} \widehat{DoD}_{p,(\beta)} := \left(\frac{nm}{n+m}\right)^{p/2} \int_\beta^{1-\beta} \left|U_n^{-1}(t) - V_m^{-1}(t)\right|^p dt \rightsquigarrow \Xi_p := \int_\beta^{1-\beta} |\mathbb{G}(t)|^p \, dt$$

under the hypothesis H_0.

B.4.2 Second Part

Next, we come to the proof of Theorem 2.6 (ii). For notational convenience we restrict ourselves from now on to the case $n = m$. However, the same strategy of proof also gives the general case $n \neq m$ (for some additional details on this issue see Remark B.15).

First, we demonstrate that $\left\{\Xi_n^{U,V}(\beta)\,|\,\beta\in[0,1/2]\right\}_{n\in\mathbb{N}}\subset(C[0,1/2],||\cdot||_\infty)$, where

$$\Xi_n^{U,V}(\beta)=\frac{n}{2}\int_\beta^{1-\beta}\left(U_n^{-1}(t)-V_n^{-1}(t)\right)^2\,dt,$$

is tight. To this end, we process the subsequent steps:

1. We show that the sequence of real valued random variables $\left\{\Xi_n^{U,V}(0)\right\}_{n\in\mathbb{N}}$ is tight;
2. We establish under Condition 1.3 that for all $0<\beta\leq 1/2$
$$\sqrt{\frac{nm}{n+m}}\left(U_n^{-1}-V_m^{-1}\right)\rightsquigarrow\mathbb{G}$$
in $\ell^\infty[\beta,1-\beta]$, where $\mathbb{G}$ is the Gaussian process defined in the statement of Theorem 2.6;
3. We verify that the sequence $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}}\subset(C[0,1/2],||\cdot||_\infty)$ is measurable;
4. We control the following expectations for small β
$$\mathbb{E}\left[\int_0^\beta\left(U_n^{-1}(t)-V_n^{-1}(t)\right)^2\,dt\right]\text{ and }\mathbb{E}\left[\int_{1-\beta}^1\left(U_n^{-1}(t)-V_n^{-1}(t)\right)^2\,dt\right].$$

In order to establish the first step of the above strategy, we prove that Condition 1.3 implies

$$J_2\left(\mu^U\right)=\int_{-\infty}^\infty\frac{U(t)(1-U(t))}{u(t)}\,dt<\infty\text{ and }J_2\left(\mu^V\right)=\int_{-\infty}^\infty\frac{V(t)(1-V(t))}{v(t)}\,dt<\infty.$$

This will allow us to conclude the tightness of $\left\{\Xi_n^{U,V}(0)\right\}_{n\in\mathbb{N}}$ by Theorem 2.5.

Lemma B.9. *Assume Setting 1.1. Then, Condition 1.3 particularly implies that*

$$J_2\left(\mu^U\right)=\int_{-\infty}^\infty\frac{U(t)(1-U(t))}{u(t)}\,dt<\infty\tag{B.13}$$

as well as

$$J_2\left(\mu^V\right)=\int_{-\infty}^\infty\frac{V(t)(1-V(t))}{v(t)}\,dt<\infty.\tag{B.14}$$

Proof of Lemma B.9. By Corollary A.22 of Bobkov and Ledoux [8] it holds for all $p\geq 1$ that

$$J_p\left(\mu^U\right)=\int_{-\infty}^\infty\frac{[U(t)(1-U(t))]^{p/2}}{(u(t))^{p-1}}\,dt=\int_0^1\left|(U^{-1})'(t)\right|^p(t(1-t))^{p/2}\,dt.$$

In consequence, Condition 1.3 yields that

$$J_2\left(\mu^U\right) \leq \int_0^1 c_U t^{2\gamma_1+1}(1-t)^{2\gamma_2+1}\, dt \overset{(i)}{=} c_U \frac{\Gamma(2\gamma_1+2)\Gamma(2\gamma_1+2)}{\Gamma(2\gamma_1+2\gamma_1+2)} < \infty,$$

where (i) follows as $\gamma_1, \gamma_2 > -1$ by assumption. This gives (B.13). Clearly, (B.14) follows by the analogue arguments. □

Next, we come to the second step of the afore mentioned strategy. The proof of the convergence stated boils down to the observation that Condition 1.3 implies Condition 1.2 for all $\beta > 0$ and an application of Lemma B.7.

Lemma B.10. *Let $0 < \beta < 1/2$, assume Setting 1.1, suppose that Condition 1.3 is met, let $\mu^U = \mu^V$ and let $m, n \to \infty$ such that $\frac{n}{n+m} \to \lambda \in (0,1)$. Then, it follows*

$$\sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - V_m^{-1}\right) \rightsquigarrow \mathbb{G}$$

in $\ell^\infty[\beta, 1-\beta]$, where $\mathbb{G}$ is the Gaussian process defined in Theorem 2.5.

Proof of Lemma B.10. Since $\mu^U = \mu^V$, i.e., $U^{-1} = V^{-1}$, it holds

$$\begin{aligned}\sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - V_m^{-1}\right) &= \sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - U^{-1} + U^{-1} - V_m^{-1}\right)\\ &= \sqrt{\frac{nm}{n+m}}\left(U_n^{-1} - U^{-1}\right) - \sqrt{\frac{nm}{n+m}}\left(V_m^{-1} - V^{-1}\right).\end{aligned}$$

Condition 1.3 implies that for all $\beta > 0$ there is an $\epsilon > 0$ such that u is strictly positive and continuous on $[U^{-1}(\beta) - \epsilon, U^{-1}(1-\beta) + \epsilon]$. Further, U_n^{-1} and V_m^{-1} are independent and hence this lemma is a direct consequence of Lemma B.7. □

Now, we demonstrate the measurability of the stochastic process $\Xi_n^{U,V}$.

Lemma B.11. *Assume Setting 1.1. Then, the random element*

$$\Xi_n^{U,V}(\beta) = \int_\beta^{1-\beta} \left|\sqrt{\frac{n}{2}}\left(U_n^{-1}(s) - V_n^{-1}(s)\right)\right|^2 ds$$

is measurable for $\beta \in [0, 1/2]$ and $n \in \mathbb{N}$. Furthermore, the process $\Xi_n^{U,V}$ is measurable as a random element of $(C[0,1/2], ||\cdot||_\infty)$ for $n \in \mathbb{N}$.

Proof of Lemma B.11. We begin by proving the first statement. Since the process $\mathbb{Q}_n^{U,V} = \sqrt{\frac{n}{2}}(U_n^{-1} - V_n^{-1})$ is left-continuous for $n \in \mathbb{N}$, it is measurable as a function from $\Omega \times [0,1] \to$

$\mathbb{R}$ [13, Chap. 2]. Thus, Theorem 18.3 of Billingsley [6] induces that $\Xi_n^{U,V}(\beta)$ is measurable for $\beta \in [0, 1/2]$ and $n \in \mathbb{N}$.

The second statement is a direct consequence of the first [7, Sec. 7]. □

Finally, we come to the last step of the previously presented strategy. Using a technically somewhat more involved variation of the partitioning idea of the proof of Theorem 2.5, we can demonstrate the subsequent bounds for the expectations $\mathbb{E}\left[\int_0^\beta \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2 \, dt\right]$ and $\mathbb{E}\left[\int_{1-\beta}^1 \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2 \, dt\right]$. In order to increase the readability of this section, we postpone its proof to Section B.6.

Lemma B.12. *Suppose Setting 1.1 and Condition 1.3 are met. Let $\mu^U = \mu^V$, let $n \geq 100$, let $0 \leq \beta = \beta_n < 1/6$ and let $n\beta > 8$. Then, it follows that*

$$\mathbb{E}\left[\int_0^\beta \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2 \, dt\right] \leq \frac{2C_1}{n-1}\left(4\beta\left(1 + 2\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)\right)^{2\gamma_1+2} + o\left(n^{-1}\right) \quad \text{(B.15)}$$

as well as

$$\mathbb{E}\left[\int_{1-\beta}^1 \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2 \, dt\right] \leq \frac{2C_2}{n-1}\left(4\beta\left(1 + 2\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)\right)^{2\gamma_2+2} + o\left(n^{-1}\right), \quad \text{(B.16)}$$

where C_1 and C_2 denote finite constants that are independent of β.

Before we come to the tightness of $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}} \subset \left(C[0,1/2], ||\cdot||_\infty\right)$, we use Lemma B.10 and Lemma B.12 to show one further technical lemma.

Lemma B.13. *Under Condition 1.3, it holds for all $\epsilon > 0$ that*

$$\lim_{\delta\to 0^+} \limsup_{n\to\infty} \mathbb{P}\left(\omega(\Xi_n^{U,V}, \delta) > \epsilon)\right) = 0,$$

where $\omega(\cdot,\cdot)$ is defined for $f : [0, 1/2] \to \mathbb{R}$ and $0 < \delta \leq 1/2$ as $\omega(f, \delta) = \sup_{|s-t|\leq\delta} |f(s) - f(t)|$.

Proof. Let $0 < \delta < 1/20$ and let $0 \le s, t \le 1/2$. We have that

$$\begin{aligned}
&\mathbb{P}\left(\omega(\Xi_n^{U,V}, \delta) > \epsilon\right) = \mathbb{P}\left(\sup_{|s-t|\le\delta} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon\right) \\
=&\mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t<2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon \text{ or } \sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon\right) \\
\le&\mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t<2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon\right) + \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon\right) \\
=& \,\mathbf{I} + \mathbf{II}. \qquad\qquad (B.17)
\end{aligned}$$

In the following, we consider both summands separately.

First Summand: Since $\Xi_n^{U,V}(\beta)$ is monotonically decreasing in β,

$$\begin{aligned}
\mathbf{I} \le&\mathbb{P}\left(\Xi_n^{U,V}(0) - \Xi_n^{U,V}(3\delta) > \epsilon\right) \\
=&\mathbb{P}\left(\int_0^{3\delta} |\mathbb{Q}_n^{U,V}(s)|^2\, ds + \int_{1-3\delta}^1 |\mathbb{Q}_n^{U,V}(s)|^2\, ds > \epsilon\right),
\end{aligned}$$

where $\mathbb{Q}_n^{U,V} = \sqrt{\frac{n}{2}}\left(U_n^{-1} - V_n^{-1}\right)$. Further, we obtain that

$$\begin{aligned}
&\mathbb{P}\left(\int_0^{3\delta} |\mathbb{Q}_n^{U,V}(s)|^2\, ds + \int_{1-3\delta}^1 |\mathbb{Q}_n^{U,V}(s)|^2\, ds > \epsilon\right) \\
\le& \frac{2}{\epsilon}\left(\mathbb{E}\left[\int_0^{3\delta} |\mathbb{Q}_n^{U,V}(s)|^2\, ds\right] + \mathbb{E}\left[\int_{1-3\delta}^1 |\mathbb{Q}_n^{U,V}(s)|^2\, ds\right]\right).
\end{aligned}$$

As $3\delta < 1/6$, we can conclude with Lemma B.12 that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{2}{\epsilon}\mathbb{E}\left[\int_0^{3\delta} \left|\mathbb{Q}_n^{U,V}(s)\right|^2\, ds\right] \\
\le& \limsup_{n\to\infty} \frac{n}{\epsilon}\left(\frac{2C_1}{n-1}\left(6\delta\left(1 + 2\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)\right)^{2\gamma_1+2} + o\left(n^{-1}\right)\right) \le C_2\delta^{2\gamma_1+2}.
\end{aligned}$$

Here, C_1 and C_2 denote finite constants that are independent of δ. Similarly,

$$\limsup_{n\to\infty} \mathbb{E}\left[\int_{1-3\delta}^1 |\mathbb{Q}_n^{U,V}(s)|^2\, ds\right] \le C_3\delta^{2\gamma_2+2},$$

where C_3 denotes a finite constant independent of δ. Since we have by assumption that $\gamma_1, \gamma_2 > -1$, i.e., $2\gamma_1 + 2 > 0$ and $2\gamma_2 + 2 > 0$, it follows that

$$\lim_{\delta \to 0+} \limsup_{n \to \infty} \mathbb{P}\left(\sup_{\substack{|s-t| \leq \delta, \\ t < 2\delta}} \left| \Xi_n^{U,V}(s) - \Xi_n^{U,V}(t) \right| > \epsilon \right) = 0.$$

Second Summand: In order to handle **II** in (B.17), we want to make use of the fact that for $\delta > 0$ the process $\mathbb{Q}_n^{U,V} = \sqrt{\frac{n}{2}}(U_n^{-1} - V_n^{-1})$ converges to a Gaussian process in $\ell^\infty[\delta, 1-\delta]$ given Condition 1.3 (see Lemma B.10 in the supplement). To this end, we verify that the function

$$\Upsilon : \ell^\infty[\delta, 1-\delta] \times \ell^\infty[\delta, 1-\delta] \to \mathbb{R},$$
$$(f, g) \mapsto \sup_{\substack{|s-t| \leq \delta, \\ t \geq 2\delta}} \left| \int_s^{1-s} f^2(x)\, dx - \int_t^{1-t} g^2(x)\, dx \right|$$

is continuous.

Let in the following $||\cdot||_\infty$ denote the norm of $\ell^\infty[\delta, 1-\delta]$. Let $((f_n, g_n))_{n \in \mathbb{N}} \subset \ell^\infty[\delta, 1-\delta] \times \ell^\infty[\delta, 1-\delta]$ be a sequence such that $(f_n, g_n) \to (f, g)$ with respect to the product norm $||(f,g)|| := ||f||_\infty + ||g||_\infty$. Then, the inverse triangle inequality yields

$$\begin{aligned}
&|\Upsilon(f_n, g_n) - \Upsilon(f, g)| \\
\leq& \sup_{\substack{|s-t| \leq \delta, \\ t \geq 2\delta}} \left| \int_s^{1-s} f_n^2(x) - f^2(x)\, dx - \int_t^{1-t} g_n^2(x) - g^2(x)\, dx \right| \\
\leq& \sup_{\substack{|s-t| \leq \delta, \\ t \geq 2\delta}} \left| \int_s^{1-s} \|f_n + f\|_\infty \|f_n - f\|_\infty\, dx + \int_t^{1-t} \|g + g_n\|_\infty \|g - g_n\|_\infty\, dx \right| \\
\leq& (1-\delta) \max\left\{ \|f_n\|_\infty + \|f\|_\infty,\ \|g_n\|_\infty + \|g\|_\infty \right\} \|(f - f_n, g - g_n)\| \overset{n \to \infty}{\to} 0.
\end{aligned}$$

Thus, we have shown $\lim_{n \to \infty} \Upsilon(f_n, g_n) = \Upsilon(f, g)$, i.e., that Υ is sequentially continuous. Hence, a combination of Lemma B.10 and the Continuous Mapping Theorem [40, Thm. 1.3.6] yields that

$$\Upsilon\left(\mathbb{Q}_n^{U,V}, \mathbb{Q}_n^{U,V}\right) \rightsquigarrow \Upsilon(\mathbb{G}, \mathbb{G}),$$

where $\mathbb{G}$ denotes the centered Gaussian process defined in Theorem 2.5. Further, Lemma B.11 shows that $\Xi_n^{U,V}(\beta)$ is measurable for $\beta \in [\delta, 1-\delta]$ and $n \in \mathbb{N}$. As $\Xi_n^{U,V}$ is continuous in β this induces the measurability of $\Upsilon\left(\mathbb{Q}_n^{U,V}, \mathbb{Q}_n^{U,V}\right)$ for $n \in \mathbb{N}$. Thus, we find that

$$\Upsilon\left(\mathbb{Q}_n^{U,V}, \mathbb{Q}_n^{U,V}\right) \Rightarrow \Upsilon(\mathbb{G}, \mathbb{G}).$$

Let $A = [\epsilon, \infty) \subset \mathbb{R}$. Then, the set A is closed and $(\epsilon, \infty) \subset A$. Hence, an application of the Portmanteau-Theorem [7, Thm. 2.1] yields that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon \right) \le \limsup_{n\to\infty} \mathbb{P}\left(\Upsilon\left(\mathbb{Q}_n^{U,V}, \mathbb{Q}_n^{U,V}\right) \in A\right)$$
$$\le \mathbb{P}\left(\Upsilon\left(\mathbb{G}, \mathbb{G}\right) \ge \epsilon\right).$$

Next, we remark that

$$\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| = \sup_{\substack{|s-t|\le\delta,\\ s\le t, t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right|.$$

Hence, we can assume for the treatment of this summand that $s \le t$. With this, we obtain that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon \right)$$
$$\le \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left| \int_s^{1-s} \mathbb{G}^2(x)\,dx - \int_t^{1-t} \mathbb{G}^2(x)\,dx \right| \ge \epsilon \right)$$
$$\le \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \int_s^t \mathbb{G}^2(x)\,dx \ge \frac{\epsilon}{2} \right) + \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \int_{1-t}^{1-s} \mathbb{G}^2(x)\,dx \ge \frac{\epsilon}{2} \right).$$

In the following, we focus on the first term. As $0 < \delta < 1/20$ and $t \le 1/2$, it holds

$$\mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \int_s^t \mathbb{G}^2(x)\,dx \ge \frac{\epsilon}{2} \right) \le \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \left(\sup_{x\in[\delta,1-\delta]} \mathbb{G}^2(x) \right)(t-s) \ge \frac{\epsilon}{2} \right)$$
$$= \mathbb{P}\left(\sup_{x\in[\delta,1-\delta]} \mathbb{G}^2(x) \ge \frac{\epsilon}{\delta} \right) \le \sqrt{\frac{\delta}{\epsilon}}\,\mathbb{E}\left[\sup_{x\in[\delta,1-\delta]} |\mathbb{G}(x)| \right].$$

By Lemma F.11 the Gaussian process $\mathbb{G}$ is continuous on $[\delta, 1-\delta]$ under the assumptions made, i.e., almost surely bounded on $[\delta, 1-\delta]$. Thus, Theorem 2.1.1 of Adler and Taylor [1] ensures that $\mathbb{E}\left[\sup_{x\in[\delta,1-\delta]} |\mathbb{G}(x)|\right] < \infty$. Hence, we find that

$$\lim_{\delta\to 0+} \mathbb{P}\left(\sup_{\substack{|s-t|\le\delta,\\ t\ge 2\delta}} \int_s^t \mathbb{G}^2(x)\,dx \ge \frac{\epsilon}{2} \right) \le \lim_{\delta\to 0+} \sqrt{\frac{\delta}{\epsilon}}\,\mathbb{E}\left[\sup_{x\in[\delta,1-\delta]} |\mathbb{G}(x)| \right] = 0.$$

Analogously,

$$\lim_{\delta\to 0+} \mathbb{P}\left(\sup_{\substack{|s-t|\leq\delta,\\ t\geq 2\delta}} \int_{1-t}^{1-s} \mathbb{G}^2(x)\,dx > \frac{\epsilon}{2}\right) = 0.$$

This concludes the treatment of the second summand in (B.17).

Combining the results for **I** and **II**, we find that

$$\lim_{\delta\to 0+} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{|s-t|\leq\delta} \left|\Xi_n^{U,V}(s) - \Xi_n^{U,V}(t)\right| > \epsilon\right) = 0.$$

Thus, we have proven Lemma B.13. □

Now, we obtain the tightness of the sequence $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}}$ in $C[0,1/2]$ as a simple consequence of the above results.

Corollary B.14. *Under Condition 1.3, the sequence* $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}}$ *is tight in the function space* $(C[0,1/2], ||\cdot||_\infty)$.

Proof of Corollary B.14. By Theorem 7.3 in Billingsley [7] (and a rescaling argument) it is sufficient to prove that the sequence $\left\{\Xi_n^{U,V}(0)\right\}_{n\in\mathbb{N}}$ is tight in $\mathbb{R}$ and that

$$\lim_{\delta\to 0+} \limsup_{n\to\infty} \mathbb{P}\left(\omega(\Xi_n^{U,V},\delta)) > \epsilon)\right) = 0.$$

We have already noted that $\left\{\Xi_n^{U,V}(0)\right\}_{n\in\mathbb{N}}$ is tight (by Theorem 2.5, which is applicable due to Lemma B.9) and thus Lemma B.13 yields Corollary B.14. □

We conclude the proof of Theorem 2.6 (ii) by using the Skorohod Representation Theorem [7, Thm. 6.7] to verify that the tightness of $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}}$ induces that

$$\frac{n}{2}\int_0^1 \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2\, dt \rightsquigarrow \int_0^1 (\mathbb{G}(t))^2\, dt.$$

Proof of Theorem 2.6 (ii). By Corollary B.14 the sequence

$$\{\Xi_n^{U,V}\}_{n\in\mathbb{N}} = \left\{\int_\beta^{1-\beta} \left|\sqrt{\frac{n}{2}}(U_n^{-1}(s) - V_n^{-1}(s))\right|^2 ds \,\middle|\, \beta\in[0,1/2]\right\}_{n\in\mathbb{N}}. \tag{B.18}$$

is tight in $(C[0,1/2], ||\cdot||_\infty)$. As tightness implies relative compactness [6, Thm. 5.1], this means that every subsequence of $\left\{\Xi_n^{U,V}\right\}_{n\in\mathbb{N}}$ contains a subsequence that is weakly convergent in $C[0,1/2]$. Let $\left\{\Xi_{n_k}^{U,V}\right\}_{n_k\in\mathbb{N}}$ be such a convergent subsubsequence, whose limit we denote by Ξ. Since the elements of the sequence $\left\{\Xi_{n_k}^{U,V}\right\}_{n_k\in\mathbb{N}}$ are measurable by Lemma B.11 and $(C[0,1/2], ||\cdot||_\infty)$ is a Polish space, it follows that there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ and random variables $\left(\Xi_{n_k}^{U,V}\right)' \stackrel{\mathcal{D}}{=} \Xi_{n_k}^{U,V}$, $n_k \in \mathbb{N}$, and $\Xi' \stackrel{\mathcal{D}}{=} \Xi$ such that $\left(\Xi_{n_k}^{U,V}\right)' \to \Xi'$ for every $\omega \in \widetilde{\Omega}$, as $n_k \to \infty$ [7, Thm. 6.7]. By a slight abuse of notation we will drop the additional prime in the following.

By construction we have that $\Xi_{n_k}^{U,V}$ converges to a limit in $(C[0,1/2], ||\cdot||_\infty)$ as $n \to \infty$. This particularly implies that the sequences $\Xi_{n_k}^{U,V}(\beta)$ converge in $\mathbb{R}$ for $\beta \in [0,1/2]$ and $\omega \in \widetilde{\Omega}$. Under the assumptions made it follows for $\beta \in (0,1/2)$ by Theorem 2.5 (i) that

$$\Xi_{n_k}^{U,V} = \int_\beta^{1-\beta} \left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s) - V_{n_k}^{-1}(s)\right)\right|^2 ds \to \int_\beta^{1-\beta} |\mathbb{G}(s)|^2 \, ds,$$

as $n_k \to \infty$ for all $\omega \in \widetilde{\Omega}$. In the following, we aim to identify the limit of $\Xi_{n_k}^{U,V}(0)$.

Since

$$\begin{aligned}\Xi_{n_k}^{U,V}(0) &= \int_0^1 \left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s) - V_{n_k}^{-1}(s)\right)\right|^2 ds \\ &\geq \int_\beta^{1-\beta} \left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s) - V_{n_k}^{-1}(s)\right)\right|^2 ds = \Xi_{n_k}^{U,V}(\beta),\end{aligned}$$

we find

$$\lim_{n_k\to\infty} \Xi_{n_k}^{U,V}(0) = \liminf_{n_k\to\infty} \Xi_{n_k}^{U,V}(0) \geq \liminf_{n_k\to\infty} \Xi_{n_k}^{U,V}(\beta) = \int_\beta^{1-\beta} |\mathbb{G}(s)|^2 \, ds,$$

for all $\omega \in \widetilde{\Omega}$. Now, letting $\beta \searrow 0$, monotone convergence yields

$$\lim_{n_k\to\infty} \Xi_{n_k}^{U,V}(0) \geq \int_0^1 |\mathbb{G}(s)|^2 \, ds.$$

If, for some $\omega \in \widetilde{\Omega}$, we had that

$$\liminf_{n_k\to\infty} \int_0^1 \left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s) - V_{n_k}^{-1}(s)\right)\right|^2 ds > \int_0^1 |\mathbb{G}(s)|^2 \, ds,$$

this would imply

$$\liminf_{n_k\to\infty}\left\{\int_0^{\beta_0}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds+\int_{1-\beta_0}^{1}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right\} \tag{B.19}$$
$$>\int_0^{\beta_0}|\mathbb{G}(s)|^2\,ds+\int_{1-\beta_0}^{1}|\mathbb{G}(s)|^2\,ds,$$

for some $\beta_0>0$. Assume that (B.19) holds for some $\omega\in\widetilde{\Omega}$. Then, there exists $\Delta=\Delta(\omega)>0$ such that

$$\liminf_{n_k\to\infty}\left\{\int_0^{\beta_0}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds+\int_{1-\beta_0}^{1}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right\}$$
$$=\int_0^{\beta_0}|\mathbb{G}(s)|^2\,ds+\int_{1-\beta_0}^{1}|\mathbb{G}(s)|^2\,ds+\Delta. \tag{B.20}$$

At the same time, we have

$$\liminf_{n_k\to\infty}\left\{\int_0^{\beta_1}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds+\int_{1-\beta_1}^{1}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right\}$$
$$=\int_0^{\beta_1}|\mathbb{G}(s)|^2\,ds+\int_{1-\beta_1}^{1}|\mathbb{G}(s)|^2\,ds+\Delta,$$

for any fixed $0<\beta_1<\beta_0$ and therefore

$$\inf_{0<\beta\le\beta_0}\liminf_{n_k\to\infty}\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\ge\Delta, \tag{B.21}$$

since $\inf_{0<\beta\le\beta_0}\left\{\int_0^{\beta}|\mathbb{G}(s)|^2\,ds+\int_{1-\beta}^{1}|\mathbb{G}(s)|^2\,ds\right\}\ge 0$. We find that

$$\mathbb{E}\left[\inf_{0<\beta\le\beta_0}\liminf_{n_k\to\infty}\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right]$$
$$=\mathbb{E}\left[\lim_{\beta\to0+}\liminf_{n_k\to\infty}\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right],$$

where the last equality holds due to monotonicity with respect to β. Further, it holds

$$\begin{aligned}
&\mathbb{E}\left[\inf_{0<\beta\leq\beta_0}\liminf_{n_k\to\infty}\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right]\\
&\overset{(i)}{=}\lim_{\beta\to0+}\mathbb{E}\left[\liminf_{n_k\to\infty}\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right]\\
&\overset{(ii)}{\leq}\lim_{\beta\to0+}\liminf_{n_k\to\infty}\mathbb{E}\left[\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right].
\end{aligned}$$

Here, (i) follows by monotone convergence and (ii) by Fatou's Lemma [6, Thm. 16.3]. We have already considered the last limit in the proof of Lemma B.13. There, we have shown that under the assumptions made

$$\lim_{\beta\to0+}\liminf_{n_k\to\infty}\mathbb{E}\left[\int_{[0,\beta]\cup[1-\beta,1]}\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\right]=0.$$

Hence, (B.19) can only hold on a set of measure zero. Thus, we have demonstrated that

$$\int_0^1\left|\sqrt{\frac{n_k}{2}}\left(U_{n_k}^{-1}(s)-V_{n_k}^{-1}(s)\right)\right|^2 ds\Rightarrow\int_0^1|\mathbb{G}(s)|^2\,ds.$$

The subsequence was chosen arbitrarily. Hence, we conclude that every subsequence of $\left\{\Xi_n^{U,V}(0)\right\}_{n\in\mathbb{N}}$ has a subsequence that converges weakly to $\int_0^1|\mathbb{G}(s)|^2\,ds$. This induces that

$$\int_0^1\left|\sqrt{\frac{n}{2}}(U_n^{-1}(s)-V_n^{-1}(s))\right|^2 ds\Rightarrow\int_0^1|\mathbb{G}(s)|^2\,ds$$

and thus we have demonstrated the claim. □

Remark B.15. Reconsideration of the above proof and the ones of the auxiliary results required highlights that only in the proof of Lemma B.12 we actually make use of the assumption $n=m$. However, it is evident that with similar arguments as used in the proof of Theorem 2.5 we can extend Lemma B.12 to the case $n\neq m$ (see Section B.6). This allows us to apply the same strategy of proof to establish Theorem 2.6 (ii) for the case $n\neq m$.

Remark B.16. Finally, we comment on the implications of our results on the statistics

$$\left(\frac{nm}{n+m}\right)^{p/2}\widehat{DoD}_p\left(\mathcal{X}_n,\mathcal{Y}_m\right):=\left(\frac{nm}{n+m}\right)^{p/2}\int_0^1\left|U_n^{-1}(t)-V_m^{-1}(t)\right|^p dt,$$

where $p\in[1,\infty)$. A careful reconsideration of the above arguments highlights that the proof of Theorem 2.6 (ii) does depend on the choice of $p=2$ via Lemma B.12. However, if we slightly change Condition 1.3, namely demand that there are constants $-2/p < \gamma_1, \gamma_2 < \infty$ and $c_U > 0$ such that

$$\left|(U^{-1})'(t)\right| \leq c_U t^{\gamma_1}(1-t)^{\gamma_2}$$

for $t \in (0,1)$, then we can adapt the proof of Lemma B.12 and thus the one of Theorem 2.6 (ii) to $p \in [1,\infty)$.

B.5 Proof of Theorem 2.7

In this subsection, we derive Theorem 2.7. Its proof follows along the lines of the proof of Theorem 2 of Munk and Czado [32], where the limit distribution of a (truncated) empirical Kantorovich distance under the assumption that the underlying true measures are not equal is derived. However, while Munk and Czado [32] work with classical empirical quantile processes, we have to deal with empirical U-quantile processes in the present setting. As in the previous subsection, we begin with the proof of the first part of Theorem 2.7 and consider its second statement afterwards.

B.5.1 First Part

Just like the proof of Theorem 2.6 (i) (see Section B.4) the one of Theorem 2.7 (i) is based on Lemma B.7. We start by establishing the subsequent theorem.

Theorem B.17. *Let $\beta \in (0,1/2)$, assume Setting 1.1 and let Condition 1.2 hold. Let $DoD_{(\beta)} \neq 0$ and let $n,m \to \infty$ such that $\frac{n}{m+n} \to \lambda \in (0,1)$. Then, it holds*

$$\sqrt{\frac{nm}{n+m}}\left(\int_\beta^{1-\beta}\left(U_n^{-1}(t)-V_m^{-1}(t)\right)^2\,dt - \int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)^2\,dt\right)$$
$$\rightsquigarrow 2\sqrt{\lambda}\int_\beta^{1-\beta}(U^{-1}(t)-V^{-1}(t))\mathbb{G}_1(t)dt - 2\sqrt{1-\lambda}\int_\beta^{1-\beta}(U^{-1}(t)-V^{-1}(t))\mathbb{G}_2(t)dt, \quad \text{(B.22)}$$

where $\mathbb{G}_1$ and $\mathbb{G}_2$ denote the centered, independent Gaussian processes defined in the statement of Lemma B.7.

Proof of Theorem B.17. We will use the 3 following steps in order to prove the claim.

Step 1: Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be the Gaussian processes defined in the theorem's statement. We show that for $n \to \infty$

$$\sqrt{n}\int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(U_n^{-1}(t)-U^{-1}(t)\right)\,dt$$
$$\rightsquigarrow \int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\mathbb{G}_1(t)\,dt$$

as well as that for $m \to \infty$

$$\sqrt{m}\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(V_m^{-1}(t)-V^{-1}(t)\right)\,dt$$
$$\rightsquigarrow \int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\mathbb{G}_2(t)\,dt.$$

Step 2: We prove that for $n,m\to\infty$ such that $\frac{n}{m+n}\to\lambda\in(0,1)$ we have

$$\sqrt{\frac{nm}{n+m}}\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt \xrightarrow{\mathbb{P}} 0$$

as well as

$$\sqrt{\frac{nm}{n+m}}\int_{\beta}^{1-\beta}\left(V_m^{-1}(t)-V^{-1}(t)\right)^2\,dt \xrightarrow{\mathbb{P}} 0.$$

Step 3: Finally, we demonstrate that

$$\begin{aligned}&\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-V_m^{-1}(t)\right)^2\,dt-\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)^2\,dt\\ =&2\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(U_n^{-1}(t)-U^{-1}(t)\right)\,dt\\ -&2\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(V_m^{-1}(t)-V^{-1}(t)\right)\,dt+o_p\left(\left(\frac{mn}{n+m}\right)^{1/2}\right).\end{aligned}$$

Step 1 essentially follows by Lemma B.7. An application of the former lemma gives that for $n\to\infty$

$$\mathbb{U}_n^{-1}=\sqrt{n}\left(U_n^{-1}-U^{-1}\right)\rightsquigarrow\mathbb{G}_1 \tag{B.23}$$

in $\ell^\infty[\beta,1-\beta]$. Further, as U and V are assumed to be continuous, U^{-1} and V^{-1} are continuous as well. This means that

$$\sup_{t\in[\beta,1-\beta]}\left|U^{-1}(t)\right|=C<\infty \quad\text{and}\quad \sup_{t\in[\beta,1-\beta]}\left|V^{-1}(t)\right|=C'<\infty.$$

With this it is straight forward to verify that the map $\phi_1:\ell^\infty[\beta,1-\beta]\to\mathbb{R}$, $f\mapsto\int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)f(t)\,dt$ is Lipschitz continuous. In consequence, the Continuous Mapping Theorem [40, Thm. 1.3.6] yields that as n grows to infinity we have

$$\sqrt{n}\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(U_n^{-1}(t)-U^{-1}(t)\right)\,dt=\phi_1(\mathbb{U}_n^{-1})$$
$$\rightsquigarrow\phi_1(\mathbb{G}_1)=\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\mathbb{G}_1(t)\,dt.$$

The analogous arguments give the corresponding result for

$$\sqrt{m}\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\left(V_m^{-1}(t)-V^{-1}(t)\right)\,dt.$$

This concludes **Step 1**.

In order to show **Step 2**, we will use similar arguments as in the previous step to verify that

$$\sqrt{n}\sqrt{\frac{nm}{n+m}}\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt=O_p(1). \tag{B.24}$$

This in turn suggests that

$$\sqrt{\frac{nm}{n+m}}\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt=o_p(1).$$

We have already seen that under the assumptions made (B.23) holds. Moreover, it is easy to verify that the function $\phi_2:\ell^\infty[\beta,1-\beta]\to\mathbb{R}$, $f\mapsto\int_\beta^{1-\beta}(f(t))^2\,dt$ is continuous. Thus, the Continuous Mapping Theorem [40, Thm. 1.3.6] suggests that for $n\to\infty$ it holds

$$n\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt=\phi_2(\mathbb{U}_n^{-1})\rightsquigarrow\phi_2(\mathbb{G}_1)=\int_{\beta}^{1-\beta}\left(\mathbb{G}_1(t)\right)^2\,dt.$$

As $\mathbb{R}$ is separable and complete, the random variable $\phi_2(\mathbb{G}_1)$ is tight [40, Lemma 1.3.2]. In consequence, we have that

$$n\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt=O_p(1).$$

As $m/(n+m)\to 1-\lambda$ for $m,n\to\infty$, this induces (B.24). The same arguments yield the analogous statement for $\sqrt{\frac{nm}{n+m}}\int_\beta^{1-\beta}(V_m^{-1}(t)-V^{-1}(t))^2\,dt$.

In the end, we come to **Step 3**. For notational purposes we write in this step $\int f$ instead of $\int_\beta^{1-\beta}f(t)\,dt$. We have

$$\begin{aligned}
&\int_{\beta}^{1-\beta}\left(U_n^{-1}(t)-V_m^{-1}(t)\right)^2\,dt-\int_{\beta}^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)^2\,dt\\
=&\int\Big[\left(U_n^{-1}\right)^2-2U_n^{-1}V_m^{-1}+\left(V_m^{-1}\right)^2-\left(U^{-1}\right)^2+2U^{-1}V^{-1}-\left(V^{-1}\right)^2\Big]\\
=&2\int\Big[U_n^{-1}U^{-1}+V_m^{-1}V^{-1}-\left(U^{-1}\right)^2-\left(V^{-1}\right)^2-U_n^{-1}V_m^{-1}+U^{-1}V^{-1}\Big]\\
&+\int\Big[\left(U_n^{-1}-U^{-1}\right)^2+\left(V_m^{-1}-V^{-1}\right)^2\Big].
\end{aligned}$$

By **Step 2** we have

$$\int \left(U_n^{-1} - U^{-1}\right)^2 = \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right)$$

and

$$\int \left(V_m^{-1} - V^{-1}\right)^2 = \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right).$$

Continuing with our calculations leads to

$$\begin{aligned}
=&2\int \left[U_n^{-1}U^{-1} + V_m^{-1}V^{-1} - \left(U^{-1}\right)^2 - \left(V^{-1}\right)^2 - U_n^{-1}V_m^{-1} + U^{-1}V^{-1}\right] \\
&+ \int \left[\left(U_n^{-1} - U^{-1}\right)^2 + \left(V_m^{-1} - V^{-1}\right)^2\right] \\
=&2\int \left[U_n^{-1}U^{-1} + V_m^{-1}V^{-1} - \left(U^{-1}\right)^2 - \left(V^{-1}\right)^2 - U_n^{-1}V_m^{-1} + U^{-1}V^{-1}\right] \\
&+ \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right) \\
=&2\int \left[U^{-1}\left(U_n^{-1} - U^{-1}\right) + V^{-1}\left(V_m^{-1} - V^{-1}\right) + U_n^{-1}\left(V^{-1} - V_m^{-1}\right)\right. \\
&\left. + V_m^{-1}\left(U^{-1} - U_n^{-1}\right)\right] + 2\int \left[\left(U^{-1} - U_n^{-1}\right)\left(V^{-1} - V_m^{-1}\right)\right] \\
&+ \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right).
\end{aligned}$$

A combination of **Step 2** and the Cauchy-Schwarz inequality induces

$$\int \left[\left(U^{-1} - U_n^{-1}\right)\left(V^{-1} - V_m^{-1}\right)\right] = \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right). \tag{B.25}$$

With this in mind, we proceed with our calculations. Taking Equation (B.25) into account, the last term becomes

$$\begin{aligned}
&2\int \left[\left(U^{-1} - V_m^{-1}\right)\left(U_n^{-1} - U^{-1}\right)\right] + 2\int \left[\left(V^{-1} - U_n^{-1}\right)\left(V_m^{-1} - V^{-1}\right)\right] \\
&+ \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right) \\
=&2\Bigg(\int \left[\left(U^{-1} - V^{-1}\right)\left(U_n^{-1} - U^{-1}\right)\right] + \int \left[\left(V^{-1} - V_m^{-1}\right)\left(U_n^{-1} - U^{-1}\right)\right] \\
&- \int \left[\left(U^{-1} - V^{-1}\right)\left(V_m^{-1} - V^{-1}\right)\right] + \int \left[\left(U^{-1} - U_n^{-1}\right)\left(V_m^{-1} - V^{-1}\right)\right]\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right)\\
=&2\int\Big[\left(U^{-1}-V^{-1}\right)\left(U_n^{-1}-U^{-1}\right)\Big] - 2\int\Big[\left(U^{-1}-V^{-1}\right)\left(V_m^{-1}-V^{-1}\right)\Big]\\
&+ \mathrm{o}_p\left(\left(\frac{mn}{n+m}\right)^{-1/2}\right),
\end{aligned}
$$

where we used again (B.25) in the last step. This concludes **Step 3**.

The claim now follows as

$$\sqrt{\frac{nm}{n+m}}\int\Big[\left(U^{-1}-V^{-1}\right)\left(U_n^{-1}-U^{-1}\right)\Big]$$

and

$$\sqrt{\frac{nm}{n+m}}\int\Big[\left(U^{-1}-V^{-1}\right)\left(V_m^{-1}-V^{-1}\right)\Big]$$

are independent and their distributional limits were calculated in **Step 1**. □

We obtain Theorem 2.7 (i) once we have verified that the limit distribution in (B.22) is normally distributed with mean zero and variance as stated.

Proof of Theorem 2.7 (i). Since the requirements of Theorem B.17 are given by assumption, it only remains to show that the limit distribution in (B.22) is normally distributed with correct mean and variance. This is a straightforward application of Theorem 8.17 of Dümbgen [19].

First of all, $\mathcal{T}=[\beta,1-\beta]$ is a compact metric space and by Lemma F.11 the process $\mathbb{G}_1$ is continuous on $\mathcal{T}$, i.e., $\mathbb{G}_1$ is uniformly continuous on $\mathcal{T}$. Let now $\mathcal{Q}_\beta$ be the uniform distribution on $[\beta,1-\beta]$ and let $\zeta(t)=U^{-1}(t)-V^{-1}(t)$ for $t\in[\beta,1-\beta]$. It follows

$$\int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\mathcal{Q}_\beta(dt)=\frac{1}{1-2\beta}\int_\beta^{1-\beta}\zeta(t)\,dt<\infty,$$

since the function ζ is continuous on $\mathcal{T}$. Consequently, Theorem 8.17 of Dümbgen [19] implies that

$$\int_\beta^{1-\beta}\left(U^{-1}(t)-V^{-1}(t)\right)\mathbb{G}_1(t)\,\mathcal{Q}_\beta(dt)=\frac{1}{1-2\beta}\int_\beta^{1-\beta}\zeta(t)\mathbb{G}_1(t)\,dt$$

is normally distributed with mean

$$\frac{1}{1-2\beta}\int_\beta^{1-\beta}\zeta(t)\mathbb{E}\left[\mathbb{G}_1(t)\right]\,dt=0$$

and variance

$$\sigma_1^2 = \int_\beta^{1-\beta}\int_\beta^{1-\beta} \zeta(t)\zeta(t')\mathrm{Cov}(\mathbb{G}_1(t),\mathbb{G}_1(t'))\,\mathcal{Q}_\beta(dt)\,\mathcal{Q}_\beta(dt')$$
$$= C^2\int_\beta^{1-\beta}\int_\beta^{1-\beta} \frac{4\zeta(t)\zeta(t')}{(u\circ U^{-1}(t))(u\circ U^{-1}(t'))}\Gamma_{d_\mathcal{X}}\left(U^{-1}(t),U^{-1}(t')\right)\,dt\,dt',$$

where $C = 1/(1-2\beta)$. Let $x = U^{-1}(t)$ and $y = U^{-1}(t')$, then

$$\frac{dx}{dt} = \frac{d}{dt}U^{-1}(t) = \frac{1}{u\circ U^{-1}(t)},$$

respectively

$$\frac{dy}{dt'} = \frac{d}{dt'}U^{-1}(t') = \frac{1}{u\circ U^{-1}(t')}.$$

As a consequence, the variance is given as

$$\sigma_1^2 = 4C^2\int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)}\int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)}(x - V^{-1}(U(x)))(y - V^{-1}(U(y)))\Gamma_{d_\mathcal{X}}\,(x,y)\;dx\,dy$$

From the above it now follows by multiplying with $2\sqrt{\lambda}(1-2\beta)$, that the term

$$2\sqrt{\lambda}\int_\beta^{1-\beta}(U^{-1}(t) - V^{-1}(t))\mathbb{G}_1(t)\,dt$$

is normally distributed with mean 0 and variance

$$16\lambda\int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)}\int_{U^{-1}(\beta)}^{U^{-1}(1-\beta)}(x - V^{-1}(U(x)))(y - V^{-1}(U(y)))\Gamma_{d_\mathcal{X}}\,(x,y)\;dx\,dy.$$

The analogue arguments for

$$\int_\beta^{1-\beta}(U^{-1}(t) \;- V^{-1}(t))\mathbb{G}_2(t)\,dt$$

and the independence of $\mathbb{G}_1$ and $\mathbb{G}_2$ yield the claim. □

B.5.2 Second Part

In order to prove Theorem 2.7 (*ii*) we will pursue a similar strategy as for the proof of Theorem 2.7 (*i*), i.e., we will first derive the analog of Theorem B.17 for $\beta = 0$ and afterwards verify that the limiting random variable is normally distributed with mean and variance as stated.

As already mentioned, the most important step of the proof of Theorem B.17 is to derive the limit distributions of

$$\int_{\beta}^{1-\beta}(U^{-1}(t)-V^{-1}(t))\mathbb{U}_n^{-1}(t)\,dt \quad\text{and}\quad \int_{\beta}^{1-\beta}(U^{-1}(t)-V^{-1}(t))\mathbb{V}_m^{-1}(t)\,dt, \tag{B.26}$$

where $\mathbb{U}_n^{-1}:=\sqrt{n}\big(U_n^{-1}-U^{-1}\big)$ and $\mathbb{V}_m^{-1}:=\sqrt{m}\big(V_m^{-1}-V^{-1}\big)$. Under Condition 1.2 these can be derived via the distributional limits for the empirical U-quantile processes $\mathbb{U}_n^{-1}$ and $\mathbb{V}_m^{-1}$ in $\ell^\infty[\beta,1-\beta]$. However, as already argued, we cannot expect $\ell^\infty(0,1)$-convergence of $\mathbb{U}_n^{-1}$ and $\mathbb{V}_m^{-1}$ if the densities u and v vanish at the border of their support. Reconsidering (B.26), we realize that $\ell^1(0,1)$-convergence of $\mathbb{U}_n^{-1}$ and $\mathbb{V}_m^{-1}$ is sufficient to derive the corresponding limiting distributions. Convergence in $\ell^1(0,1)$ is much weaker than convergence in $\ell^2(0,1)$ or $\ell^\infty(0,1)$. Indeed, it turns out that this convergence can quickly be verified. Given an interval $[a,b]\subset\mathbb{R}$ let $D[a,b]$ denote the space of cádlág functions on $[a,b]$ (equipped with the supremum norm) and $\mathbb{D}_2\subset D[a,b]$ the set of distribution functions of measures that concentrate on $(a,b]$. Using ideas from Kaji [22] we can show the following.

Lemma B.18. *Let F have compact support on $[a,b]$ and let F be continuously differentiable on its support with derivative f that is strictly positive on (a,b) (Possibly, $f(a)=0$ and/or $f(b))=0$). Then the inversion functional $\phi_{inv}:F\mapsto F^{-1}$ as a map $\mathbb{D}_2\subset D[a,b]\to\ell^1(0,1)$ is Hadamard-differentiable at F tangentially to $C[a,b]$ with derivative $\alpha\mapsto-(\alpha/f)\circ F^{-1}$.*

Proof of Lemma B.18. Let $h_n\to h$ uniformly in $D[a,b]$, where h is continuous, $t_n\to 0$ and $F+t_nh_n\in\mathbb{D}_2$ for all $n\geq 1$ sufficiently large. Let $\nabla\phi_{inv_F}(\alpha)=-(\alpha/f)\circ F^{-1}$. We have to demonstrate that

$$\left\|\frac{\phi_{inv}(F+t_nh_n)-\phi_{inv}(F)}{t_n}-\nabla\phi_{inv_F}(h)\right\|_{\ell^1(0,1)}\to 0,$$

as $n\to\infty$. We realize that for every $\epsilon>0$ there exist $a_\epsilon,b_\epsilon\in[a,b]$ such that $\max\{F(a_\epsilon),1-F(b_\epsilon)\}<\epsilon$ and f is strictly positive on $[a_\epsilon,b_\epsilon]$. It follows that

$$\begin{aligned}&\left\|\frac{\phi_{inv}(F+t_nh_n)-\phi_{inv}(F)}{t_n}-\nabla\phi_{inv_F}(h)\right\|_{\ell^1(0,1)}\\ \leq&\int_{F(a_\epsilon)+\epsilon}^{F(b_\epsilon)-\epsilon}\left|\frac{\phi_{inv}(F+t_nh_n)-\phi_{inv}(F)}{t_n}-\nabla\phi_{inv_F}(h)\right|(s)\,ds\\ +&\int_{0}^{2\epsilon}\left|\frac{\phi_{inv}(F+t_nh_n)-\phi_{inv}(F)}{t_n}-\nabla\phi_{inv_F}(h)\right|(s)\,ds\\ +&\int_{1-2\epsilon}^{1}\left|\frac{\phi_{inv}(F+t_nh_n)-\phi_{inv}(F)}{t_n}-\nabla\phi_{inv_F}(h)\right|(s)\,ds.\end{aligned}$$

Next, we treat the summands separately. The claim follows once we have shown that the first summand vanishes for all $\epsilon > 0$ as $n \to \infty$ and the other two summands become arbitrarily small for ϵ small and $n \to \infty$.

First summand: We start with the first summand. Since its requirements are fulfilled for all $\epsilon > 0$, we have by Lemma 3.9.23 of van der Vaart and Wellner [40] that

$$\sup_{s\in[F(a_\epsilon)+\epsilon, F(b_\epsilon)-\epsilon]} \left| \frac{\phi_{inv}(F+t_n h_n) - \phi_{inv}(F)}{t_n} - \nabla\phi_{inv_F}(h) \right|(s) \to 0, \tag{B.27}$$

as $n \to \infty$. Thus, the same holds for

$$\int_{F(a_\epsilon)+\epsilon}^{F(b_\epsilon)-\epsilon} \left| \frac{\phi_{inv}(F+t_n h_n) - \phi_{inv}(F)}{t_n} - \nabla\phi_{inv_F}(h) \right|(s)\, ds,$$

which is bounded by (B.27).

Second summand: We have

$$\begin{aligned} &\int_0^{2\epsilon} \left| \frac{\phi_{inv}(F+t_n h_n) - \phi_{inv}(F)}{t_n} - \nabla\phi_{inv_F}(h) \right|(s)\, ds \\ \le &\int_0^{2\epsilon} \left| \frac{\phi_{inv}(F+t_n h_n) - \phi_{inv}(F)}{t_n} \right|(s)\, ds + \int_0^{2\epsilon} |\nabla\phi_{inv_F}(h)|(s)\, ds. \end{aligned} \tag{B.28}$$

In the following, we consider both terms separately. For the first term, we find that

$$\begin{aligned} &\int_0^{2\epsilon} \left| \frac{\phi_{inv}(F+t_n h_n) - \phi_{inv}(F)}{t_n} \right|(s)\, ds \\ =&\frac{1}{|t_n|} \int_0^{2\epsilon} \left|(F+t_n h_n)^{-1}(s) - F^{-1}(s)\right|\, ds. \end{aligned}$$

Next, we realize that for $G \in \mathbb{D}_2 \subset D[a,b]$ and $s \in (0,1)$, we have that

$$G^{-1}(s) = -\int_a^b \mathbb{1}_{\{s\le G(x)\}}\, dx + b.$$

Since $F \in \mathbb{D}_2$ and $F + t_n h_n \in \mathbb{D}_2$ for all $n \in \mathbb{N}$, this yields that

$$\begin{aligned} &\frac{1}{|t_n|} \int_0^{2\epsilon} \left|(F+t_n h_n)^{-1}(s) - F^{-1}(s)\right|\, ds \\ \le&\frac{1}{|t_n|} \int_a^b \int_0^{2\epsilon} \left|\mathbb{1}_{\{s\le (F+t_n h_n)(x)\}} - \mathbb{1}_{\{s\le F(x)\}}\right|\, ds\, dx, \end{aligned}$$

where we applied the Theorem of Tonelli/Fubini [6, Thm. 18.3] in the last step. Let in the following $F_{t_n h_n} = F + t_n h_n$. Then, we obtain that

$$\begin{aligned}&\int_0^{2\epsilon} \left| \mathbb{1}_{\{s \le F_{t_n h_n}(x)\}} - \mathbb{1}_{\{s \le F(x)\}} \right| \, ds \\ \le &\begin{cases} \left|F(x) - F_{t_n h_n}(x)\right|, & \text{if } \min\{F(x), F_{t_n h_n}(x)\} \le 2\epsilon. \\ 0, & \text{else.} \end{cases}\end{aligned}$$

Let now $x \ge F^{-1}\left(2\epsilon + \|h_n t_n\|_\infty\right)$, then it follows that

$$F(x) \ge 2\epsilon + \|h_n t_n\|_\infty \ge 2\epsilon$$

as well as

$$F_{t_n h_n}(x) \ge 2\epsilon + \|h_n t_n\|_\infty + t_n h_n(x) \ge 2\epsilon.$$

Combining these findings, we obtain that

$$\begin{aligned}&\frac{1}{|t_n|} \int_0^{2\epsilon} \left|(F + t_n h_n)^{-1}(s) - F^{-1}(s)\right| \, ds \\ \le &\frac{1}{|t_n|} \int_a^{F^{-1}\left(2\epsilon + \|h_n t_n\|_\infty\right)} \left|(F + t_n h_n)(x) - F(x)\right| \, dx \\ \le &\|h_n - h\|_{\ell^1(a,b)} + \int_a^{F^{-1}\left(2\epsilon + \|h_n t_n\|_\infty\right)} |h(x)| \, dx.\end{aligned}$$

We realize that the first term goes to zero as $n \to \infty$ by construction. Further, since $t_n \to 0$ for $n \to \infty$, we obtain that $F^{-1}\left(2\epsilon + \|h_n t_n\|_\infty\right) \to F^{-1}(2\epsilon)$. Thus, for $n \to \infty$ the second term can be made arbitrarily small by the choice of ϵ.

For the second term in (B.28), we obtain by a change of variables that

$$\int_0^{2\epsilon} \left|\nabla \phi_{inv_F}(h)\right|(s) \, ds = \int_0^{2\epsilon} \left| \frac{h\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)} \right| \, du = \int_a^{F^{-1}(2\epsilon)} |h(u)| \, du.$$

Thus, this term will be arbitrarily small for ϵ small.

Third summand: The third summand can be treated with the same arguments as the second. □

Based on Lemma B.18, it is straight forward to demonstrate that $\mathbb{U}_n^{-1} \rightsquigarrow \mathbb{G}_1$ as well as $\mathbb{V}_m^{-1} \rightsquigarrow \mathbb{G}_2$ in $\ell^1(0,1)$ (see Section F.2 for more information on weak convergence of empirical U-quantile processes). Reconsidering the proof of Theorem B.17, it becomes clear that in order to adapt the previous arguments we further only require that

$$\sqrt{\frac{nm}{n+m}} \int_0^1 \left(U_n^{-1}(t) - U^{-1}(t)\right)^2 \, dt = o_p(1) \tag{B.29}$$

as well as the analogous result for $\sqrt{\frac{nm}{n+m}}\int_0^1 \left(V_m^{-1}(t)-V^{-1}(t)\right)^2\,dt$. Based on these observations, we can show th following theorem.

Theorem B.19. *Assume Setting 1.1 and let Condition 1.3 hold. Let $DoD_{(0)} \neq 0$ and let $n, m \to \infty$ such that $\frac{n}{m+n} \to \lambda \in (0,1)$. Then, it holds*

$$\sqrt{\frac{nm}{n+m}}\left(\int_0^1 \left(U_n^{-1}(t)-V_m^{-1}(t)\right)^2\,dt - \int_0^1 \left(U^{-1}(t)-V^{-1}(t)\right)^2\,dt\right)$$
$$\rightsquigarrow 2\sqrt{\lambda}\int_0^1 (U^{-1}(t)-V^{-1}(t))\mathbb{G}_1(t)\,dt - 2\sqrt{1-\lambda}\int_0^1 (U^{-1}(t)-V^{-1}(t))\mathbb{G}_2(t)\,dt. \quad \text{(B.30)}$$

Here, $\mathbb{G}_1$ and $\mathbb{G}_2$ denote centered, independent Gaussian processes with covariance structures as defined in Theorem B.17.

Proof of Theorem B.19. Let $\operatorname{supp}(U) = [0, D_1]$. By assumption we have that U is continuously differentiable and that $u(t) = U'(t) > 0$ for all $t \in (0, D_1)$. Furthermore, we have shown in the proof of Lemma B.7 that the class

$$\mathscr{F} = \left\{(x,y) \mapsto \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\leq t\}} : t \in [0, D_1]\right\}$$

meets the requirements of Theorem F.12. Thus, as all its assumptions are given, an application of Theorem F.12 gives that

$$\sqrt{n}(U_n^{-1} - U^{-1}) \rightsquigarrow \mathbb{G}_1$$

in $\ell^1(0,1)$.

It remains to demonstrate (B.29). In the following, let μ_n^U denote the empirical measure corresponding to U_n, i.e., for $A \in \mathscr{B}(\mathbb{R})$

$$\mu_n^U(A) = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} \mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\in A\}}.$$

By Lemma B.9 we have that under Condition 1.3

$$J_2\left(\mu^U\right) = \int_{-\infty}^{\infty} \frac{U(t)(1-U(t))}{u(t)}\,dt < \infty$$

(alternatively this can be assumed directly, see Remark 2.9). Hence, it follows by (B.9) in the proof of Theorem 2.5 (see Section B.3) that

$$\mathbb{E}\left[\int_0^1 \left(U_n^{-1}(t)-U^{-1}(t)\right)^2\,dt\right] = \mathbb{E}\left[\mathcal{K}_2^2\left(\mu_n^U, \mu^U\right)\right] \leq \frac{4}{n+1}J_2(\mu^U),$$

which induces (B.29).

The analogous statements for $\mathbb{V}_m^{-1}$ and $\sqrt{\frac{nm}{n+m}}\int_0^1 \left(V_m^{-1}(t) - V^{-1}(t)\right)^2\, dt$ follow by the same arguments. As already argued, this yields the claim. □

We complete the proof of Theorem 2.7 (ii) by showing that (B.30) is normally distributed.

Proof of Theorem 2.7 (ii). In order to prove that the limit distribution defined in (B.30) is normally distributed, we consider the summands

$$\Psi_1 := 2\sqrt{\lambda}\int_0^1 (U^{-1}(t) - V^{-1}(t))\mathbb{G}_1(t)\, dt$$

and

$$\Psi_2 := 2\sqrt{1-\lambda}\int_0^1 (U^{-1}(t) - V^{-1}(t))\mathbb{G}_2(t)\, dt.$$

separately. First of all, we observe that

$$\Psi_1 = 2\sqrt{\lambda}\int_0^1 (U^{-1}(t) - V^{-1}(t))\mathbb{G}_1(t)\, dt \overset{D}{=} 2\sqrt{\lambda}\int_0^1 (U^{-1}(t) - V^{-1}(t))\frac{\mathbb{G}_3\left(U^{-1}(t)\right)}{u(U^{-1}(t))}\, dt,$$

where $\mathbb{G}_3$ denotes a mean zero Gaussian process with covariance

$$\mathrm{Cov}(\mathbb{G}_3(t), \mathbb{G}_3(t')) = 4\Gamma_{d_{\mathcal{X}}}(t, t').$$

In consequence, a change of variables $(x = U^{-1}(t))$ yields that

$$\Psi_1 \overset{D}{=} 2\sqrt{\lambda}\int_0^{D_{\mathcal{X}}} (x - V^{-1}(U(x)))\mathbb{G}_3\,(t)\; dt,$$

where $\mathrm{supp}\,(U) = [0, D_{\mathcal{X}}]$. It is important to note that metric measure spaces are compact by definition and hence $D_{\mathcal{X}} < \infty$. Further, we observe that the empirical U-process $\mathbb{U}_n = \sqrt{n}\,(U_n - U)$ converges in distribution against $\mathbb{G}_3$, i.e., $\mathbb{U}_n \rightsquigarrow \mathbb{G}_3$ in $\ell^\infty[0, D_{\mathcal{X}}]$. Consequently, we obtain by Corollary F.8 that $\mathbb{G}_3$ is continuous (alternatively this can be easily verified directly via [17, Thm. 7.1]). Further, we realize that $\mathcal{T} = [0, D_{\mathcal{X}}]$ is a compact metric space and hence $\mathbb{G}_1$ is uniformly continuous on $\mathcal{T}$. Since the function $x \mapsto x - V^{-1}(U(x))$ is integrable on $[0, D_{\mathcal{X}}]$, we can again apply Theorem 8.17 from Dümbgen [19] in combination with the arguments used in the proof of Theorem 2.7 (i) to derive that Ψ_1 is normally distributed with mean zero and variance

$$\sigma_1 = 16\lambda \int_0^{D_{\mathcal{X}}}\int_0^{D_{\mathcal{X}}} (x - V^{-1}(U(x)))(y - V^{-1}(U(y)))\Gamma_{d_{\mathcal{X}}}(x, y)\, dxdy.$$

The analogous arguments yield that Ψ_2 is normally distributed with mean zero and variance

$$\sigma_2 = 16(1-\lambda) \int\limits_0^{D_{\mathcal{Y}}} \int\limits_0^{D_{\mathcal{Y}}} (U^{-1}(V(x)) - x))(U^{-1}(V(y)) - y)\Gamma_{d_{\mathcal{Y}}}(x,y)\, dxdy,$$

where $\operatorname{supp}(V) = [0, D_{\mathcal{Y}}]$. Since Ψ_1 and Ψ_2 are independent, this yields the claim.

□

Remark B.20. Although one can try to adopt a similar strategy for the derivation of the limit behavior of

$$\widehat{DoD}_{p,(\beta)}(\mathcal{X}_n, \mathcal{Y}_m) = \int_{\beta}^{1-\beta} \left|U_n^{-1}(t) - V_m^{-1}(t)\right|^p dt$$

under the alternative H_1, $\beta \in [0, 1/2)$, $p \in [1, \infty)$, especially the calculations for the proof of the auxiliary results Theorem B.17 and Theorem B.19 heavily rely on the fact that we have chosen $p = 2$. Thus, Theorem 2.7 is not as easily generalized as Theorem 2.6.

B.6 Proof of Lemma B.12

In this section, we provide a full proof for Lemma B.12. Throughout the following let $\mathfrak{S}(B)$ denote the set of all permutations of the finite set B. In the special case $B = \{1, \dots, n\}$, we write $\mathfrak{S}_n$ instead of $\mathfrak{S}(\{1, \dots, n\})$. We begin this section with the verification of an auxiliary result.

Lemma B.21. *Let $X_1, \dots, X_m, Y_1, \dots, Y_m$, $m \in \mathbb{N}_{\geq 2}$, be independent, identically distributed with differentiable distribution function F and let $(F_m^X)^{-1}$ and $(F_m^Y)^{-1}$ denote the empirical quantile functions of $X_1, \dots, X_m$ and $Y_1, \dots, Y_m$, respectively. Let $\beta \in (0, 1/2)$. Assume that there exist constants $-1 < \gamma_1, \gamma_2 < \infty$ and $c_F > 0$ such that*

$$\left|(F^{-1})'(s)\right| \leq c_F \cdot s^{\gamma_1}(1-s)^{\gamma_2}. \tag{B.31}$$

Then, it holds

$$\mathbb{E}\left[\int_0^{\beta} \left|\sqrt{m}((F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s))\right|^2 ds\right] \leq \frac{C_1}{m+2}\left\{\left(\beta\left(1 + \frac{\sqrt{\log(m)}}{\sqrt{m}}\right)\right)^{2\gamma_1+2} + \frac{C_2}{m^{2\beta^2}}\right\}$$

as well as

$$\mathbb{E}\left[\int_{1-\beta}^{1} \left|\sqrt{m}((F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s))\right|^2 ds\right] \leq \frac{C_1}{m+2}\left\{\left(\beta\left(1 + \frac{\sqrt{\log(m)}}{\sqrt{m}}\right)\right)^{2\gamma_2+2} + \frac{C_2}{m^{2\beta^2}}\right\},$$

where C_1, C_2 denote finite positive constants that are independent of β.

Proof of Lemma B.21. In the following, we only prove the first inequality. The second inequality follows by employing the analogous arguments.

Recall that

$$(F_m^X)^{-1}(s) = X_{(k)} \quad \text{for} \quad \frac{k-1}{m} < s \leq \frac{k}{m},\ k = 1, \dots, m.$$

Hence, we have that

$$\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right] \leq \mathbb{E}\left[\frac{1}{m}\sum_{k=1}^{\lceil m\beta\rceil} |X_{(k)} - Y_{(k)}|^2\right].$$

Let $U_1, \dots, U_m$ and $V_1, \dots, V_m$ be two independent i.i.d. $U[0,1]$ samples. Then

$$\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right] \leq \mathbb{E}\left[\frac{1}{m}\sum_{k=1}^{\lceil m\beta\rceil} |F^{-1}(U_{(k)}) - F^{-1}(V_{(k)})|^2\right].$$

Since $U_{(k)} \overset{\mathcal{D}}{=} V_{(k)} \overset{\mathcal{D}}{=} \mathrm{Beta}(k, m-k+1)$ and $U_{(k)}$ and $V_{(k)}$ are independent, we may write

$$\begin{aligned}&\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right]\\ &\leq \frac{1}{m}\sum_{k=1}^{\lceil m\beta\rceil} \int_0^1\int_0^1 |F^{-1}(s) - F^{-1}(t)|^2\, dB_{k,m-k+1}(s)\, dB_{k,m-k+1}(t).\end{aligned}$$

An application of Proposition B.8 in Bobkov and Ledoux (2016) yields

$$\begin{aligned}&\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right]\\ &\leq \left(\frac{10}{\sqrt{m+2}}\right)^2 \frac{1}{m}\sum_{k=1}^{\lceil m\beta\rceil} \int_0^1 x(1-x)|(F^{-1})'(x)|^2\, dB_{k,m-k+1}(x).\end{aligned}$$

Next, we use that

$$dB_{k,m-k+1}(x) = m\binom{m-1}{k-1} x^{k-1}(1-x)^{m-k}\, dx$$

and obtain

$$\begin{aligned}&\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right]\\ &\leq \frac{100}{m+2}\sum_{k=1}^{\lceil m\beta\rceil} \binom{m-1}{k-1} \int_0^1 x(1-x)|(F^{-1})'(x)|^2 x^{k-1}(1-x)^{m-k}\, dx.\end{aligned}$$

In a further step we split the integral into two parts, one for small $x \lesssim \beta$ and one for the remaining larger values of x. The maximum of the function $x \mapsto x^k(1-x)^{m-k+1}$ is attained at $k/(m+1)$, where

$$\frac{k}{m+1} \leq \frac{\lceil \beta m \rceil}{m+1} \leq \frac{\beta m + 1}{m+1} \leq \beta + \frac{1-\beta}{m+1}.$$

In particular, for small values $k \ll \lceil \beta m \rceil$, integration over $x \lesssim \beta$ contains most of the "mass" of the function $x \mapsto x^k(1-x)^{m-k+1}$, whereas the "mass" for $x \gtrsim \beta$ becomes negligible for fixed $\beta > 0$ as $m \to \infty$.

We resort to the following two observations:

$$\sum_{k=1}^{\lceil m\beta \rceil} \binom{m-1}{k-1} x^{k-1}(1-x)^{m-k} \leq 1 \quad \text{for all } x \in [0,1], \tag{B.32}$$

by the binomial formula and, for $x > \frac{m\beta}{m-1}\left(1 + \sqrt{\frac{\log(m)}{m}}\right)$,

$$\mathbb{P}\left(\mathrm{Bin}(m-1,x) \leq \lceil m\beta \rceil - 1\right) \leq \exp\left(-2\frac{(m\beta(1+\sqrt{\log(m)/m}) - m\beta)^2}{m}\right) = m^{-2\beta^2}, \tag{B.33}$$

by Hoeffding's inequality. This gives the estimate

$$\begin{aligned}
&\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s) - (F_m^Y)^{-1}(s)|^2\, ds\right] \\
&\leq \tfrac{100}{m+2} \int_0^{\frac{\beta m}{m-1}\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)} x(1-x)|(F^{-1})'(x)|^2\, dx \\
&+ \tfrac{100}{m+2} \int_{\frac{\beta m}{m-1}\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)}^1 x(1-x)|(F^{-1})'(x)|^2 \cdot \mathbb{P}\left(\mathrm{Bin}(m-1,x) \leq \lceil m\beta \rceil - 1\right) dx \\
&\leq \tfrac{100}{m+2} \left\{ \int_0^{\frac{\beta m}{m-1}\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)} x(1-x)|(F^{-1})'(x)|^2\, dx + \frac{1}{m^{2\beta^2}} \int_0^1 x(1-x)|(F^{-1})'(x)|^2\, dx \right\}.
\end{aligned}$$

In the remainder of this proof C_1 and C_2 denote finite, positive constants that are independent of β and may vary from line to line. Using (B.31) we obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^\beta |(F_m^X)^{-1}(s)-(F_m^Y)^{-1}(s)|^2\,ds\right]\\
\leq&\frac{100}{m+2}\left\{\int_0^{\frac{\beta m}{m-1}\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)} c_F x^{2\gamma_1+1}(1-x)^{2\gamma_2+1}\,dx+\frac{1}{m^{2\beta^2}}\int_0^1 c_F x^{2\gamma_1+1}(1-x)^{2\gamma_2+1}\,dx\right\}\\
\leq&\frac{100}{m+2}\left\{\int_0^{\frac{\beta m}{m-1}\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)} C_1 x^{2\gamma_1+1}\,dx+\frac{C_2}{m^{2\beta^2}}\right\}\\
=&\frac{C_1}{m+2}\left\{\left(\beta\left(1+\frac{\sqrt{\log(m)}}{\sqrt{m}}\right)\right)^{2\gamma_1+2}+\frac{C_2}{m^{2\beta^2}}\right\},
\end{aligned}
$$

where we used that $2\gamma_i+1>-1$, $i\in\{1,2\}$, by assumption and that $m/(m-1)\leq 2$ for $m\in\mathbb{N}_{\geq 2}$. Thus, we have proven the claim. □

With Lemma B.21 at our disposal and the ideas developed for proving Lemma B.3 (see Section B.3), we can start the proof of Lemma B.12.

Proof of Lemma B.12. Here, we only derive (B.15). The statement (B.16) follows by the analogue arguments.

Let $X_1,\dots X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and $Y_1,\dots,Y_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Let $m=\left\lceil\frac{\beta n(n-1)}{2}\right\rceil$. Further, denote by $\left\{d_{(i)}^{\mathcal{X}}\right\}_{i=1}^{n(n-1)/2}$ the ordered sample of $\{d_{\mathcal{X}}(X_i,X_j)\}_{1\leq i<j\leq n}$ and $\left\{d_{(i)}^{\mathcal{Y}}\right\}_{i=1}^{n(n-1)/2}$ the ordered sample of $\{d_{\mathcal{Y}}(Y_i,Y_j)\}_{1\leq i<j\leq n}$. We have by Lemma 4.1 of Bobkov and Ledoux [8] that

$$
\begin{aligned}
\mathbb{E}\left[\int_0^\beta \left|U_n^{-1}(t)-V_n^{-1}(t)\right|^2\,dt\right]&\leq\mathbb{E}\left[\sum_{i=1}^m\frac{2}{n(n-1)}\left|d_{(i)}^{\mathcal{X}}-d_{(i)}^{\mathcal{Y}}\right|^2\right]\\
&=\frac{2}{n(n-1)}\mathbb{E}\left[\inf_{\sigma\in\mathfrak{S}_m}\sum_{i=1}^m\left|d_i^U-d_{\sigma(i)}^V\right|^2\right],
\end{aligned}
\tag{B.34}
$$

where $\{d_i^U\}_{i=1}^m$ denotes the unordered sample of $\left\{d_{(i)}^{\mathcal{X}}\right\}_{i=1}^m$ and $\{d_i^V\}_{i=1}^m$ is defined analogously. In order to further bound (B.34), we want to make use of similar ideas as in the proof of Theorem B.5. Thus, we once again divide the random variables $\{d_i^U\}_{i=1}^m$ and $\{d_i^V\}_{i=1}^m$ into groups of independent random variables which are easier to handle. Let

$$
k=\max\left\{l\in\{1,\dots,n\}: m\geq\sum_{i=1}^l(n-i)\right\}
$$

and define

$$k^* = m - \sum_{i=1}^{k}(n-i).$$

By the definition of k follows that

$$\frac{\beta(n-1)}{2} \overset{(i)}{\leq} k \overset{(ii)}{\leq} \beta n.$$

Concerning (i): We have by definition that

$$\sum_{i=1}^{k}(n-i) \leq m \Leftrightarrow kn - \frac{k(k+1)}{2} \leq m$$

Clearly, it follows for $k \geq \beta n$ that

$$\begin{aligned} kn - \frac{k(k+1)}{2} \geq& \beta n^2 - \frac{\beta n(\beta n+1)}{2} = \frac{\beta n(n-1)}{2} + \frac{\beta n^2}{2}(1-\beta) \\ \overset{(*)}{\geq}& \frac{\beta n(n-1)}{2} + \frac{5\beta n}{12} \overset{(**)}{>} \frac{\beta n(n-1)}{2} + 1 \geq m. \end{aligned}$$

Here, $(*)$ follows as $\beta < 1/6$ and $(**)$ holds as $n\beta > 8$.

Concerning (ii): We have that $k \geq \beta(n-1)/2$, since

$$\frac{\beta n(n-1)}{2} - \frac{\frac{\beta(n-1)}{2}\left(\frac{\beta(n-1)}{2}+1\right)}{2} \leq \frac{\beta n(n-1)}{2} \leq \left\lceil \frac{\beta n(n-1)}{2} \right\rceil = m.$$

Additionally, we note that the assumptions $n\beta > 8$ and $\beta < 1/6$ guarantee that $k > 3$ and that

$$n - k \geq n - \beta n = (1-\beta)n > \beta n \geq k.$$

In order to prove the claim, we further have to distinguish the cases k even and k odd.

Case 1: Let k be odd. We show that we once again can divide $\left\{d_i^U\right\}_{i=1}^m$ into n groups of independent variables. To this end, we consider a worst case scenario, in which the number of dependencies between the considered number of random variables is maximized. Such a worst case scenario is given by

$$\begin{aligned} \left\{d_i^U\right\}_{i=1}^m = \big\{ & d_{\mathcal{X}}(X_1,X_2),\ldots,d_{\mathcal{X}}(X_1,X_n),d_{\mathcal{X}}(X_2,X_3),\ldots,d_{\mathcal{X}}(X_2,X_n),\ldots,d_{\mathcal{X}}(X_k,X_n), \\ & d_{\mathcal{X}}(X_{k+1},X_{k+2}),\ldots,d_{\mathcal{X}}(X_{k+1},X_{k+1+k^*})\big\}. \end{aligned}$$

We divide $\left\{d_i^U\right\}_{i=1}^m$ into two blocks

$$\{d_{\mathcal{X}}(X_i,X_j)\}_{1\leq i<j\leq k} \cup \{d_{\mathcal{X}}(X_{k+1},X_j)\}_{k+2\leq j\leq k+1+k^*}$$

and

$$\{d_{\mathcal{X}}(X_i, X_j)\}_{1\le i\le k, k+1\le j\le n}.$$

As $k+1>3$ is even, it follows by Lemma G.2 that the first block can be partitioned into k blocks of size $(k+1)/2$ or $(k+1)/2-1$.

Due to its simple dependency structure the second block can be split into $\max\{n-k,k\} = n-k$ groups of size $\min\{n-k,k\}=k$.

Example B.22. Let $n=6$ and $k=2$. Then, we have

$$\begin{array}{ccc|ccc} d_{\mathcal{X}}(X_1,X_4) & d_{\mathcal{X}}(X_1,X_5) & d_{\mathcal{X}}(X_1,X_6) & d_{\mathcal{X}}(X_1,X_4) & d_{\mathcal{X}}(X_1,X_5) & d_{\mathcal{X}}(X_1,X_6)\\ d_{\mathcal{X}}(X_2,X_4) & d_{\mathcal{X}}(X_2,X_5) & d_{\mathcal{X}}(X_2,X_6) & d_{\mathcal{X}}(X_2,X_4) & d_{\mathcal{X}}(X_2,X_5) & d_{\mathcal{X}}(X_2,X_6)\\ d_{\mathcal{X}}(X_3,X_4) & d_{\mathcal{X}}(X_3,X_5) & d_{\mathcal{X}}(X_3,X_6) & d_{\mathcal{X}}(X_3,X_4) & d_{\mathcal{X}}(X_3,X_5) & d_{\mathcal{X}}(X_3,X_6). \end{array}$$

We stress that this example is only used to highlight the simple dependency structure of $\{d_{\mathcal{X}}(X_i, X_j)\}_{1\le i\le k, k+1\le j\le n}$.

To summarize, if k is odd, we can divide $\{d_i^U\}_{i=1}^m$ in the worst case scenario into n groups of $(k+1)/2, (k+1)/2-1$ or k independent random variables. Thus, the same statement holds true for any other scenario.

Case 2: Let k be even. We demonstrate that also in this case we can split $\{d_i^U\}_{i=1}^m$ into n groups of independent variables. We consider the same worst case scenario as for the previous case, i.e.,

$$\{d_i^U\}_{i=1}^m = \{d_{\mathcal{X}}(X_1,X_2),\dots,d_{\mathcal{X}}(X_1,X_n),d_{\mathcal{X}}(X_2,X_3),\dots,d_{\mathcal{X}}(X_2,X_n),\dots,d_{\mathcal{X}}(X_k,X_n), \\ d_{\mathcal{X}}(X_{k+1},X_{k+2}),\dots,d_{\mathcal{X}}(X_{k+1},X_{k+1+k^*})\}$$

and apply similar arguments. As previously, we divide $\{d_i^U\}_{i=1}^m$ into the two blocks

$$\{d_{\mathcal{X}}(X_i,X_j)\}_{1\le i<j\le k} \cup \{d_{\mathcal{X}}(X_{k+1},X_j)\}_{k+2\le j\le k+1+k^*} \tag{B.35}$$

and

$$\{d_{\mathcal{X}}(X_i, X_j)\}_{1\le i\le k, k+1\le j\le n}.$$

As $k+1>3$ is odd it follows by Lemma G.2 that the first block can be partitioned into $k+1$ groups of $k/2$ or $k/2-1$ independent random variables and the same arguments as for Case 1 yield that the second block can be split into $n-k$ blocks of independent random variables of size k. This so far yields $n+1$ groups of independent random variables. However, the first part of the proof of Lemma G.2 indicates that we can pick one group of the partition of $\{d_{\mathcal{X}}(X_i, X_j)\}_{1\le i\le k, k+1\le j\le n}$ and split it over the $k+1$ groups that form (B.35) in such a way that the resulting groups still consist of independent random variables.
In summary, we find that we can partition $\{d_i^U\}_{i=1}^m$ into n groups of at least $k/2-1$ and at most k elements.

By the previous arguments there exist partitions $(P_l)_{l=1}^n$ of $\{d_i^U\}_{i=1}^m$ and $(Q_l)_{l=1}^n$ of $\{d_i^V\}_{i=1}^m$ consisting of independent random variables with $|P_l| = |Q_l| =: p_l$ for $1 \leq l \leq n$.

Let $\left\{d_i^{U,P_l}\right\}_{i=1}^{p_l}$ denote the elements of P_l, $1 \leq l \leq n$ and let $\left\{d_i^{V,Q_l}\right\}_{i=1}^{p_l}$ denote the ones of Q_l. By the first step of the proof of Lemma B.4 it follows

$$\frac{2}{n(n-1)}\mathbb{E}\left[\inf_{\sigma\in\mathfrak{S}_m}\sum_{i=1}^m\left(d_i^U - d_{\sigma(i)}^V\right)^2\right] \leq \frac{2}{n(n-1)}\mathbb{E}\left[\sum_{l=1}^n\inf_{\sigma\in\mathfrak{S}_{p_l}}\sum_{i=1}^{p_l}\left(d_i^{U,P_l} - d_{\sigma(i)}^{V,Q_l}\right)^2\right]$$

With this we get using Lemma 4.1 of Bobkov and Ledoux [8] that

$$\begin{aligned}
&\frac{2}{n(n-1)}\mathbb{E}\left[\sum_{l=1}^n\inf_{\sigma\in\mathfrak{S}_{p_l}}\sum_{i=1}^{p_l}\left(d_i^{U,P_l} - d_{\sigma(i)}^{V,Q_l}\right)^2\right]\\
&= \frac{2}{n(n-1)}\mathbb{E}\left[\sum_{l=1}^n\sum_{i=1}^{p_l}\left(d_{(i)}^{U,P_l} - d_{(i)}^{V,Q_l}\right)^2\right]\\
&= \frac{2}{n(n-1)}\sum_{l=1}^n\mathbb{E}\left[\sum_{i=1}^{p_l}\left(d_{(i)}^{U,P_l} - d_{(i)}^{V,Q_l}\right)^2\right]
\end{aligned}$$

Next, we connect the above expectations to a difference of two empirical quantile functions. We once again distinguish the cases k odd and k even.

Case 1: Let k be odd. For $1 \leq l \leq n$, we have that

$$(k+1)/2 - 1 \leq p_l \leq k \text{ and } \frac{\beta(n-1)}{2} \leq k \leq \beta n.$$

Additionally, as $m = \left\lceil\frac{\beta n(n-1)}{2}\right\rceil$, there are $\left\lfloor\frac{(1-\beta)n(n-1)}{2}\right\rfloor$ distances that are greater than all elements in P_l. Of these at least

$$\begin{aligned}
\left\lfloor\frac{(1-\beta)n(n-1)}{2}\right\rfloor - np_l \geq& \frac{(1-\beta)n(n-1)}{2} - 1 - \beta n^2\\
=& \frac{(1-3\beta)n(n-2)}{2} + \frac{(1-5\beta)n}{2} - 1\\
\overset{(iii)}{\geq}& \frac{(1-3\beta)n(n-2)}{2}
\end{aligned}$$

are independent of the elements of P_l. Here, (iii) holds as $\beta < 1/6$ and $n \geq 100$. In consequence, we can find at least $\left\lfloor\frac{(1-3\beta)(n-2)}{2}\right\rfloor$ more random variables in the set $\{d_i^U\}_{i=m+1}^{n(n-1)/2}$ that are independent of the ones in P_l. Thus, we can consider P_l as the first p_l order statistics of a sample with

$$\frac{(1-2.5\beta)n}{2} - \frac{5-5.5\beta}{2} \leq p_l + \left\lfloor\frac{(1-3\beta)(n-2)}{2}\right\rfloor \leq \frac{(1-\beta)n}{2}$$

elements. Hence, we find that there exists $C_{l_1} = C_{l_1}(n, \beta)$ such that

$$C_{l_1} n = p_l + \left\lfloor \frac{(1-3\beta)(n-2)}{2} \right\rfloor .$$

with

$$\frac{4}{15} \leq \frac{7}{24} - \frac{5}{200} \leq \frac{1-2.5\beta}{2} - \frac{5-5.5\beta}{2n} \leq C_{l_1} \leq \frac{1-\beta}{2} \leq \frac{1}{2}.$$

The analogous statement holds for Q_l. Consequently, we obtain for $1 \leq l \leq n$ that

$$\mathbb{E}\left[\sum_{i=1}^{p_l}\left(d_{(i)}^{U,P_l} - d_{(i)}^{V,Q_l}\right)^2\right] = C_{l_1} n \mathbb{E}\left[\int_0^{\frac{p_l}{C_{l_1} n}} \left(\left(F_{C_{l_1} n}^{P_l}\right)^{-1}(t) - \left(F_{C_{l_1} n}^{Q_l}\right)^{-1}(t)\right)^2 dt\right],$$

where $\left(F_{C_{l_1} n}^{P_l}\right)^{-1}$ denotes an empirical quantile function based on an i.i.d. sample of U of size $C_{l_1} n$ and $\left(F_{C_{l_1} n}^{Q_l}\right)^{-1}$ is defined analogously. Using the previously derived bounds for C_{l_1} and p_l, we obtain

$$\begin{aligned} & C_{l_1} n \mathbb{E}\left[\int_0^{\frac{p_l}{C_{l_1} n}} \left(\left(F_{C_{l_1} n}^{P_l}\right)^{-1}(t) - \left(F_{C_{l_1} n}^{Q_l}\right)^{-1}(t)\right)^2 dt\right] \\ \leq & \frac{n}{2} \mathbb{E}\left[\int_0^{\frac{15\beta}{4}} \left(\left(F_{C_{l_1} n}^{P_l}\right)^{-1}(t) - \left(F_{C_{l_1} n}^{Q_l}\right)^{-1}(t)\right)^2 dt\right]. \end{aligned}$$

Applying now Lemma B.21 to the above expectation gives that

$$\begin{aligned} & \frac{n}{2} \mathbb{E}\left[\int_0^{\frac{15\beta}{4}} \left(\left(F_{C_{l_1} n}^{P_l}\right)^{-1}(t) - \left(F_{C_{l_1} n}^{Q_l}\right)^{-1}(t)\right)^2 dt\right] \\ & \leq C_1 \left(\frac{15}{4}\beta\left(1 + \frac{\sqrt{15\log(n)}}{2\sqrt{n}}\right)\right)^{2\gamma_1+2} + o(1), \end{aligned}$$

where C_1 denotes a finite constant that is independent of β.

In consequence, we obtain that

$$\begin{aligned} \mathbb{E}\left[\int_0^{\beta} \left(U_n^{-1}(t) - V_n^{-1}(t)\right)^2 dt\right] \leq & \frac{2}{n(n-1)} \mathbb{E}\left[\sum_{l=1}^{n} \inf_{\sigma \in \mathfrak{S}_{p_l}} \sum_{i=1}^{p_l} \left(d_i^{U,P_l} - d_{\sigma(i)}^{V,Q_l}\right)^2\right] \\ \leq & \frac{2}{n(n-1)} \sum_{l=1}^{n} \mathbb{E}\left[\sum_{i=1}^{p_l}\left(d_{(i)}^{U,P_l} - d_{(i)}^{V,Q_l}\right)^2\right] \\ \leq & \frac{2C_1}{n-1}\left(\frac{15}{4}\beta\left(1 + \frac{\sqrt{15\log(n)}}{2\sqrt{n}}\right)\right)^{2\gamma_1+2} + o\left(n^{-1}\right). \end{aligned}$$

Case 2: Let k be even. Then, we obtain with the same arguments as previously that

$$\mathbb{E}\left[\sum_{i=1}^{p_l}\left(d_{(i)}^{U,P_l}-d_{(i)}^{V,Q_l}\right)^2\right]=C_{l_2}n\mathbb{E}\left[\int_0^{\frac{p_l}{C_{l_2}n}}\left(\left(F_{C_{l_2}n}^{P_l}\right)^{-1}(t)-\left(F_{C_{l_2}n}^{Q_l}\right)^{-1}(t)\right)^2\,dt\right],$$

where $\left(F_{C_{l_2}n}^{P_l}\right)^{-1}$ and $\left(F_{C_{l_2}n}^{Q_l}\right)^{-1}$ denote empirical quantile functions based on an i.i.d. sample of size $C_{l_2}n$. Here, it holds for $C_{l_2}=C_{l_2}(n,\beta)$ that

$$\frac{1}{4}\leq\frac{7}{24}-\frac{6}{200}\leq\frac{1-2.5\beta}{2}-\frac{6-5.5\beta}{2n}\leq C_{l_2}\leq\frac{1-\beta}{2}\leq\frac{1}{2}.$$

Thus, we find that

$$\begin{aligned}&C_{l_2}n\mathbb{E}\left[\int_0^{\frac{p_l}{C_{l_2}n}}\left(\left(F_{C_{l_2}n}^{P_l}\right)^{-1}(t)-\left(F_{C_{l_2}n}^{Q_l}\right)^{-1}(t)\right)^2\,dt\right]\\ &\leq C_{l_2}n\mathbb{E}\left[\int_0^{4\beta}\left(\left(F_{C_{l_2}n}^{P_l}\right)^{-1}(t)-\left(F_{C_{l_2}n}^{Q_l}\right)^{-1}(t)\right)^2\,dt\right].\end{aligned}$$

Just as in the previous case an application of Lemma B.21 in combination with the derived estimates yields that also in this case

$$\mathbb{E}\left[\int_0^{\beta}\left(U_n^{-1}(t)-V_n^{-1}(t)\right)^2\,dt\right]\leq\frac{2C_1}{n-1}\left(4\beta\left(1+2\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)\right)^{2\gamma_1+2}+o\left(n^{-1}\right),$$

where C_1 denotes a finite constant that is independent of β. □

Remark B.23. As mentioned previously, it is straight forward to extend Lemma B.12 to the case $n\neq m$. In the present setting, we have $\mu^U=\mu^V$, i.e., $U^{-1}=V^{-1}$. Hence, we can employ similar arguments as in the proof of Theorem 2.5 (cf. Section B.3) to show that

$$\begin{aligned}\mathbb{E}\left[\int_0^{\beta}\left(U_n^{-1}(t)-V_m^{-1}(t)\right)^2\,dt\right]\leq&2\Bigg(\mathbb{E}\left[\int_0^{\beta}\left(U_n^{-1}(t)-(U_n')^{-1}(t)\right)^2\,dt\right]\\ &+\mathbb{E}\left[\int_0^{\beta}\left(V_m^{-1}(t)-(V_m')^{-1}(t)\right)^2\,dt\right]\Bigg),\end{aligned}$$

where $(U_n')^{-1}$ and $(V_m')^{-1}$ denote independent copies of U_n^{-1} and V_m^{-1}, respectively. Both summands can now be bounded by an application of Lemma B.12.

B.7 The DTM-Signature

As already illustrated in Section 5.1 of the main document, Brécheteau [9] has proposed a statistical test, denoted as Φ_{DTM}, for the comparison of two metric measure spaces $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}}, \mu_{\mathcal{Y}})$. While our test Φ^*_{DoD} is based on an empirical version of the statistic $DoD_p(\mathcal{X}, \mathcal{Y})$ defined in (2) of the paper, the test Φ_{DTM} is based on an empirical version of a quantity $T_\kappa(\mathcal{X}, \mathcal{Y})$, which will be defined below. Both $DoD_p(\mathcal{X}, \mathcal{Y})$ and $T_\kappa(\mathcal{X}, \mathcal{Y})$ are different lower bounds of the Gromov-Wasserstein distance, as we will elaborate in the following.

We start by defining the *DTM-signature* (as done in Brécheteau [9]). For $x \in \mathcal{X}$ let $B_{\mathcal{X}}(x, r) = \{x' \in \mathcal{X} \,|\, d_{\mathcal{X}}(x, x') \le r\}$ and define

$$F_x(t) := \mu_{\mathcal{X}}(B_{\mathcal{X}}(x, t)) = \mathbb{P}\left(d_{\mathcal{X}}(x, X_1) \le t\right),$$

where $X_1 \sim \mu_{\mathcal{X}}$. Let F_x^{-1} denote the quantile function of F_x. Then, the *DTM-function* with mass parameter $\kappa \in [0, 1]$ associated to $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ is defined as

$$\delta_{\mathcal{X},\kappa}(x) := \frac{1}{\kappa} \int_{l=0}^{\kappa} \inf\left\{r > 0 \,|\, \mu_{\mathcal{X}}\left(B_{\mathcal{X}}(x, r)\right) > l\right\} dl = \frac{1}{\kappa} \int_0^{\kappa} F_x^{-1}(l)\, dl.$$

Further, the *DTM-signature* of $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ is given as

$$D_{\mathcal{X},\kappa} := \delta_{\mathcal{X},\kappa}(X_2),$$

where $X_2 \sim \mu_{\mathcal{X}}$. It is important to note that $D_{\mathcal{X},\kappa}$ is a real valued random variable. Let G_y, G_y^{-1}, $\delta_{\mathcal{Y},\kappa}(y)$, $y \in \mathcal{Y}$, and $D_{\mathcal{Y},\kappa}$ be defined analogously. Brécheteau [9] demonstrates that

$$T_\kappa(\mathcal{X}, \mathcal{Y}) := \mathcal{K}_1(D_{\mathcal{X},\kappa}, D_{\mathcal{Y},\kappa}) \le \frac{2}{m} \mathcal{GW}_1\left(\mathcal{X}, \mathcal{Y}\right),$$

where $\mathcal{K}_1$ denotes the Kantorovich (transport) distance of order one (see e.g. [42, Def. 6.1] for a formal definition). The test Φ_{DTM} is constructed based on a (subsampled) empirical version of $T_\kappa(\mathcal{X}, \mathcal{Y})$.

As the underlying signatures strongly influence the behavior of the corresponding test, we will investigate how the signatures $T_\kappa(\mathcal{X}, \mathcal{Y})$ and $DoD_p(\mathcal{X}, \mathcal{Y})$ relate to each other. By the definition of $D_{\mathcal{X},\kappa}$, we see that $T_\kappa(\mathcal{X}, \mathcal{Y})$ is sensitive to local changes. Hence, $T_\kappa(\mathcal{X}, \mathcal{Y})$ discriminates between the metric measure spaces in Figure 7 of Mémoli [27] for $\kappa > 1/4$, whereas $DoD_p\left(\mathcal{X}, \mathcal{Y}\right)$ is always zero for this example. On the other hand, for $\kappa \le 1/2$, $T_\kappa(\mathcal{X}, \mathcal{Y})$ cannot distinguish between the metric measure spaces displayed in Figure B.2 above, whereas $DoD_p(\mathcal{X}, \mathcal{Y})$ can become arbitrarily large, if the represented triangles are moved further apart. More precisely, DoD_p scales as the p'th power of the distance between the triangles.

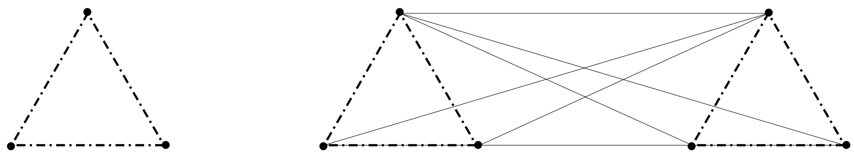

Fig. B.2: Different metric measure spaces: Representation of two different, discrete metric measure spaces that are both equipped with the respective uniform distribution. Left: Three points with the same pairwise distances (dash-dotted lines). Right: Two translated copies that are further than one side length apart.

B.7.1 Connections to other Lower Bounds

As it puts the lower bound $T_\kappa(\mathcal{X},\mathcal{Y})$ into a broader perspective, we will relate it to other known lower bounds of the Gromov-Wasserstein distance in the remainder of this section. For this purpose, we first show that, just like $DoD_1(\mathcal{X},\mathcal{Y})$ [28, Sec. 2.2], $\frac{\kappa}{2}T_\kappa(\mathcal{X},\mathcal{Y})$ is a lower bound for $\mathbf{TLB}_1(\mathcal{X},\mathcal{Y})$ (defined in the subsequent lemma), which is itself a lower bound of the Gromov-Wasserstein distance [27, Sec. 6].

Lemma B.24. *Let $(\mathcal{X},d_{\mathcal{X}},\mu_{\mathcal{X}})$ and $(\mathcal{Y},d_{\mathcal{Y}},\mu_{\mathcal{Y}})$ be two metric measure spaces. Then, it holds for $\kappa\in[0,1]$ that*

$$T_\kappa(\mathcal{X},\mathcal{Y})\le\frac{2}{\kappa}\mathbf{TLB}_1(\mathcal{X},\mathcal{Y}):=\frac{1}{\kappa}\inf_{\pi\in\mathcal{M}(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})}\int_{\mathcal{X}\times\mathcal{Y}}\Omega_1(x,y)\,d\pi(x,y),\tag{B.36}$$

where

$$\Omega_1(x,y)=\inf_{\pi'\in\mathcal{M}(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})}\int_{\mathcal{X}\times\mathcal{Y}}\left|d_{\mathcal{X}}(x,x')-d_{\mathcal{Y}}(y,y')\right|\,d\pi'(x',y').$$

Proof. For any $\pi\in\mathcal{M}(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})$ it holds that

$$\begin{aligned}
&\mathcal{K}_1(D_{\mathcal{X},\kappa},D_{\mathcal{Y},\kappa})\\
\le&\int_{\mathcal{X}\times\mathcal{Y}}\left|\delta_{\mathcal{X},\kappa}(x)-\delta_{\mathcal{Y},\kappa}(y)\right|\,d\pi(x,y)\\
\le&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\kappa\left|\inf\left\{r>0\,|\,\mu_{\mathcal{X}}\left(B_{\mathcal{X}}(x,r)\right)>l\right\}-\inf\left\{r>0\,|\,\mu_{\mathcal{Y}}\left(B_{\mathcal{Y}}(y,r)\right)>l\right\}\right|\,dl\,d\pi(x,y)\\
=&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\kappa\left|\int_0^\infty\left(\mathbb{1}_{\{\mu_{\mathcal{X}}(B_{\mathcal{X}}(x,r))\le l\}}-\mathbb{1}_{\{\mu_{\mathcal{Y}}(B_{\mathcal{Y}}(y,r))\le l\}}\right)\,dr\right|\,dl\,d\pi(x,y)\\
\le&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\infty\int_0^\kappa\left|\mathbb{1}_{\{\mu_{\mathcal{X}}(B_{\mathcal{X}}(x,r))\le l\}}-\mathbb{1}_{\{\mu_{\mathcal{Y}}(B_{\mathcal{Y}}(y,r))\le l\}}\right|\,dl\,dr\,d\pi(x,y)\\
\le&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\infty\left|\mu_{\mathcal{X}}\left(B_{\mathcal{X}}(x,r)\right)\wedge\kappa-\mu_{\mathcal{Y}}\left(B_{\mathcal{Y}}(y,r)\right)\wedge\kappa\right|\,dr\,d\pi(x,y)\\
\le&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\infty\left|\int_{\mathcal{X}\times\mathcal{Y}}\left(\mathbb{1}_{\{d_{\mathcal{X}}(x,x')\le r\}}-\mathbb{1}_{\{d_{\mathcal{Y}}(y,y')\le r\}}\right)\,d\pi^*(x',y')\right|\wedge\kappa\,dr\,d\pi(x,y),
\end{aligned}$$

where

$$\pi^* \in \operatorname*{arg\,min}_{\pi' \in \mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})} \int_{\mathcal{X}\times\mathcal{Y}} \left|d_\mathcal{X}(x,x') - d_\mathcal{Y}(y,y')\right| \, d\pi'(x',y').$$

The minimum is attained, as the set $\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})$ is compact [41]. With this, we obtain

$$\begin{aligned}
&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\infty \left|\int_{\mathcal{X}\times\mathcal{Y}} \left(\mathbb{1}_{\{d_\mathcal{X}(x,x')\le r\}} - \mathbb{1}_{\{d_\mathcal{Y}(y,y')\le r\}}\right) \, d\pi^*(x',y')\right| \wedge \kappa \, dr \, d\pi(x,y)\\
\le&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^\infty \left|\mathbb{1}_{\{d_\mathcal{X}(x,x')\le r\}} - \mathbb{1}_{\{d_\mathcal{Y}(y,y')\le r\}}\right| \, dr \, d\pi^*(x',y') \, d\pi(x,y)\\
=&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\left(\inf_{\pi'\in\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})}\int_{\mathcal{X}\times\mathcal{Y}}\left|d_\mathcal{X}(x,x') - d_\mathcal{Y}(y,y')\right| \, d\pi'(x',y')\right) \, d\pi(x,y).
\end{aligned}$$

This gives the claim. □

However, $T_\kappa(\mathcal{X},\mathcal{Y})$ is not only a lower bound of $\mathbf{TLB}_1(\mathcal{X},\mathcal{Y})$, but also (different from $DoD_1(\mathcal{X},\mathcal{Y})$) strongly connected to the lower bound $\mathbf{FLB}_1(\mathcal{X},\mathcal{Y})$, defined in the following. Let

$$s_\mathcal{X} : \mathcal{X} \to \mathbb{R}_+, \quad x \mapsto ||d_\mathcal{X}(x,\cdot)||_{\ell^1(\mu_\mathcal{X})} = \int_\mathcal{X} d_\mathcal{X}(x,x') \, d\mu_\mathcal{X}(x')$$

and define $S_\mathcal{X} : \mathbb{R} \to [0,1]$ by $t \mapsto \mu_\mathcal{X}(\{x \in \mathcal{X} \,|\, s_\mathcal{X}(x) \le t\})$. Let $s_\mathcal{Y}$ and $S_\mathcal{Y}$ be defined analogously. Then,

$$\mathbf{FLB}_1(\mathcal{X},\mathcal{Y}) = \frac{1}{2}\inf_{\pi\in\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})}\int_{\mathcal{X}\times\mathcal{Y}} |s_\mathcal{X}(x) - s_\mathcal{Y}(y)| \, d\pi(x,y) = \frac{1}{2}\int_\mathbb{R} |S_\mathcal{X}(t) - S_\mathcal{Y}(t)| \, dt$$

defines a lower bound of $\mathbf{TLB}_1(\mathcal{X},\mathcal{Y})$, i.e., in particular of $\mathcal{GW}_1(\mathcal{X},\mathcal{Y})$ (see Mémoli [27, Rem. 6.5] and Chowdhury and Mémoli [14, Thm. 3.1]).

Lemma B.25. *Let $(\mathcal{X},d_\mathcal{X},\mu_\mathcal{X})$ and $(\mathcal{Y},d_\mathcal{Y},\mu_\mathcal{Y})$ be two metric measure spaces. Then, it holds for $\kappa \in [0,1]$ that*

$$\begin{aligned}
T_\kappa(\mathcal{X},\mathcal{Y}) \le&\frac{1}{\kappa}\inf_{\pi\in\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})}\int_{\mathcal{X}\times\mathcal{Y}}\left|s_\mathcal{X}^{(F_x^{-1}(\kappa))}(x) - s_\mathcal{Y}^{(F_x^{-1}(\kappa))}(y)\right| \, d\pi(x,y) &\text{(B.37)}\\
&+\frac{1}{\kappa}\sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\int_{F_x^{-1}(\kappa)}^{G_y^{-1}(\kappa)} G_y(t) - \kappa \, dt\right|, &\text{(B.38)}
\end{aligned}$$

where $s_\mathcal{X}^{(F_x^{-1}(\kappa))}(x) = ||d_\mathcal{X}(x,\cdot) \wedge F_x^{-1}(\kappa)||_{\ell^1(\mu_\mathcal{X})}$ and $s_\mathcal{Y}^{(F_x^{-1}(\kappa))}(y)$ is defined analogously. In particular, we have for $\kappa = 1$ that

$$T_\kappa(\mathcal{X},\mathcal{Y}) = 2\mathbf{FLB}_1(\mathcal{X},\mathcal{Y}).$$

Remark B.26. In the proof of Lemma B.25 below, we observe that the roles of $F_x^{-1}(\kappa)$ and $G_y^{-1}(\kappa)$ in (B.37) and (B.38) are interchangeable, i.e., the same line of proof also implies that

$$\begin{aligned}T_\kappa(\mathcal{X},\mathcal{Y}) \le & \frac{1}{\kappa}\inf_{\pi\in\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})}\int_{\mathcal{X}\times\mathcal{Y}}\left|s_\mathcal{X}^{(G_y^{-1}(\kappa))}(x)-s_\mathcal{Y}^{(G_y^{-1}(\kappa))}(y)\right|\,d\pi(x,y)\\ &+\frac{1}{\kappa}\sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\int_{G_y^{-1}(\kappa)}^{F_x^{-1}(\kappa)}\kappa-F_x(t)\,dt\right|.\end{aligned}$$

Proof. We start with proving the first part of the statement. For any $\pi\in\mathcal{M}(\mu_\mathcal{X},\mu_\mathcal{Y})$ it holds that

$$\begin{aligned}\mathcal{K}_1(D_{\mathcal{X},\kappa},D_{\mathcal{Y},\kappa}) \le & \int_{\mathcal{X}\times\mathcal{Y}}\left|\delta_{\mathcal{X},\kappa}(x)-\delta_{\mathcal{Y},\kappa}(y)\right|\,d\pi(x,y)\\ =&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^\kappa F_x^{-1}(l)\,dl-\int_0^\kappa G_y^{-1}(l)\,dl\right|\,d\pi(x,y)\\ =&\frac{1}{\kappa}\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}G_x(l)-F_y(l)\,dl+\int_{F_x^{-1}(\kappa)}^{G_y^{-1}(\kappa)}G_y(l)-\kappa\,dl\right|\,d\pi(x,y),\end{aligned}$$

where the last step follows by Lemma H.1. Consequently, we obtain that

$$\begin{aligned}&\mathcal{K}_1(D_{\mathcal{X},\kappa},D_{\mathcal{Y},\kappa})\\ \le&\frac{1}{\kappa}\left(\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}F_x(l)-G_y(l)\,dl\right|\,d\pi(x,y)+\sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\int_{F_x^{-1}(\kappa)}^{G_y^{-1}(\kappa)}G_y(t)-\kappa\,dt\right|\right).\end{aligned}$$

It remains to show that

$$\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}F_x(l)-G_y(l)\,dl\right|\,d\pi(x,y)\le\int_{\mathcal{X}\times\mathcal{Y}}\left|s_\mathcal{X}^{(\kappa)}(x)-s_\mathcal{Y}^{(\kappa)}(y)\right|\,d\pi(x,y).$$

To this end, we observe that

$$\begin{aligned}&\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}F_x(l)-G_y(l)\,dl\right|\,d\pi(x,y)\\ =&\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}\int_\mathcal{X}\mathbb{1}_{\{d_\mathcal{X}(x,x')\le l\}}\,d\mu_\mathcal{X}(x')-\int_\mathcal{Y}\mathbb{1}_{\{d_\mathcal{Y}(y,y')\le l\}}\,d\mu_\mathcal{Y}(y')\,dl\right|\,d\pi(x,y)\\ =&\int_{\mathcal{X}\times\mathcal{Y}}\left|\int_0^{F_x^{-1}(\kappa)}\int_{\mathcal{X}\times\mathcal{Y}}\mathbb{1}_{\{d_\mathcal{X}(x,x')\le l\}}-\mathbb{1}_{\{d_\mathcal{Y}(y,y')\le l\}}\,d\pi'(x',y')\,dl\right|\,d\pi(x,y)\end{aligned}$$

for any $\pi' \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$. The theorem of Tonelli/Fubini [6, Thm. 18.3] yields that

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{Y}} \left| \int_0^{F_x^{-1}(\kappa)} F_x(l) - G_y(l)\, dl \right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_{\mathcal{X}\times\mathcal{Y}} \int_0^{F_x^{-1}(\kappa)} \mathbb{1}_{\{d_{\mathcal{X}}(x,x')\le l\}} - \mathbb{1}_{\{d_{\mathcal{Y}}(y,y')\le l\}}\, dl\, d\pi'(x',y') \right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_{\mathcal{X}\times\mathcal{Y}} \big(d_{\mathcal{Y}}(y,y') \wedge F_x^{-1}(\kappa)\big) - \big(d_{\mathcal{X}}(x,x') \wedge F_x^{-1}(\kappa)\big)\, d\pi'(x',y') \right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| s_{\mathcal{X}}^{(F_x^{-1}(\kappa))}(x) - s_{\mathcal{Y}}^{(F_x^{-1}(\kappa))}(y) \right| \, d\pi(x,y),
\end{aligned}$$

which gives the first part of the claim.

We come to the second part of the statement. Let $\kappa = 1$. We observe that for any $\pi \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ it follows that

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{Y}} \left|\delta_{\mathcal{X},1}(x) - \delta_{\mathcal{Y},1}(y)\right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_0^1 F_x^{-1}(l)\, dl - \int_0^1 G_y^{-1}(l)\, dl \right| \, d\pi(x,y) \\
=& \frac{1}{\kappa} \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_0^{\infty} F_x(l) - G_y(l)\, dl \right| \, d\pi(x,y),
\end{aligned}$$

where the last step follows by Lemma H.1. Hence, we obtain

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{Y}} \left|\delta_{\mathcal{X},1}(x) - \delta_{\mathcal{Y},1}(y)\right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_0^{\infty} \int_{\mathcal{X}} \mathbb{1}_{\{d_{\mathcal{X}}(x,x')\le l\}}\, d\mu_{\mathcal{X}}(x') - \int_{\mathcal{Y}} \mathbb{1}_{\{d_{\mathcal{Y}}(y,y')\le l\}}\, d\mu_{\mathcal{Y}}(y')\, dl \right| \, d\pi(x,y) \\
=& \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_0^{\infty} \int_{\mathcal{X}\times\mathcal{Y}} \mathbb{1}_{\{d_{\mathcal{X}}(x,x')\le l\}} - \mathbb{1}_{\{d_{\mathcal{Y}}(y,y')\le l\}}\, d\pi'(x',y')\, dl \right| \, d\pi(x,y),
\end{aligned}$$

for any $\pi' \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$. Since the spaces $\mathcal{X}$ and $\mathcal{Y}$ are compact, the Theorem of

Tonelli/Fubini [6, Thm. 18.3] gives that

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{Y}} |\delta_{\mathcal{X},1}(x) - \delta_{\mathcal{Y},1}(y)| \, d\pi(x,y) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_{\mathcal{X}\times\mathcal{Y}} \int_0^\infty \mathbb{1}_{\{d_{\mathcal{X}}(x,x')\le l\}} - \mathbb{1}_{\{d_{\mathcal{Y}}(y,y')\le l\}} \, dl \, d\pi'(x',y') \right| \, d\pi(x,y) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} \left| \int_{\mathcal{X}\times\mathcal{Y}} d_{\mathcal{X}}(x,x') - d_{\mathcal{Y}}(y,y') \, d\pi'(x',y') \right| \, d\pi(x,y) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} |s_{\mathcal{X}}(x) - s_{\mathcal{Y}}(y)| \, d\pi(x,y).
\end{aligned}$$

Since by Lemma 28 in Chowdhury and Mémoli [14] it follows that

$$\mathcal{K}_1(D_{\mathcal{X},1}, D_{\mathcal{Y},1}) = \mathcal{K}_1(\delta_{\mathcal{X},1}\#\mu_{\mathcal{X}}, \delta_{\mathcal{Y},1}\#\mu_{\mathcal{Y}}) = \inf_{\pi\in\mathcal{M}(\mu_{\mathcal{X}},\mu_{\mathcal{Y}})} \int_{\mathcal{X}\times\mathcal{Y}} |\delta_{\mathcal{X},1}(x) - \delta_{\mathcal{Y},1}(y)| \, d\pi(x,y),$$

minimizing over $\pi \in \mathcal{M}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ yields the claim. □

C Distributions of Euclidean Distances on k-Ahlfors Regular Sets

In Section B.2 we derived sufficient conditions for a metric measure space $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$, where $\mathcal{X} \subset \mathbb{R}^d$ has positive Lebesgue measure, $d_{\mathcal{X}}$ denotes the Euclidean distance and $\mu_{\mathcal{X}}$ the uniform distribution, to fulfill Condition 1.2 and Condition 1.3. Naturally, it is of interest to extend these results to metric measure spaces, where $\mathcal{X}$ is a curve or a hypersurface. In the following, we focus on a simple class of hypersurfaces equipped with the Euclidean distance and the uniform distribution and verify that the corresponding metric measure spaces meet the requirements of Condition 1.2 and Condition 1.3 under specific assumptions. To this end, we first concisely introduce some important concepts from geometric measure theory in Section C.1 (see [20, 31] for more information) and derive sufficient conditions for Condition 1.2 and Condition 1.3 if the metric measure space $\mathcal{X}$ (equipped with Euclidean distance and the uniform distribution) is the image of a *k-Ahlfors* regular set (see Definition C.6) under a sufficiently smooth diffeomorphism in Section C.2.

C.1 Preliminary Results

First of all, let us introduce the *Hausdorff measure*, which is a general area measure for subsets of $\mathbb{R}^d$.

Definition C.1 (Morgan [31, Sec. 2.3])**.** Given $A \subset \mathbb{R}^d$, define the *k-dimensional Hausdorff measure* $\mathscr{H}^k(A)$ by

$$\mathscr{H}^k(A) := \lim_{\delta\to 0} \inf_{\substack{A\subseteq\bigcup S_i,\\ \text{diam}(S_i)\leq\delta}} \sum \alpha_k \left(\frac{\text{diam}\,(S_i)}{2}\right)^k,$$

where the infimum is taken over all countable coverings S_i of A with $\text{diam}\,(S_i) < \delta$ and $\alpha_k = \Gamma^k\left(\frac{1}{2}\right)/\Gamma\left(\frac{k}{2}+1\right)$ denotes the volume of the unit ball in $\mathbb{R}^k$.

Remark C.2. The Hausdorff measure is an outer measure, which is countably additive on the Borel sets of $\mathbb{R}^d$ and gives the correct area for C^1-manifolds of $\mathbb{R}^d$ (see [20, Chap. 3]).

The Hausdorff measure allows us to define probability measures on very general subsets of $\mathbb{R}^d$ and can be used to define complex metric measure spaces. However, as already mentioned, we restrict our considerations to the distribution of Euclidean distances on sets equipped with a "uniform distribution". This is made precise in the following definition.

Definition C.3. Let $\mathcal{X} \subset \mathbb{R}^d$ be a Borel set such that there exists $0 < k \leq d$ with $0 < \mathscr{H}^k(\mathcal{X}) < \infty$. Then, the *standard Euclidean metric measure space* induced by $\mathcal{X}$ denotes the triple $(\mathcal{X}, d_\mathcal{X}, \mu_\mathcal{X})$, where $d_\mathcal{X}$ denotes the Euclidean distance and $\mu_\mathcal{X} = \frac{\mathscr{H}^k|_\mathcal{X}}{\mathscr{H}^k(\mathcal{X})}$. Given a standard Euclidean metric measure space, we denote by F_x, $x \in \mathcal{X}$, the distribution function

$$F_x(t) = \mathbb{P}\left(||x - X|| \leq t\right),$$

where $X \sim \mu_\mathcal{X}$. Further, we recall that in this setting the distribution function U of $d_\mathcal{X}(X, X')$, $X, X' \overset{i.i.d.}{\sim} \mu_\mathcal{X}$, is given as

$$U(t) = \int F_x(t)\, d\mu_\mathcal{X}(x) = \mathbb{P}\left(||X - X'|| \leq t\right). \tag{C.1}$$

The basic area formula [20, Sec. 3.2.5] is a useful extension of the classical change of variables formula and essential for our considerations. However, we will not apply it in its full generality and mostly work with the following simple corollary.

Corollary C.4. *Let $W_1 \subset \mathbb{R}^k$ be an open set and $A \subseteq W_1$ a Borel set. Let $h : W_1 \to W_2 \subset \mathbb{R}^d$ be bi-Lipschitz, $k \leq d$. Then, the Jacobian matrix J_h of h exists almost everywhere and the Gram determinant $G_h : A \to [0, \infty)$, defined by*

$$G_h(z) = det(J_h(z)^T J_h(z)),$$

is bounded away from 0 and ∞. Furthermore, h is injective and the relation

$$\int_{h(A)} g(x)\, d\mathscr{H}^k(x) = \int_A g(h(z)) G_h^{1/2}(z)\, d\lambda_k(z)$$

holds for any measurable function $g : h(A) \to [0, \infty]$.

Proof. The differentiability of h outside of a null set is subject to Rademacher's theorem [20, Thm. 3.1.6] and the claim follows by the area formula [20, Sec. 3.2.5] once we have shown that there are constants c_1, c_2 such that $0 < c_1 \leq G_h(z) \leq c_2 < \infty$ for $z \in A$. In order to bound the Gram determinant, we consider a point z where h is differentiable and assume that $t > 0$ and $w \in \mathbb{R}^k$ with $||w|| = 1$. Since h is bi-Lipschitz continuous, there are global constants $0 < c < C < \infty$ such that

$$ct \leq ||h(z + tw) - h(z)|| \leq Ct$$

holds for t sufficiently small (such that $z + tw \in W_1$). Dividing by t and taking the limit $t \to 0$ yields that

$$c \leq ||J_h(z)w|| \leq C,$$

which states that $J_h(z)w$ is uniformly bounded for any unit vector w and point of differentiability $z \in W_1$. This lets us control the spectrum of the Gram matrix $J_h(z)^T J_h(z)$ for which

$$\sigma_{\min} = \min_{||w||=1} \|J_h(z)w\|^2 \geq c^2 \text{ and } \sigma_{\max} = \max_{||w||=1} \|J_h(z)w\|^2 \leq C^2$$

are the smallest and largest eigenvalues. Hence, we find that

$$c^{2d} \leq \det\left(J_h(z)^T J_h(z)\right) \leq C^{2d}.$$

As already argued, this yields the claim. □

The subsequent example shows, how we can use Corollary C.4 to calculate the cumulative distribution function of the distribution of distances in some specific settings.

Example C.5. Let $W_1 \subset \mathbb{R}^k$ be open and let $h : W_1 \to W_2 \subset \mathbb{R}^d$, $k \leq d$, be bi-Lipschitz. Let $\mathcal{X} = h(A) \subset \mathbb{R}^d$, where $A \subseteq W_1$ is Borel and $\lambda_k(A) > 0$. Let $0 < \mathscr{H}^k(\mathcal{X}) < \infty$ and consider the standard Euclidean metric measure space induced by $\mathcal{X}$. Further, let $x \in \mathcal{X}$ and $X, X' \sim \mu_{\mathcal{X}}$. Then, it follows by Corollary C.4 that

$$\begin{aligned}F_x(t) = \mathbb{P}\left(||x - X|| \leq t\right) =& \frac{1}{\mathscr{H}^k(\mathcal{X})} \int_{h(A)} \mathbb{1}_{\{||x-z||)\leq t\}} \, d\mathscr{H}^k(z)\\ =& \frac{1}{\mathscr{H}^k(\mathcal{X})} \int_A \mathbb{1}_{\{||x-h(z)||\leq t\}} G_h^{1/2}(z) \, d\lambda_k(z).\end{aligned}$$

A further application of Corollary C.4 yields that in this case

$$\begin{aligned}U(t) =& \mathbb{P}\left(d_{\mathcal{X}}(X, X') \leq t\right)\\ =& \frac{1}{(\mathscr{H}^k(\mathcal{X}))^2} \int_A \int_A \mathbb{1}_{\{||h(z_1)-h(z_2)||\leq t\}} G_h^{1/2}(z_1) G_h^{1/2}(z_2) \, d\lambda_k(z_1) d\lambda_k(z_2).\end{aligned}$$

In Example C.5 we have assumed that $\mathcal{X}$ is the bi-Lipschitz image of a Borel set A. However, as Borel sets can be quite general, it is clear that we need some more restrictions on the set A (resp. $h(A)$) in order to ensure that Condition 1.2 and Condition 1.3 are met. One useful regularity condition is to assume that the balls in the considered set roughly behave like balls in a d-dimensional plane (see the subsequent definition).

Definition C.6 (David and Semmes [16, Sec.2])**.** A set $\mathcal{X} \subset \mathbb{R}^d$ is called *k-Ahlfors* regular for $k > 0$ if it is closed and there exists a constant $C_H > 1$ such that

$$\frac{1}{C_H} r^k \leq \mathscr{H}^k(\mathcal{X} \cap B(x,r)) \leq C_H r^k \tag{C.2}$$

for all x in $\mathcal{X}$. Here, $B(x,r)$ denotes the open ball with center x and radius r.

Remark C.7. Classical examples of Ahlfors regular sets are balls, spheres, compact domains with C^2-boundary (or their boundaries). For more examples see [16, 35] and the references therein.

The following lemma contains some useful facts about k-Ahlfors regular sets (see Troscheit [39, Sec. 4.2] and Mattila and Saaranen [26, Sec.1]).

Lemma C.8. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a bounded k-Ahlfors regular set. Then, it holds:*

(i) $\mathscr{H}^k(\mathcal{X}) < \infty$, i.e., $\mu_{\mathcal{X}} = \frac{\mathscr{H}^k|_{\mathcal{X}}}{\mathscr{H}^k(\mathcal{X})}$ is well defined;

(ii) $\Psi(\mathcal{X})$ is k-Ahlfors regular and bounded, if Ψ is a bi-Lipschitz map.

Remark C.9. In particular, Lemma C.8 implies that every bounded, k-Ahlfors regular subset of $\mathbb{R}^d$ induces a standard Euclidean metric measure space and that the same is true for bi-Lipschitz images of bounded k-Ahlfors regular sets.

The following lemma demonstrates that the assumption that $\mathcal{X}$ is k-Ahlfors, $k \geq 2$, already imposes some regularity for the distribution of distances of the induced standard Euclidean metric measure space.

Lemma C.10. *Let $\mathcal{X}$ be a bounded k-Ahlfors regular set, $k \geq 2$, and consider the standard, Euclidean metric measure space induced by $\mathcal{X}$. Then, we obtain:*

1. *The function $F_x(t)$ is differentiable at $t = 0$ with $F_x'(0) = 0$ and the same is true for $U(t)$.*
2. *Suppose there exists $\delta > 0$ such that $F_x(t)$ is differentiable and convex on $[0, \delta)$ for $\mu_{\mathcal{X}}$-almost all $x \in \mathcal{X}$. Further, assume that for each $t \in [0, \delta)$ there exists $g \in \ell^1(\mu_{\mathcal{X}})$, $\epsilon > 0$ and $\mathcal{X}_t \subseteq \mathcal{X}$ with $\mu_{\mathcal{X}}(\mathcal{X}_t) = 1$ and $f_x(s) = F_x'(s) \leq g(x) < \infty$ for $s \in (t-\epsilon, t+\epsilon)$ and all $x \in \mathcal{X}_t$. Then, it follows that U is differentiable on $[0, \delta)$ and that there exists a constant c_U such that*

$$\left|(U^{-1})'(t)\right| \leq c_U t^{-\frac{k-1}{k}} \tag{C.3}$$

for $t \in [0, \delta)$.

Proof. We start by proving the first statement. Clearly, we have that $F_x(t) = 0$ and $U(t) = 0$ for all $t \leq 0$. Furthermore, it is easy to see that

$$F_x(t) = \frac{\mathscr{H}^k(\mathcal{X} \cap B(x,t))}{\mathscr{H}^k(\mathcal{X})}.$$

As $\mathcal{X}$ is k-Ahlfors regular we obtain that

$$0 \leq f_x(0) = F_x'(0) = \lim_{h \searrow 0} \frac{F_x(h) - F_x(0)}{h} = \lim_{h \searrow 0} \frac{\frac{\mathscr{H}^k(\mathcal{X} \cap B(x,h)}{\mathscr{H}^k(\mathcal{X})}}{h} \leq \lim_{h \searrow 0} \frac{\frac{C_H h^k}{\mathscr{H}^k(\mathcal{X})}}{h} = 0.$$

Since

$$0 \leq u(0) = U'(0) = \lim_{h \searrow 0} \frac{\int F_x(t)\, d\mu_{\mathcal{X}}(x)}{h} \leq \lim_{h \searrow 0} \frac{\frac{C_H h^k}{\mathscr{H}^k(\mathcal{X})}}{h} = 0,$$

the first statement follows.

Next, we prove the second statement. Let $x \in \mathcal{X}$ be such that $F_x(t)$ is differentiable and convex on $[0, \delta)$. The convexity of F_x on $[0, \delta)$ implies that

$$F_x(t) \geq F_x(s) + f_x(s)(t - s)$$

for all $s, t \in [0, \delta)$. In consequence, we find for $t < s$ that

$$f_x(s) \geq \frac{F_x(s) - F_x(t)}{s - t}.$$

Let now $t = \frac{s}{(2C_H^2)^{\frac{1}{k}}}$, where C_H is as in (C.2). Since $C_H > 1$, we find that $t < s$ and hence

$$f_x(s) \geq \frac{F_x(s) - F_x\left(\frac{s}{(2C_H^2)^{\frac{1}{k}}}\right)}{\left(1 - \frac{1}{(2C_H^2)^{\frac{1}{k}}}\right) s} \geq \frac{\frac{s^k}{2\mathscr{H}^k(\mathcal{X}) C_H}}{\left(1 - \frac{1}{(2C_H^2)^{\frac{1}{k}}}\right) s} = C_1 s^{k-1}, \tag{C.4}$$

where we have used that F_x is monotonically increasing and C_1 denotes some constant that only depends on k, $\mathscr{H}^k(\mathcal{X})$ and C_H. In consequence, we have for all $s \in [0, \delta)$ that

$$f_x(s) \geq C_1 s^{k-1} \geq C_1 \left(C_H s^k\right)^{\frac{k-1}{k}} \geq C_2 \left(F_x(t)\right)^{\frac{k-1}{k}},$$

where C_2 denotes some constant. In the last inequality, we have used that the set $\mathcal{X}$ is k-Ahlfors regular and thus

$$F_x(t) = \frac{\mathscr{H}^k(\mathcal{X} \cap B(x,t))}{\mathscr{H}^k(\mathcal{X})} \leq \frac{C_H}{\mathscr{H}^k(\mathcal{X})} t^k. \tag{C.5}$$

The next step in order to show the second claim is to demonstrate that for $t \in [0, \delta)$

$$u(t) = \int f_x(t)\, d\mu_{\mathcal{X}}(x). \tag{C.6}$$

Since the requirements of Lemma H.3 are met by assumption for all $t \in [0, \delta)$, (C.6) follows by the representation of $U(t)$ in (C.1) and an application of the aforementioned theorem. Since (C.6) in combination with (C.4) yields that

$$u(t) \geq \int C_1 t^{k-1}\, d\mu_{\mathcal{X}}(x) = C_1 t^{k-1}$$

and by (C.1) and (C.5)

$$U(t) \leq \int \frac{C_H}{\mathscr{H}^k(\mathcal{X})} t^k\, d\mu_{\mathcal{X}}(x) = \frac{C_H}{\mathscr{H}^k(\mathcal{X})} t^k,$$

we find that

$$u(t) \geq C_1 t^{k-1} \geq C_1 \left(C_H t^k\right)^{\frac{k-1}{k}} \geq C_2 \left(U(t)\right)^{\frac{k-1}{k}}.$$

As already noted in Section B.2, this is equivalent to (C.3) and thus we have shown the claim. □

Remark C.11. The case of 1-Ahlfors regular sets is special, in the sense that the same arguments as previously applied show that

$$\lim_{h \searrow 0} \frac{F_x(h) - F_x(0)}{h} \geq \lim_{h \searrow 0} \frac{\frac{C_H^{-1} h}{\mathscr{H}^k(\mathcal{X})}}{h} = \frac{1}{C_H \mathscr{H}^k(\mathcal{X})}$$

and

$$\lim_{h \searrow 0} \frac{F_x(h) - F_x(0)}{h} \leq \frac{C_H}{\mathscr{H}^k(\mathcal{X})}.$$

This means that $f_x(0)$ (if it exists) is bounded and bounded away from zero. Furthermore, we observe that the same holds true for $u(0) = U'(0)$. As discussed previously in Section 1, this simplifies our asymptotic considerations (provided that we can verify the required differentiability).

We conclude the preliminaries with the introduction of one result on the differentiability of general probability functions of the form

$$P(t) = \int_{\phi(x)\leq t} p(x)\, d\lambda_d(x),$$

where $\phi(x) : \mathbb{R}^d \to \mathbb{R}$ denotes a Lipschitz continuous function and p a Lebesgue density. The following theorem will be essential for verifying that F_x is differentiable in the setting of Example C.5.

Theorem C.12. *Let $p : \mathbb{R}^d \to \mathbb{R}$ be a Lebesgue density. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function such that*

$$g(x) = \begin{cases} \frac{p(x)}{||\nabla\phi(x)||} \mathbb{1}_{\phi^{-1}(\mathbb{R})}(x), & \text{if } p(x) > 0 \\ 0, & \text{else.} \end{cases}$$

is in $\ell^1(\mathbb{R}^d)$. Then, it follows that

$$P(t) = \int_{\phi(x)\leq t} p(x)\, d\lambda_d(x) = \int_{-\infty}^{t} \int_{\phi^{-1}(y)} \frac{p(x)}{||\nabla\phi(x)||}\, d\mathscr{H}^{d-1}(x)\, d\lambda(y). \tag{C.7}$$

Proof. The differentiability of ϕ outside a null set follows by Rademacher's theorem [20, Thm. 3.1.6]. In the following, we will verify (C.7) by an application of the Co-Area Formula [20, Sec. 3.2.12]. To this end, let Z denote a random variable with density p and define for $B \in \mathscr{B}(\mathbb{R}^d)$

$$g_B(x) := \begin{cases} \frac{p(x)}{||\nabla\phi(x)||} \mathbb{1}_{\phi^{-1}(B)}(x), & \text{if } p(x) > 0 \\ 0, & \text{else.} \end{cases}$$

Obviously, $|g_B(x)| \leq |g(x)|$ and hence $g_B \in \ell^1(\mathbb{R}^d)$ for any B. Consequently, the Co-Area Formula [20, Sec. 3.2.12] yields that

$$\begin{aligned} \mathbb{P}(\phi(Z) \in B) = \mathbb{P}\left(Z \in \phi^{-1}(B)\right) &= \int_{\phi^{-1}(B)} p(x)\, d\lambda_d(x) \\ &= \int_{\mathbb{R}^d} g_B(x) ||\nabla\phi(x)||\, d\lambda_d(x) \\ &= \int_{\mathbb{R}} \int_{\phi^{-1}(y)} \frac{p(x)}{||\nabla\phi(x)||} \mathbb{1}_{\phi^{-1}(B)}(x)\, d\mathscr{H}^{d-1}(x)\, d\lambda(y). \end{aligned}$$

If $y \notin B$, we have that $\phi^{-1}(y) \cap \phi^{-1}(B) = \emptyset$. Consequently, we find that

$$\mathbb{P}(\phi(Z) \in B) = \int_B \int_{\phi^{-1}(y)} \frac{p(x)}{||\nabla\phi(x)||}\, d\mathscr{H}^{d-1}(x)\, d\lambda(y).$$

Hence, the claim follows for $B = (-\infty, t]$. □

The subsequent example establishes that the above theorem can be applied to derive the density f_x of $F_x(t) = \mathbb{P}(d_{\mathcal{X}}(x, X) \leq t)$ in the setting of Example C.5.

Example C.13. Let $W_1 \subset \mathbb{R}^k$ be open and let $h : W_1 \to W_2 \subset \mathbb{R}^d$, $k \leq d$, be a C^1-diffeomorphism. Let $\mathcal{X} = h(A) \subset \mathbb{R}^d$, where $A \subseteq W_1$ is bounded and k-Ahlfors regular (this implies $\lambda_k(A) > 0$). Let $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ denote the standard Euclidean metric measure space induced by $\mathcal{X}$. Fix $x \in \mathcal{X}$. Then, it follows by Example C.5 that

$$F_x(t) = \frac{1}{\mathscr{H}^k(\mathcal{X})} \int_{||x-h(z)|| \leq t} G_h^{1/2}(z) \mathbb{1}_A(z)\, d\lambda_k(z).$$

Additionally define $\Psi_x(z) \coloneqq ||x - h(z)||$ and assume that the function

$$g(z) = \begin{cases} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla_z \Psi_x(z)||} \mathbb{1}_{A \cap \Psi_x^{-1}(\mathbb{R})}(z), & \text{if } G_h^{1/2}(z)\mathbb{1}_A(z) > 0 \\ 0, & \text{else.} \end{cases}$$

is contained in $\ell^1(\mathbb{R}^k)$. In the following, we will demonstrate that it is possible to derive a density for F_x under the assumptions made.

Since the function Ψ_x is Lipschitz continuous as the composition of two Lipschitz continuous functions, it follows by Theorem C.12 it that

$$F_x(t) = \int_0^t \int_{\Psi_x^{-1}(y)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla_z \Psi_x(z)||} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z)\, d\lambda(y).$$

Now, we observe that $\nabla_z \Psi_x(z)$ is well defined for all $z \in A$ and $x \in \mathcal{X}$ such that $h(x) \neq z$ (i.e. $z \notin \Psi_x^{-1}(0)$). Further, Lemma C.10 suggests that for $k \geq 2$ the choice $f_x(0) = 0$ is natural. Hence, we find that the density of F_x is given as

$$f_x(t) = \begin{cases} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla_z \Psi_x(z)||} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z), & \text{if } t > 0 \\ 0, & \text{else.} \end{cases}$$

Remark C.14. In the setting of Example C.13 with $k = 1$, the analogous arguments give that

$$F_x(t) = \int_0^t \int_{\Psi_x^{-1}(y)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla_z \Psi_x(z)||} \mathbb{1}_A(z)\, d\mathscr{H}^0(z)\, d\lambda(y),$$

which means that also in this case a density f_x exists. However, it is, different from the case $k \geq 2$, not clear that F_x is differentiable at 0 and hence the choice of $f_x(0)$ is not obvious.

C.2 Main Results

In the remainder of this section, we will derive sufficient assumptions for a standard Euclidean metric measure space $\mathcal{X}$, which is the image of a k-Ahlfors regular set, $k \geq 2$, under a sufficiently smooth diffeomorphism, to fulfill Condition 1.2 (see Theorem C.16) and Condition 1.3 (Theorem C.18). Similar arguments hold for the images of 1-Ahlfors regular sets, which are slightly different as pointed out previously, and we briefly comment on this issue at the end of the section. However, before we come to that, we need to establish some regularity properties of f_x in the setting considered.

Lemma C.15. *Let $W_1 \subset \mathbb{R}^k$ be open and let $h : W_1 \to W_2 \subset \mathbb{R}^d$, $k \leq d$, be a C^2-diffeomorphism. Let $\mathcal{X} = h(A) \subset \mathbb{R}^d$, where $A \subseteq W_1$ is bounded and k-Ahlfors regular (i.e. $\lambda_k(A) > 0$). Let $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ denote the standard Euclidean metric measure space induced by $\mathcal{X}$. Let F_x be supported on $[0, D_x]$, $x \in \mathcal{X}$. We recall that $\Psi_x(z) := ||x - h(z)||$ and assume that the function*

$$g_x(z) = \begin{cases} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla_z \Psi_x(z)||} \mathbb{1}_{A \cap \Psi_x^{-1}(\mathbb{R})}(z), & \text{if } G_h^{1/2}(z)\mathbb{1}_A(z) > 0 \\ 0, & \text{else.} \end{cases}$$

is contained in $\ell^1(\mathbb{R}^k)$.

(i) Suppose that there exists $\epsilon > 0$ such that $||\nabla_z \Psi_x(z)||$ is bounded and bounded away from zero on $\Psi_x^{-1}(0, \epsilon]$. Further, assume that $\lim_{t \searrow 0} \mathscr{H}^{k-1}(A \cap \Psi_x^{-1}(t)) = 0$. Then, it follows that

$$\lim_{t \searrow 0} f_x(t) = 0.$$

(ii) Assume there exists $0 < t_1 \leq t \leq t_2 \leq D_x$ such that $||\nabla_z \Psi_x(z)||$ is bounded away from zero on $\Psi_x^{-1}[t_1, t_2]$. Further, assume that $\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap \partial A) = 0$. Then, it follows that f_x is continuous at t.

Proof. Under the assumptions made, Example C.13 yields that for all $t > 0$

$$f_x(t) = \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X}) \, ||\nabla_z \Psi_x(z)||} \mathbb{1}_A(z) \, d\mathscr{H}^{k-1}(z). \tag{C.8}$$

Based on this identity, we will prove the two statements.

We start by proving (i). It holds that

$$0 \leq \lim_{t \searrow 0} f_x(t) = \lim_{t \searrow 0} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X}) \, ||\nabla_z \Psi_x(z)||} \mathbb{1}_A(z) d\mathscr{H}^{k-1}(z) \leq \lim_{t \searrow 0} C \int_{\Psi_x^{-1}(t) \cap A} d\mathscr{H}^{k-1}(z),$$

where C denotes a constant. The last inequality follows since there are constants C_1, C_2, C_3 such that $0 < C_1 \leq G_h(z) \leq C_2 < \infty$ (see Corollary C.4) and $\|\nabla_z \Psi_x(z)\| \geq C_3 > 0$ (by assumption) for all $z \in \Psi_x^{-1}(0, \epsilon]$. Obviously,

$$\lim_{t \searrow 0} C \int_{\Psi_x^{-1}(t) \cap A} d\mathscr{H}^{k-1}(z) = 0$$

and hence $f_x(t)$ is continuous at 0.

Next, we verify the statement (ii). To this end, we adapt some arguments from the proof of Lemma 46 in Merigot and Thibert [29]. By Cauchy-Lipschitz's theory (see [2, 38]) there exist $t_1 \leq t_1^* < t_2^* \leq t_2$ and $\epsilon' > 0$ such that one can construct a flow $\Phi : [-\epsilon', \epsilon'] \times \Omega \to \mathbb{R}^k$ with

$$\begin{cases} \frac{d}{ds}\Phi(s,z) = \frac{\nabla_z \Psi_x(\Phi(s,z))}{||\nabla_z \Psi_x(\Phi(s,z))||^2} \\ \Phi(0,z) = z, \end{cases}$$

where $\Omega \subset \mathbb{R}^k$ is an open set that contains $\Psi_x^{-1}[t_1^*, t_2^*]$. Differentiating $s \mapsto \Psi_x(\Phi(s,z))$ immediately shows that $\Psi_x(\Phi(s,z)) = \Psi_x(z) + s$. This implies that $\Phi(s, \Psi_x^{-1}(t)) = \Psi_x^{-1}(t+s)$. Furthermore, since $z \mapsto \nabla_z \Psi_x(z)/||\nabla_z \Psi_x(z)||^2$ is in $C^1(\Psi_x^{-1}[t_1, t_2])$, it follows that Φ is in C^1 and that $\Phi(s, \cdot)$ converges pointwise in a C^1 sense to the identity as $s \to 0$. In consequence, we find by a change of variables (see [29, Thm. 56]) that

$$\begin{aligned} \lim_{s^* \to 0} f_x(t + s^*) &= \lim_{s^* \to 0} \int_{\Phi(s^*, \Psi_x^{-1}(t))} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\, \|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z) d\mathscr{H}^{k-1}(z) \\ &= \lim_{s^* \to 0} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(\Phi(s^*, z))}{\mathscr{H}^k(\mathcal{X})\, \|(\nabla_z \Psi_x) \circ \Phi(s^*, z)\|} \mathbb{1}_A(\Phi(s^*, z)) J_{\Phi(s^*, \cdot)}(z) d\mathscr{H}^{k-1}(z), \end{aligned}$$

where $J_{\Phi(s^*, \cdot)}$ denotes the *Jacobian determinant* of $\Phi(s^*, \cdot)$ (see Merigot and Thibert [29] for a formal definition). We have by our previous considerations that there exists constants C_2, C_3 such that $G_h(z) \leq C_2 < \infty$ (see Corollary C.4) and $||\nabla_z \Psi_x(z)|| > C_3 > 0$ (by assumption) for all $z \in \Psi_x^{-1}[t_1, t_2]$. Consequently, the Dominated Convergence Theorem yields that

$$\begin{aligned} \lim_{s^* \to 0} f_x(t + s^*) = \lim_{s^* \to 0} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(\Phi(s^*, z))}{\|(\nabla_z \Psi_x) \circ \Phi(s^*, z)\|} \mathbb{1}_A(\Phi(s^*, z)) J_{\Phi(s^*, \cdot)}(z)\, d\mathscr{H}^{k-1}(z) \\ = \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z) = f_x(t), \end{aligned}$$

where we have used that $\Phi(s, \cdot)$ converges pointwise in a C^1-sense to the identity as $s \to 0$ and that we have by assumption $\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap \partial A) = 0$. In consequence, we have shown that $f_x(t)$ is continuous at t under the assumptions made. □

With Lemma C.15 established, we can finally give sufficient requirements for metric measure spaces constructed as in Lemma C.15 to fulfill Condition 1.2.

Theorem C.16. *Let $W_1 \subset \mathbb{R}^k$ be open and let $h : W_1 \to W_2 \subset \mathbb{R}^d$, $k \leq d$, be a C^2-diffeomorphism. Let $\mathcal{X} = h(A) \subset \mathbb{R}^d$, where $A \subseteq W_1$ is bounded and k-Ahlfors regular (i.e. $\lambda_k(A) > 0$). Let $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ denote the standard Euclidean metric measure space induced by $\mathcal{X}$. Let U be supported on $[0, D]$ and recall that $\Psi_x(z) := ||x - h(z)||$, $x \in \mathcal{X}$. Further, assume that there exists $\mathcal{X}' \subseteq \mathcal{X}$ with $\mu_{\mathcal{X}}(\mathcal{X}') = \mu_{\mathcal{X}}(\mathcal{X})$ such that the function*

$$g_x(z) = \begin{cases} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})||\nabla\Psi_x(z)||}\mathbb{1}_{A\cap\Psi_x^{-1}(\mathbb{R})}(z), & \text{if } G_h^{1/2}(z)\mathbb{1}_A(z) > 0 \\ 0, & \text{else.} \end{cases} \tag{C.9}$$

is contained in $\ell^1(\mathbb{R}^k)$ for all $x \in \mathcal{X}'$.

Let $\int_{\mathcal{X}}\int_{\Psi_x^{-1}(t)\cap A} d\mathscr{H}^{k-1}(z)\, d\mu_{\mathcal{X}}(x) > 0$ for all $t \in (0, D)$, $\sup_{x\in\mathcal{X}'}\sup_{z\in A}||\nabla\Psi_x(z)|| < \infty$ and assume:

(i) There exists $\epsilon > 0$, $C > 0$ and $\mathcal{X}_0 \subseteq \mathcal{X}'$ with $\mu_{\mathcal{X}}(\mathcal{X}_0) = \mu_{\mathcal{X}}(\mathcal{X}')$ such that it holds for all $x \in \mathcal{X}_0$ that $\inf_{z\in\Psi_x^{-1}(0,\epsilon]}||\nabla_z\Psi_x^{-1}(z)|| > C$, $\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap A) > 0$ for all $t \in (0, \epsilon]$ and $\lim_{t\searrow 0}\mathscr{H}^{k-1}(A \cap \Psi_x^{-1}(t)) = 0$.

(ii) For each $0 < t < D$ there exists $g \in \ell^1(\mu_{\mathcal{X}})$, $\epsilon > 0$ with $0 < t - \epsilon < t + \epsilon < D$ and $\mathcal{X}_t \subseteq \mathcal{X}'$ with $\mu_{\mathcal{X}}(\mathcal{X}_t) = \mu_{\mathcal{X}}(\mathcal{X}')$ such that for all $x \in \mathcal{X}_t$ it holds

$$\int_{\Psi_x^{-1}(s)} \frac{1}{||\nabla_z\Psi_x(z)||}\mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z) < g(x) \tag{C.10}$$

for all $s \in [t - \epsilon, t + \epsilon]$ and either $||\nabla_z\Psi_x^{-1}(z)|| \geq c_x > 0$ on $\Psi_x^{-1}[t - \epsilon, t + \epsilon]$ and $\int_{\Psi_x^{-1}(t)\cap\partial A} d\mathscr{H}^{k-1}(z) = 0$ or $\Psi_x^{-1}(t - \delta_x, t + \delta_x) \cap A = \emptyset$ for some $0 < \delta_x < \epsilon$, where c_x and δ_x may depend on $x \in \mathcal{X}_t$.

(iii) There exits $\mathcal{X}_D \subseteq \mathcal{X}'$ with $\mu_{\mathcal{X}}(\mathcal{X}_D) = \mu_{\mathcal{X}}(\mathcal{X}')$ such that

$$\lim_{t\nearrow D}\sup_{x\in\mathcal{X}_D}\int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\,||\nabla_z\Psi_x(z)||}\mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z) = C < \infty \tag{C.11}$$

and for all $x \in \mathcal{X}_D$ we have either $\Psi_x^{-1}(D - \delta_x, D) \cap A = \emptyset$ for some $0 < \delta_x < D$, where δ_x may depend on $x \in \mathcal{X}_D$, or

$$\lim_{t\nearrow D}\int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\,||\nabla_z\Psi_x(z)||}\mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z) = f_x(D).$$

Then, the metric measure space $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ meets the requirements of Condition 1.2 for all $\beta \in (0, 1/2)$.

Proof. First of all, we will establish the continuity of u. To this end, we recall that for $t \in \mathbb{R}$

$$U(t) = \int F_x(t)\, d\mu_{\mathcal{X}}(x)$$

and that the derivative f_x of F_x exists under the assumptions made (see Theorem C.12 and Example C.13) for all t and $x \in \mathcal{X}'$ and is defined as

$$f_x(t) = \begin{cases} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z), & \text{if } t > 0 \\ 0, & \text{else.} \end{cases} \tag{C.12}$$

We will verify the existence of $u = U'$ by applying Lemma H.3 for each t and demonstrate its continuity via the Dominated Convergence Theorem.

Let us consider $t = 0$. Since there are constants C, C_1, C_2, such that $0 < C_1 < G_h(z) < C_2 < \infty$ for all $z \in A$ (see Corollary C.4) and $\|\nabla_z \Psi_x(z)\| > C > 0$ for all $z \in \Psi_x^{-1}(0, \epsilon]$ and $x \in \mathcal{X}_0$, it follows that

$$|f(x, s)| \leq C_3 \int_{\Psi_x^{-1}(s) \cap A} d\mathscr{H}^{k-1}(z) < C_3 < \infty,$$

for all $x \in \mathcal{X}_0$ and $s \in [0, \epsilon]$, where C_3 denotes a constant. In consequence, we know that $|f_x(s)| \leq C_3$ for all $s \in (-\epsilon, \epsilon)$ and $x \in \mathcal{X}_0$. Hence, Lemma H.3 yields that

$$u(0) = \int f_x(0)\, d\mu_{\mathcal{X}}(x).$$

Furthermore, the Dominated Convergence Theorem in combination with Lemma C.15 yields that

$$\lim_{s \searrow 0} u(s) = \int \lim_{s \searrow 0} f_x(s)\, d\mu_{\mathcal{X}}(x) = \int f_x(0)\, d\mu_{\mathcal{X}}(x) = u(0).$$

Now, let $t \in (0, D)$. Since $0 < C_1 < G_h(z) < C_2 < \infty$, the combination of (C.10) and (C.12) yields that there exists $g^* \in \ell^1(\mu_{\mathcal{X}})$ such that $|f_x(s)| \leq g^*(x)$ for all $s \in (t - \epsilon, t + \epsilon)$ and $x \in \mathcal{X}_t$. Hence, we obtain by Lemma H.3 that

$$u(t) = \int f_x(t)\, d\mu_{\mathcal{X}}(x).$$

Next, we consider f_x for one $x \in \mathcal{X}_t$. If $\Psi_x^{-1}(t - \delta_x, t + \delta_x) \cap A = \emptyset$ for some $\delta_x > 0$, then it obviously follows that $\lim_{s \to t} f_x(s) = 0$. Otherwise, we have that the assumptions of Lemma C.15 are met and hence f_x is continuous at t. Thus, the Dominated Convergence Theorem yields that

$$\lim_{s \to t} u(s) = \int \lim_{s \to t} f_x(s)\, d\mu_{\mathcal{X}}(x) = \int f_x(t)\, d\mu_{\mathcal{X}}(x) = u(t),$$

which means that u is continuous at t.

Now, we come to the case $t = D$. By Equation (C.11), it follows that there exist an $\epsilon > 0$ such that for all $s \in [D-\epsilon, D]$ it holds $\sup_{x\in\mathcal{X}_D} f_x(t) \leq 2C_4$. Therefore, Lemma H.3 implies that

$$u(D) = \int f_x(D)\, d\mu_{\mathcal{X}}(x).$$

Further, the Dominated Convergence Theorem yields that

$$\lim_{s\nearrow D} u(s) = \int \lim_{s\nearrow D} f_x(s)\, d\mu_{\mathcal{X}}(x) = \int f_x(D)\, d\mu_{\mathcal{X}}(x) = u(D),$$

where we have use that by assumption either $\Psi_x^{-1}(t-\delta_x, t+\delta_x) \cap A = \emptyset$ for some $\delta_x > 0$ (i.e. $\lim_{s\nearrow D} f_x(s) = 0$) or $\lim_{s\nearrow D} f_x(s) = f_x(D) < \infty$. This establishes the continuity of u.

It remains to verify that $u > 0$ on $(0, D)$. For this purpose, we realize that $0 < C_1 < G_h(z) < C_2 < \infty$ for all $z \in A$ (see Corollary C.4) and that $\sup_{x\in\mathcal{X}'}\sup_{z\in A} ||\nabla\Psi_x(z)|| < \infty$ by assumption. Consequently, there exists a constant C_5 such that for all $t \in (0, D)$

$$\begin{aligned} u(t) &= \int\int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\,\|\nabla_z\Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z)\, d\mu_{\mathcal{X}}(x) \\ &\geq C_5 \int_{\mathcal{X}}\int_{\Psi_x^{-1}(t)\cap A} d\mathscr{H}^{k-1}(z)\, d\mu_{\mathcal{X}}(x) > 0, \end{aligned}$$

where the last inequality follows by the assumptions made. As already argued, this gives the claim. □

Remark C.17. While Assumption (i) of Theorem C.16 seems to be met in many simple examples (see e.g. Example C.20), it is evident that this requirement can be slightly relaxed. Reconsidering the above proof highlights that Assumptions (i) can be replaced by the following:

(i') There exits $\mathcal{X}_0 \subseteq \mathcal{X}'$ with $\mu_{\mathcal{X}}(\mathcal{X}_0) = \mu_{\mathcal{X}}(\mathcal{X}')$ such that

$$\lim_{t\searrow 0}\sup_{x\in\mathcal{X}_0} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X})\,\|\nabla_z\Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z) = 0. \tag{C.13}$$

Finally, we come to the second main theorem of this section, which yields sufficient assumptions for a standard Euclidean metric measure space that is constructed as described in Theorem C.16 to fulfill Condition 1.3

Theorem C.18. *Suppose that the assumption of Theorem C.16 are met. Additionally, let there exist $\delta > 0$ and $\eta > 0$ such that the following requirements are fulfilled:*

(i) The function

$$t \mapsto \frac{1}{\mathscr{H}^k(\mathcal{X})} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z)$$

is monotonically increasing for $t \in (0, \delta)$ *and* $x \in \mathcal{X}$*.*

(ii) The function

$$t \mapsto \frac{1}{(\mathscr{H}^k(\mathcal{X}))^2} \int_{\mathcal{X}} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z)\, d\mathscr{H}^k(x)$$

is monotonically decreasing (or bounded away from zero) for $t \in [D - \delta, D]$*.*

(iii) It holds for $t \in [D - \delta, D]$ *that*

$$\int_{\mathcal{X}} \int_{\Psi_x^{-1}(t) \cap A} \frac{1}{\|\nabla_z \Psi_x(z)\|}\, d\mathscr{H}^{k-1}(z)\, d\mathscr{H}^k(x) \geq c_U (D - t)^{\eta}.$$

Then, $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ *also fulfills the requirements of Condition 1.3.*

Proof. By assumption (and the fact that $f_x(0) = 0$), we have that $f_x(t)$ is monotonically increasing on $[0, \delta)$, $x \in \mathcal{X}$. We have also shown in the proof of Theorem C.16 that for each $t \in [0, D]$ there exists $\epsilon > 0$, $g \in \ell^1(\mu_{\mathcal{X}})$ and a set $\mathcal{X}_t \subseteq \mathcal{X}'$ such that $f_x(s) \leq g(x)$ for all $s \in [t - \epsilon, t + \epsilon]$ and $x \in \mathcal{X}_t$. Thus, we can apply Lemma C.10 and obtain that there exists c_U such that

$$\left|(U^{-1})'(t)\right| \leq c_U t^{-\frac{k-1}{k}}.$$

Further, we have already shown that under the assumptions made $u(t) > 0$ for all $t \in (0, D)$. Hence, the proof of Lemma 2.3 (see Section B.2) suggests that it is sufficient to verify that for a (potentially different) constant c_U

$$\left|(U^{-1})'(t)\right| \leq c_U (1 - t)^{-\frac{\eta}{\eta+1}} \Leftrightarrow u(t) \geq c_U (1 - U(t))^{\frac{\eta}{\eta+1}}$$

for all $t \in [D - \delta, D]$. We observe that

$$(1 - U(t))^{\frac{\eta}{\eta+1}} = \left(\int_0^D u(s)\, ds - \int_0^t u(s)\, ds \right)^{\frac{\eta}{\eta+1}} = (u(s^*)(D - t))^{\frac{\eta}{\eta+1}},$$

for some $s^* \in (t, D)$ by the Mean Value Theorem. Since

$$u(t) = \frac{1}{(\mathscr{H}^k(\mathcal{X}))^2} \int_{\mathcal{X}} \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z)\, d\mathscr{H}^{k-1}(z)\, d\mathscr{H}^k(x),$$

which is monotonically decreasing on $[D-\delta, D]$ by assumption, it follows that

$$(1-U(t))^{\frac{\eta}{\eta+1}} \leq c_U \left(u(t)(D-t)\right)^{\frac{\eta}{\eta+1}},$$

where c_U is a constant that may vary from line to line. Recall that there exists $C_1 > 0$ such that $G_h(z) \geq C_1$ (cf. Corollary C.4). Hence, we find that

$$u(t) \geq c_U \int_{\mathcal{X}} \int_{\Psi_x^{-1}(t)\cap A} \frac{1}{\|\nabla_z \Psi_x(z)\|} \, d\mathscr{H}^{k-1} \, d\mathscr{H}^k(x) \geq c_U (D-t)^\eta,$$

where the last inequality makes use of our assumptions. Plugging in our findings yields that

$$(1-U(t))^{\frac{\eta}{\eta+1}} \leq c_U \left(u^{\frac{\eta+1}{\eta}}(t)\right)^{\frac{\eta}{\eta+1}} = c_U u(t),$$

which as already argued yields the claim. □

Remark C.19. Carefully rereading the proofs of Theorem C.16 and Theorem C.18 highlights that the assumption $k \geq 2$ is mainly made for convenience. If we make suitable differentiability assumptions for F_x at 0, it will be possible to adapt the arguments made. Since it follows in this case by Remark C.11, that $u(0) > 0$ we can even drop the first assumption of Theorem C.18. However, it is important to note that there are some subtle differences. For example, it is noteworthy that the measure $\mathscr{H}^0$ is the *counting measure* (i.e. for a finite set E it holds $\mathscr{H}^0(E) = |E|$). This implies for example that

$$f_x(t) = \int_{\Psi_x^{-1}(t)} \frac{G_h^{1/2}(z)}{\mathscr{H}^k(\mathcal{X}) \, \|\nabla_z \Psi_x(z)\|} \mathbb{1}_A(z) \, d\mathscr{H}^{k-1}(z)$$

is bounded on $[t_1, t_2]$ if and only if $\inf_{z \in \Psi_x^{-1}[t_1,t_2]\cap A} \|\nabla_z \Psi_x(z)\| \geq c > 0$ for some constant c. This is quite different for $k \geq 2$.

Finally, we show that the assumptions of Theorem C.16, which are much easier to check than the even more technical assumptions of Theorem C.18, are met in some simple examples.

Example C.20. In the following, we consider $W_1 = (0.9, 2.1) \times (0.9, 2.1)$, $A = [1,2] \times [1,2] \subset \mathbb{R}^2$ and

$$h : W_1 \subset \mathbb{R}^2 \to \mathbb{R}^3, \ (z_1, z_2) \mapsto (z_1, z_2, z_2^2).$$

It is easy to check that h is a C^2-diffeomorphism from W_1 onto $W_2 = h(W_1)$. In the following, we consider the standard Euclidean metric measure space induced by $\mathcal{X} = h(A)$ and verify that it meets the requirements of Theorem C.16. To this end, we observe $\Psi_x(z) = \sqrt{(x_1-z_1)^2 + (x_2-z_2)^2 + (x_2^2 - z_2^2)^2}$, $x \in \mathcal{X}$, and that for $z \in A \cap (\Psi_x^{-1}(0))^c$

$$\nabla_z \Psi_x(z) = \begin{pmatrix} \frac{x_1 - z_1}{\sqrt{(x_1-z_1)^2+(x_2-z_2)^2+(x_2^2-z_2^2)^2}} \\ \frac{2z_2^3+(1-2x_2^2)z_2-x_2}{\sqrt{(x_1-z_1)^2+(x_2-z_2)^2+(x_2^2-z_2^2)^2}} \end{pmatrix}.$$

Next, we will verify that there exists a constant $C > 0$ such that $||\nabla_z \Psi_x(z)|| \geq C$ for all $z \in A \cap (\Psi_x^{-1}(0))^c$ and $x \in \mathcal{X}$. To this end, we observe that

$$\begin{aligned}\frac{1}{||\nabla_z \Psi_x(z)||^2} &= \frac{(x_1 - z_1)^2 + (x_2 - z_2)^2 + (x_2^2 - z_2^2)^2}{(x_1 - z_1)^2 + (2z_2^3 + (1 - 2x_2^2)z_2 - x_2)^2} \\ &= \frac{(x_1 - z_1)^2 + (x_2 - z_2)^2 + ((x_2 - z_2)(x_2 + z_2))^2}{(x_1 - z_1)^2 + ((z_2 - x_2)(2x_2z_2 + 2z_2^2 + 1))^2} \\ &\leq \frac{(x_1 - z_1)^2 + 17(x_2 - z_2)^2}{(x_1 - z_1)^2 + 25(z_2 - x_2)^2} \leq 1,\end{aligned}$$

which implies the existence of C as described previously. Analogously to the above calculations, one can prove that it holds $\sup_{x\in\mathcal{X}} \sup_{z\in A} ||\nabla_z \Psi_x(z)|| < \infty$. Now, define for $x \in \mathcal{X}$ the function g_x as done in (C.9). Clearly, the previous calculations yield that $g_x \in \ell^1(\lambda_k)$ for all x. Calculating $\Psi_x^{-1}(t)$ is usually cumbersome. Fortunately, it is not strictly necessary in order to check the conditions of Theorem C.16. Since

$$\begin{aligned}\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap A) =& \mathscr{H}^{k-1}(\{z \in A : ||x - h(z)|| = t\}) \\ =& \mathscr{H}^{k-1}(\{y \in \mathcal{X} : ||x - y|| = t\}),\end{aligned} \tag{C.14}$$

it is straightforward to verify that for all $t \in (0, D)$ there exist a set $\mathcal{X}' \subset \mathcal{X}$ with positive measure such that $\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap A) > 0$ for all $x \in \mathcal{X}'$. Further, (C.14) allows us to demonstrate the existence of $\epsilon > 0$ such that $\mathscr{H}^{k-1}(\Psi_x^{-1}(t) \cap A) > 0$ for all $t \in (0, \epsilon]$ and $x \in \mathcal{X}$. By (C.14), it additionally follows that $\lim_{t\searrow 0} \mathscr{H}^{k-1}(A \cap \Psi_x^{-1}(t)) = 0$ for all $x \in \mathcal{X}$ in this setting. Let $[0, D_x]$ denote the support of F_x, $x \in \mathcal{X}$. It is easily verifiable that D_x is $\mu_{\mathcal{X}}$-almost surely unique. This in combination with the fact that $||\nabla_z \Psi_x(z)|| \geq C > 0$ for all $z \in A$ and $x \in \mathcal{X}$ yields that Assumptions (i)-(iii) of Theorem C.16 are fulfilled. In consequence, we find that Theorem C.16 is applicable and $\mathcal{X}$ meets the requirements of Condition 1.2.

In the remainder of this section, we consider curves of the kind

$$h : A \to \mathbb{R}^2, z \mapsto (z, \varphi(z)),$$

where $A \subset \mathbb{R}$ is a closed interval, and verify that the metric measure spaces induced by these curves meet the requirements of Condition 1.2 if the function φ twice differentiable and either monotone or a contraction.

Example C.21. Suppose that $A = [a_1, a_2] \subset W_1$, where $-\infty < a_1 < a_2 < \infty$ and W_1 denotes an open interval. Further, assume that $\varphi : W_1 \to \mathbb{R}$ is a twice continuously differentiable function that is either monotone or a contraction. Then, it is straight forward to verify that

$$h : W_1 \to \mathbb{R}^2, z \mapsto (z, \varphi(z))$$

is a C^2-diffeomorphism onto its image. Denote by $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ the standard Euclidean metric measure space induced by $\mathcal{X} = h(A)$. Clearly, $\mathcal{X}$ is a 1-Ahlfors regular set. Since φ is twice continuously differentiable, it can easily be verified that the distribution function F_x is continuously differentiable at 0 for $\mu_{\mathcal{X}}$-almost all $x \in \mathcal{X}$. Hence, we can try to apply Theorem C.16 in order to check that Condition 1.2 is met in this framework (see Remark C.19, where 1-Ahlfors regular sets are addressed). Clearly, we have that $\Psi_x(z) = \sqrt{(x-z)^2 + (\varphi(x) - \varphi(z))^2}$, $x \in \mathcal{X}$ and that for $z \in A \cap (\Psi_x^{-1}(0))^c$

$$\nabla_z \Psi_x(z) = -\frac{(x-z) + \varphi'(z)(\varphi(x) - \varphi(z))}{\sqrt{(x-z)^2 + (\varphi(x) - \varphi(z))^2}}.$$

In consequence, we find that for $x \in \mathcal{X}$ and $z \in A \cap (\Psi_x^{-1}(0))^c$

$$\left| \frac{1}{\nabla_z \Psi_x(z)} \right| = \left| \frac{|x-z| \sqrt{1 + \left(\frac{\varphi(x) - \varphi(z)}{x-z} \right)^2}}{|x-z| \left(\operatorname{sign}(x-z) + \varphi'(z) \frac{\varphi(x) - \varphi(z)}{|x-z|} \right)} \right|$$

Just as in Example C.20, we demonstrate that there exist a constant $C_1 < \infty$ such that $\left| \frac{1}{\nabla_z \Psi_x(z)} \right| < C_1$ for $x \in \mathcal{X}$ and $z \in A \cap (\Psi_x^{-1}(0))^c$. Since φ is continuously differentiable, it is sufficient to show that

$$\left| \operatorname{sign}(x-z) + \varphi'(z) \frac{\varphi(x) - \varphi(z)}{|x-z|} \right| > C_2 > 0 \tag{C.15}$$

for all $x \in \mathcal{X}$ and $z \in A \cap (\Psi_x^{-1}(0))^c$, where C_2 denotes some constant. For this purpose, we consider three cases.

Case 1: Suppose that φ is a monotonically increasing function. Let $x \in \mathcal{X}$. We observe that $\varphi'(z) \geq 0$ for all $z \in A \cap (\Psi_x^{-1}(0))^c$. Since $z \neq x$ $(x \in \Psi_x^{-1}(0))$, it holds that either $x < z$ or $z > x$.

$x < z$: Here, we have $\operatorname{sign}(x-z) = -1$, $\varphi' \geq 0$ and $\varphi(x) - \varphi(z) < 0$. In consequence,

$$\left| \operatorname{sign}(x-z) + \varphi'(z) \frac{\varphi(x) - \varphi(z)}{|x-z|} \right| \geq |-1| = 1.$$

$x > z$: It holds that $\operatorname{sign}(x-z) = 1$, $\varphi' \geq 0$ and $\varphi(x) - \varphi(z) > 0$. In consequence,

$$\left| \operatorname{sign}(x-z) + \varphi'(z) \frac{\varphi(x) - \varphi(z)}{|x-z|} \right| \geq |1| = 1.$$

Since $x \in \mathcal{X}$ was arbitrary, we have shown (C.15). As already argued, this yields that $\left| \frac{1}{\nabla_z \Psi_x(z)} \right| < C_1 < \infty$ for $z \in A \cap (\Psi_x^{-1}(0))^c$.

Case 2: Suppose that φ is a monotonically decreasing function. This case can be treated with the same arguments as Case 1.

Case 3: Suppose that φ is a contraction. In this case, we have that $|\varphi'(z)| \leq \kappa_1 < 1$ as well as $\frac{|\varphi(x)-\varphi(z)|}{|x-z|} \leq \kappa_2 < 1$ for all $x, z \in \mathcal{X}$. Hence, we find that for $x \in \mathcal{X}$ and $z \in A \cap (\Psi_x^{-1}(0))^c$

$$\begin{aligned} &\left|\text{sign}(x-z) + \varphi'(z)\frac{\varphi(x)-\varphi(z)}{|x-z|}\right| \\ =&\left|\text{sign}(x-z) + \text{sign}(\varphi(z)-\varphi(x))\text{sign}\left(\varphi'(z)\right)|\varphi'(z)|\frac{|\varphi(x)-\varphi(z)|}{|x-z|}\right| \\ \geq& \, |1-\kappa_1\kappa_2| > 0. \end{aligned}$$

In consequence, we have proven (C.15).

Furthermore, we observe that we clearly have that $\left|\frac{1}{\nabla_z \Psi_x(z)}\right| > C > 0$ for some constant C and for all $x \in \mathcal{X}$ and $z \in A \cap (\Psi_x^{-1}(0))^c$. This allows us to argue along the lines of Example C.20 in order to verify that all requirements for an application of (a version of) Theorem C.16 are met (see Remark C.19).

D Bootstrap Approximation

In this section, we use the general results on bootstrapping from Section G to prove that bootstrapping the quantiles of the limit distribution of $\widehat{DoD}_{(\beta)}$ under H_0 is viable for $\beta \in (0, 1/2)$. To this end, we investigate the bootstrap empirical U-process and the bootstrap empirical U-quantile process defined in Example G.1 with the specific kernel function $h(x, y, t) = \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\leq t\}}$.

First, recall that $(\mathcal{X}, d_{\mathcal{X}}, \mu_{\mathcal{X}})$ denotes a metric measure space and that the random variables $X_1, \dots, X_n$ are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Furthermore, we recall that $U(t) = \mathbb{P}(d_{\mathcal{X}}(X, X') \leq t)$, $X, X' \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$, that

$$U_n(t) := \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\leq t\}},$$

that U^{-1} denotes the quantile function of U and that U_n^{-1} stands for the empirical quantile function of U_n. Let μ_n designate the empirical measure based on the sample $X_1, \dots, X_n$. Given the sample values, let $X_1^*, \dots, X_{n_B}^*$ be an independent, identically distributed sample of size n_B from μ_n. Then, the bootstrap empirical U-distribution, $U_{n_B}^*(t)$, $t \in \mathbb{R}$, considered in this section is defined as

$$U^*_{n_B}(t) := \frac{2}{n_B(n_B-1)} \sum_{1\le i<j\le n_B} \mathbb{1}_{\{d_{\mathcal{X}}(X_i^*,X_j^*)\le t\}},$$

the corresponding bootstrap empirical U-process is given as $\mathbb{U}^*_{n_B} = \sqrt{n_B}\left(U^*_{n_B} - U_n\right)$ and the corresponding bootstrap U-quantile process as $\left(\mathbb{U}^*_{n_B}\right)^{-1} = \sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1} - U_n^{-1}\right)$. It is important to note that it is possible to regard $U^*_{n_B}$ (and thus also $\mathbb{U}^*_{n_B}$ and $\left(\mathbb{U}^*_{n_B}\right)^{-1}$) as a functional depending on the sample $\mathcal{X}_n = \{X_1,\dots,X_n\}$ and a random weight vector M_n that is independent of $\mathcal{X}_n$ (see Section G and in particular Example G.1).

We can now formally state the goal of this section. Let $\beta \in (0,1/2)$ be fixed. We aim to approximate the quantiles of $\Xi = \Xi(\beta)$, where

$$\Xi = \int_\beta^{1-\beta} (\mathbb{G}(t))^2 \, dt$$

is defined in (7) as the limit, after proper scaling, of the test statistic $\widehat{DoD}_{(\beta)}$ under the hypothesis. Let Ξ_n and $\Xi^*_{n_B}$ denote an empirical version and a bootstrap empirical version of Ξ, respectively, i.e.,

$$\Xi_n = \int_\beta^{1-\beta} \left(\mathbb{U}_n^{-1}(t)\right)^2 \, dt,$$

and

$$\Xi^*_{n_B} = \int_\beta^{1-\beta} \left(\left(\mathbb{U}^*_{n_B}\right)^{-1}(t)\right)^2 \, dt. \tag{D.1}$$

Given a sample $\mathcal{X}_n = \{X_1,\dots,X_n\}$, we denote by $\Xi^{*,(1)}_{n_B},\dots,\ \Xi^{*,(R)}_{n_B}$ different, independent bootstrap empirical versions of Ξ, as defined in (D.1), where the corresponding bootstrap empirical U-quantile processes $\left(\mathbb{U}^{*,(i)}_n\right)^{-1}$, $1\le i\le R$, are based on independent bootstrap samples of $\mathcal{X}_n$.
We intend to demonstrate that the empirical α-quantile of the sample $\Xi^{*,(1)}_{n_B},\dots,\ \Xi^{*,(R)}_{n_B}$, denoted by $\xi^{(R)}_{n_B,\alpha}$, can be used for testing as proposed in Section 3. As $\Xi^*_{n_B}$ is a functional of the bootstrap empirical U-quantile process, we require some regularity assumptions on U to achieve this. The subsequent conditions comprise the regularity assumptions of Condition 1.2 on U and they are restated for convenience only.

Condition D.1. *Let $\beta \in (0,1/2)$ and let U be continuously differentiable on an interval*

$$[C_1,C_2] = [U^{-1}(\beta)-\epsilon, U^{-1}(1-\beta)+\epsilon]$$

for some $\epsilon > 0$ with strictly positive derivative u.

In order to validate using $\xi^{(R)}_{n_B,\alpha}$ for testing as proposed in Section 3, we pursue the subsequent strategy:

1. We verify that given Condition D.1

$$\mathbb{U}^*_{n_B} = \sqrt{n_B}\left(U^*_{n_B} - U_n\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}$$

   in $\ell^\infty[C_1, C_2]$ (see Theorem D.2), where $\mathbb{K}$ is a centered Gaussian process with covariance $\mathrm{Cov}(\mathbb{K}(s), \mathbb{K}(t)) = 4\Gamma_{d_\mathcal{X}}(s,t)$. Here, $\Gamma_{d_\mathcal{X}}$ is as defined in Lemma B.7.

2. Based on this, we prove that

$$\left(\mathbb{U}^*_{n_B}\right)^{-1} = \sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1} - U_n^{-1}\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{G}$$

   in $\ell^\infty[\beta, 1-\beta]$ (cf. Theorem D.3), where $\mathbb{G}$ is a mean zero Gaussian process with covariance as defined in Theorem 2.6.

3. Finally, we demonstrate that (cf. Theorem D.4)

$$\Xi^*_{n_B} \underset{M}{\overset{P}{\rightsquigarrow}} \Xi.$$

   This induces (see e.g. [11, Sec. 4]) that

$$\lim_{n,n_B,R\to\infty} \mathbb{P}\left(\int_\beta^{1-\beta} \left(\mathbb{U}_n^{-1}(t)\right)^2 dt \geq \xi^{(R)}_{n_B,\alpha}\right) = \alpha,$$

   which shows that it is viable to use $\xi^{(R)}_{n_B,\alpha}$ for testing.

We begin with the first step of the presented strategy and recall that the general empirical U-process indexed by certain function classes can be bootstrapped (cf. Theorem G.4). Based on this and the observations collected in Remark G.3 we prove the following.

Theorem D.2. *Let Condition D.1 be met and let $\sqrt{n_B} = o(n)$. Then, it follows that*

$$\sqrt{n_B}\left(U^*_{n_B} - U_n\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}$$

in $\ell^\infty[C_1, C_2]$. Here, $\mathbb{K}$ is a centered Gaussian process with covariance $\mathrm{Cov}\left(\mathbb{K}(t), \mathbb{K}(t')\right) = \Gamma_{d_\mathcal{X}}(t,t')$ *and $\Gamma_{d_\mathcal{X}}$ is as defined in Lemma B.7.*

Proof. Let $\mathscr{F} = \{f_t(x,y) = \mathbb{1}_{\{d_\mathcal{X}(x,y)\leq t\}} : t \in [C_1, C_2]\}$. As we have already noted in Remark G.3, we can consider $U_n(t), U(t)$ and $U^*_{n_B}(t)$ as processes indexed by $f_t \in \mathscr{F}$. With this observation the proof of the theorem's statement is straight forward and follows by an application of Theorem G.4. Therefore, we have to ensure $\mathscr{F}$ is a permissible function class and that the requirements (i)-(iii) of Theorem G.2 are met.

We begin by remarking that we have already shown in the proof of Lemma B.7 that $\mathscr{F}$ is a permissible function class.

Next, we check the assumption (i). Since $\mathscr{F}$ is a is a VC-subgraph class, as seen in the proof of Lemma B.7, we can conclude using Remark G.3 that requirement (i) is fulfilled.

The observations made in Remark G.3 are also helpful for the verification of (ii). Obviously, we have that

$$\sup_{f_t \in \mathscr{F}} |f_t(x,y)| \leq 1$$

for all $x, y \in \mathcal{X}$, i.e., $F \equiv 1$. Thus, Remark G.3 implies that also (ii) is given.

The final assumption (iii) is given by a combination of Remark F.7 and Corollary F.8, whose requirements are fulfilled in this setting.

In consequence, Theorem G.4 is applicable, which gives

$$\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}} - U_n^{\mathscr{F}}\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}^{\mathscr{F}}.$$

in $\ell^\infty(\mathscr{F})$. We have already observed (cf. Remark G.3), that this statement is equivalent to

$$\sqrt{n_B}\left(U_{n_B}^{*} - U_n\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}$$

in $\ell^\infty[C_1, C_2]$. Thus, we have proven the claim. □

Next, we come to the second step of the previously presented strategy. Based on the convergence of the process $\mathbb{U}_{n_B}^*$, we derive the next theorem applying the delta method for the bootstrap [23, Thm. 12.1] in combination with Lemma F.9.

Theorem D.3. *Suppose that Condition D.1 is met and let* $\sqrt{n_B} = o(n)$*. Then, it follows*

$$\sqrt{n_B}\left(\left(U_{n_B}^{*}\right)^{-1} - U_n^{-1}\right) \underset{M}{\overset{P}{\rightsquigarrow}} -\frac{1}{u \circ U^{-1}} \mathbb{K} \circ U^{-1} \overset{\mathcal{D}}{=} \mathbb{G}$$

in $\ell^\infty[\beta, 1-\beta]$*, where* $\mathbb{G}$ *is a centered Gaussian process with covariance defined in (11) in the paper.*

Proof. Let $D[C_1, C_2]$ denote the space of all cádlág functions on $[C_1, C_2]$ and let $\mathbb{D}_1$ be the set of all restrictions of distribution functions on $\mathbb{R}$ to $[C_1, C_2]$. We realize that the statement follows by an application of the delta method for the bootstrap [23, Thm. 12.1] with the inverse functional

$$\phi_{inv} : \mathbb{D}_1 \subset D[C_1, C_2] \to \ell^\infty[\beta, 1-\beta], \quad F \mapsto F^{-1}$$

to the bootstrap empirical process $\mathbb{U}_{n_B}^*$. In order to apply the delta method [23, Thm. 12.1] in this setting, we have to verify that

(i) the inverse functional ϕ_{inv} is Hadamard differentiable tangentially to $C[C_1, C_2]$ with derivative $\alpha \mapsto -(\alpha/f) \circ F^{-1}$;

(ii) the empirical U-distribution U_n and the bootstrap empirical U-distribution $U^*_{n_B}$ take values in $\mathbb{D}_1$;

(iii) it holds

$$\sqrt{n}\,(U_n - U) \rightsquigarrow \mathbb{K}$$

in $\ell^\infty[C_1, C_2]$, where $\mathbb{K}$ is the Gaussian process defined in Theorem D.2;

(iv) the Gaussian process $\mathbb{K}$ takes values in $C[C_1, C_2]$ and is tight in the function space $\ell^\infty[C_1, C_2]$;

(v) the maps $M_n \mapsto h\left(U^*_{n_B}\right)$ are measurable for every $h \in C_b\left(\ell^\infty\left[C_1, C_2\right]\right)$ outer almost surely (cf. Example G.1);

(vi) we have

$$\sqrt{n}\left(U^*_{n_B} - U_n\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}$$

in $\ell^\infty[C_1, C_2]$.

In the following, we check that all the above requirements are fulfilled.

Concerning the first point, we realize that the assumptions of Lemma F.9 are given. Thus, we conclude that the inversion functional ϕ_{inv} is indeed Hadamard differentiable tangentially to $C[C_1, C_2]$ with the stated derivative.

The definitions of U_n and $U^*_{n_B}$ suggest that they are distribution functions. Thus, U_n and $U^*_{n_B}$ are in $\mathbb{D}_1$ when restricted to $[C_1, C_2]$, i.e., the second requirement is fulfilled.

We obtain (iii) by Corollary F.8, which is applicable as $\mathscr{F} = \{f_t(x, y) = \mathbb{1}_{\{d_{\mathcal{X}}(x,y) \le t\}} : t \in [C_1, C_2]\}$ is a permissible VC-subgraph class (see the proof of Lemma B.7).

The Gaussian process $\mathbb{K}$ is almost surely continuous under the assumptions made (cf. Corollary F.8). This additionally guarantees that $\mathbb{K}$ is a tight random variable in $\ell^\infty[C_1, C_2]$, as it takes values in the Polish subspace $C[C_1, C_2]$ [40, Lemma 1.3.2]. Hence, also the fourth requirement is met.

Since all $h \in C_b\left(\ell^\infty[C_1, C_2]\right)$ are continuous, i.e., Borel-measurable, (v) holds, if we can show that the maps $M_n \mapsto \sqrt{n_B}(U^*_{n_B} - U_n)$ are measurable outer almost surely for each $n \in \mathbb{N}$. Let $n \in \mathbb{N}$ be arbitrary but fixed. Considering the weights M_n, we realize that they take values in the set

$$\mathbb{M}_n = \left\{ M_n = (M_{n1}, \dots) \in \mathbb{N}^\infty : \sum_{i=1}^{n} M_{ni} = n_B \text{ and } M_{ni} = 0 \text{ for } i > n \right\},$$

which is a finite subset of the countable space $\mathbb{N}^\infty$. As usual $\mathbb{N}^\infty$ is endowed with its projection σ-field. In the following this σ-field is denoted by $\mathcal{N}_\infty$. Now, as $\mathcal{N}_\infty$ contains all possible subsets of $\mathbb{M}_n$, it follows that every function on $\mathbb{M}_n$ is measurable, i.e., particularly the map $M_n \mapsto \sqrt{n_B}(U^*_{n_B} - U_n)$. Since $n \in \mathbb{N}$ was arbitrary, we get that the maps $M_n \mapsto \sqrt{n_B}(U^*_{n_B} - U_n)$ are measurable outer almost surely for $n \in \mathbb{N}$.

Finally, it follows under the assumptions made that $\sqrt{n_B}\left(U^*_{n_B} - U_n\right) \overset{P}{\underset{M}{\rightsquigarrow}} \mathbb{K}$ in $\ell^\infty[C_1, C_2]$ (cf. Theorem D.2). This yields (vi).

As all its requirements are fulfilled, the delta method for the bootstrap [23, Thm. 12.1] suggests that

$$\sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1} - U_n^{-1}\right) = \sqrt{n_B}\left(\phi_{inv}\left(U^*_{n_B}\right) - \phi_{inv}\left(U_n\right)\right) \overset{P}{\underset{M}{\rightsquigarrow}} -\frac{\mathbb{K}\circ U^{-1}}{u \circ U^{-1}} \overset{\mathcal{D}}{=} \mathbb{G}$$

in $\ell^\infty[\beta, 1-\beta]$. This yields the claim. □

We come to the third and final step of the mentioned strategy. We process this step by using a continuous mapping theorem for the bootstrap, namely Theorem 10.8 of Kosorok [23].

Theorem D.4. *Assume that Condition D.1 is fulfilled and let $\sqrt{n_B} = o(n)$. Then, it follows*

$$\int_\beta^{1-\beta} \left(\left(\mathbb{U}^*_{n_B}\right)^{-1}(t)\right)^2 dt \overset{P}{\underset{M}{\rightsquigarrow}} \Xi.$$

Proof. The statement follows by a conjunction of Theorem D.3 and Theorem 10.8 of Kosorok [23]. We remark that the conditions for Theorem D.3 are given by assumption, i.e., we have that

$$\sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1} - U_n^{-1}\right) \overset{P}{\underset{M}{\rightsquigarrow}} \mathbb{G} \tag{D.2}$$

in $\ell^\infty[\beta, 1-\beta]$. As shown by Lemma F.11, $\mathbb{G}$ takes almost surely values in $C[\beta, 1-\beta]$, which is a separable subspace of $\ell^\infty[\beta, 1-\beta]$. Thus, we can conclude that $\mathbb{G}$ is a tight random variable [40, Lemma 1.3.2].

For the application of Theorem 10.8 of Kosorok [23] we have to further ensure that the map

$$\varphi : \ell^\infty[\beta, 1-\beta] \to \mathbb{R},\ f \mapsto \int f(x)^2\, dx$$

is continuous and that the maps (cf. Example G.1)

$$M_n \mapsto h\left(\sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1} - U_n^{-1}\right)\right) \tag{D.3}$$

are measurable for every $h \in C_b(\ell^\infty[\beta, 1-\beta])$ outer almost surely. While the continuity of φ is obvious, the measurability of the maps defined in (D.3) can be established with the same arguments as in the proof of Theorem D.3. Consequently, we have by the conjunction of [23, Thm. 10.8] and (D.2) that

$$\int_\beta^{1-\beta} \left(\sqrt{n_B}\left(\left(U^*_{n_B}\right)^{-1}(t) - U_n^{-1}(t)\right)\right)^2 dt = \varphi\left(\left(\mathbb{U}^*_{n_B}\right)^{-1}\right) \underset{M}{\overset{P}{\rightsquigarrow}} \varphi(\mathbb{G}) = \int_\beta^{1-\beta} (\mathbb{G}(t))^2 \, dt,$$

as claimed. □

Remark D.5. Let $\beta \in (0, 1/2)$. As the map

$$\varphi_p : \ell^\infty[\beta, 1-\beta] \times \ell^\infty[\beta, 1-\beta] \to \mathbb{R}, \ (f, g) \mapsto \int_\beta^{1-\beta} |f(x) - g(x)|^p \, dx$$

is continuous for any $p \in [1, \infty)$, we immediately obtain that

$$\int_\beta^{1-\beta} \left|\left(\mathbb{U}^*_{n_B}\right)^{-1}(t)\right|^p dt \underset{M}{\overset{P}{\rightsquigarrow}} \Xi_p := \int_\beta^{1-\beta} |\mathbb{G}(t)|^p \, dt.$$

Hence, the quantiles Ξ_p of can be estimated consistently by the analogue bootstrap scheme.

E Detailed Simulation Results

In this section, we gather the detailed results of the numerical experiments performed in Section 4.2 and Section 5.1 of the paper.

E.1 Details of Section 4.2

In the following, we state the full results of the comparison of the space $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ with the spaces $(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})$, $1 \le i \le 6$, performed in Section 4.2. We recall that these spaces are constructed based on the evaluations of a 2-Wasserstein geodesic between the the uniform distribution on a 3D-pentagon and the uniform distribution on a torus at t_i, $1 \le i \le 6$ (for the precise construction see Section 4.2 and for an illustration see Figure 5 in Section 4.2 of the main document).

E.1.1 Empirical Power

First, we state the full results of our numerical study of the empirical power of the test Φ^*_{DoD} in this setting.

Tab. E.1: Comparison of different metric measure spaces I: The empirical power of the DoD-test Φ^*_{DoD} (1000 replications, see Section 3 of the paper for a formal definition of the test) for the comparison of the metric measure space $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ to the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Figure 5 in Section 4.2 of the paper for illustrations) for different n.

Φ^*_{DoD}						
Sample Size	$t_1 = 0$	$t_2 = 0.1$	$t_3 = 0.2$	$t_4 = 0.4$	$t_5 = 0.6$	$t_6 = 1$
100	0.007	0.012	0.033	0.289	0.774	0.998
250	0.018	0.076	0.335	0.952	1.000	1.000
500	0.038	0.236	0.763	1.000	1.000	1.000
1000	0.045	0.473	0.988	1.000	1.000	1.000

The results show that the test Φ^*_{DoD} is conservative for small sample sizes and that it nevertheless discriminates almost always between the metric measure spaces $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ and $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{j=3}^6$ for $n \geq 500$.

E.1.2 Distribution of Independent Distances

We give the detailed results for the comparison of the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$, when considering only the independent distances.

Tab. E.2: Comparison of different metric measure spaces II: The empirical power of the test based on $\widehat{D}_{\beta,ind}$ (1000 applications, see Section 4.2 for a formal definition) for the comparison of of the metric measure space $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ to the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Figure 5 in Section 4.2 of the paper for illustrations) for different n.

$\Phi_{D_{ind}}$						
Sample Size	$t_1 = 0$	$t_2 = 0.1$	$t_3 = 0.2$	$t_4 = 0.4$	$t_5 = 0.6$	$t_6 = 1$
100	0.039	0.042	0.054	0.073	0.118	0.227
250	0.045	0.045	0.066	0.139	0.270	0.738
500	0.046	0.061	0.090	0.248	0.548	0.991
1000	0.052	0.066	0.119	0.507	0.919	1.000

Table E.2 illustrates the additional power that we gain by considering all available distances (instead of only the independent ones) and carefully handling the occurring dependencies.

E.1.3 Influence of β

Further, we state the complete result of our investigation of the influence of β.

Tab. E.3: The influence of β: The empirical power of the DoD-test Φ^*_{DoD} (1000 replications, see Section 3 of the paper for a formal definition of the test) for the comparison of the metric measure space $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ to the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Figure 5 in Section 4.2 of the paper for illustrations) for $n = 500$ and different β.

Φ^*_{DoD}						
β	$t_1 = 0$	$t_2 = 0.1$	$t_3 = 0.2$	$t_4 = 0.4$	$t_5 = 0.6$	$t_6 = 1$
0	0.029	0.218	0.766	0.965	1.000	1.000
0.01	0.038	0.214	0.758	0.960	1.000	1.000
0.05	0.040	0.231	0.793	0.972	0.999	1.000
0.25	0.027	0.235	0.808	0.980	1.000	1.000

As discussed in Section 4.2 of the main document, it is apparent that the choice of the parameter β has a small but not significant impact on the comparison of the metric measure spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$.

E.2 Details of Section 5.1

In this subsection, we state detailed results of the comparison of the test Φ^*_{DoD} (see Section 3 of the paper) with the test Φ_{DTM} proposed by Brécheteau [9].

E.2.1 Torus-Pentagon Comparison

First, we give the complete results of the comparison of the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Section 4.2 of the paper for a formal definition) based on Φ_{DTM}.

Tab. E.4: Comparison of different metric measure spaces III: The empirical power of the test Φ_{DTM} (1000 applications) for the comparison of of the metric measure space $(\mathcal{W}_1, d_{\mathcal{W}_1}, \mu_{\mathcal{W}_1})$ to the spaces $\{(\mathcal{W}_i, d_{\mathcal{W}_i}, \mu_{\mathcal{W}_i})\}_{i=1}^6$ (see Figure 5 in Section 4.2 of the paper for illustrations) for different n and κ ($n_S = n/15$).

$\kappa = 0.05,\ \alpha = 0.05$						
Sample Size	$t_1 = 0$	$t_2 = 0.1$	$t_3 = 0.2$	$t_4 = 0.4$	$t_5 = 0.6$	$t_6 = 1$
100	0.046	0.067	0.059	0.056	0.053	0.053
250	0.054	0.057	0.070	0.052	0.062	0.063
500	0.065	0.059	0.063	0.072	0.068	0.069
1000	0.067	0.089	0.080	0.087	0.111	0.118
$\kappa = 0.1,\ \alpha = 0.05$						
Sample Size	$t_1 = 0$	$t_2 = 0.1$	$t_3 = 0.2$	$t_4 = 0.4$	$t_5 = 0.6$	$t_6 = 1$
100	0.040	0.051	0.052	0.064	0.059	0.058
250	0.066	0.065	0.069	0.069	0.062	0.064
500	0.064	0.086	0.083	0.118	0.091	0.099
1000	0.078	0.085	0.092	0.165	0.207	0.189

In particular, we observe that while the test Φ_{DTM} approximately holds its prespecified significance level for both choices of κ, it develops next to no power in this setting.

E.2.2 Spiral Comparison

Brécheteau [9] illustrated the empirical power of Φ_{DTM} for the comparison of different spiral types (see Figure E.1). In order to compare the two tests, we have applied Φ^*_{DoD} in the same setting. For the ease of readability, we give the construction of the spirals and additionally report the values of Brécheteau [9] for this framework (cf. Table E.5). Let $R \sim U[0, 1]$ be uniformly distributed and independent of $S, S' \overset{i.i.d.}{\sim} N(0, 1)$. Choose a significance level of $\alpha = 0.05$ and let $\beta = 0.01$. For $v = 10, 15, 20, 30, 40, 100$ we simulate samples of

$$(R\sin(vR) + 0.03S, R\cos(vR) + 0.03S') \sim \mu_v \tag{E.1}$$

and consider these to be samples from a metric measure spaces equipped with the Euclidean metric. We apply Φ^*_{DoD} with quantiles based on μ_v in order to compare μ_v with μ_v (based on different samples of size 2000) and μ_v with μ_{10}, $v = 15, \ldots, 100$. The results are presented in Table E.5.

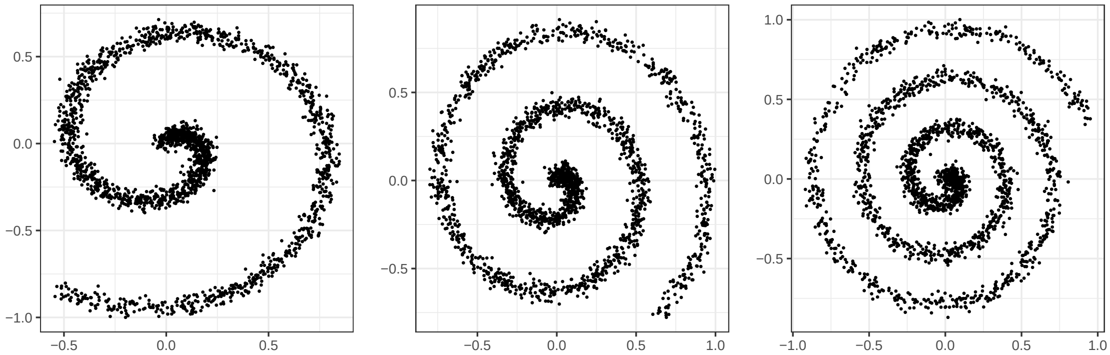

Fig. E.1: Different metric measure spaces II: Representation of samples created by (E.1) for $v = 10, 15, 20$ (from left to right).

Tab. E.5: Spiral Comparison: Empirical significance level and power of Φ^*_{DoD} and Φ_{DTM} for the comparisons the metric measure spaces represented in Figure E.1.

Φ^*_{DoD}					
v	15	20	30	40	100
Type-I error	0.036	0.051	0.051	0.054	0.048
Emp. power	1	1	1	1	1
Φ_{DTM}					
v	15	20	30	40	100
Type-I error	0.043	0.049	0.050	0.051	0.050
Emp. power	0.525	0.884	0.987	0.977	0.985

We observe that while both tests approximately hold their significance level, the test Φ^*_{DoD} differentiates more easily between the different spiral types.

E.2.3 Protein Structure Comparison

Finally, we state the full results of the protein structure comparison based on Φ_{DTM}.

Tab. E.6: Protein Comparison II: The empirical power of the test proposed by Brécheteau [9] (1000 applications) for the comparison of proteins represented in Figure 2 in Section 1.3 of the paper for different n and κ ($n_S = n/5$).

$\kappa = 0.05$, $\alpha = 0.05$			
n	5D0U vs 5D0U	5D0U vs 5JPT	5D0U vs 6FAA
100	0.055	0.068	0.109
250	0.049	0.080	0.297
500	0.037	0.090	0.690
$\kappa = 0.1$, $\alpha = 0.05$			
n	5D0U vs 5D0U	5D0U vs 5JPT	5D0U vs 6FAA
100	0.068	0.061	0.166
250	0.056	0.084	0.420
500	0.047	0.104	0.760

For the protein shape comparison, we observe that the test Φ_{DTM} approximately holds its significance level and is, for our choices of κ, more sensitive to the small local discrepancies between the structures 5D0U and 5JPT in comparison to Φ^*_{DoD}. However, it develops significantly less power for the comparison of 5D0U and 6FAA than Φ^*_{DoD}.

Part II

Supplement B: Auxiliary Results

In the second part of the supplement we gather several helpful, technical results and establish some apparently well known facts for which the the authors failed to find complete proofs. More precisely, this part of the supplement consists of three sections. In the first one, we derive distributional limits of the empirical U-quantile process, in the second one we investigate the bootstrap for empirical U-processes and in the final one we gather several technical result with a focus on measurability issues.
Throughout this part "$\mathbb{E}^*$" designates outer expectation, "$\mathbb{P}^*$" means outer probability and "$\rightsquigarrow$" stands for weak convergence in the sense of Hoffman-Jørgensen (see van der Vaart and Wellner [40, Part 1]). Let T be an arbitrary set. Then, the space $\ell^\infty(T)$ denotes the usual space off all uniformly bounded, $\mathbb{R}$-valued functions on T and $\ell^1(T)$ the one off all integrable, $\mathbb{R}$-valued functions on T. Moreover, $C(T)$ and $C_b(T)$ stand for the spaces of real valued, continuous and real valued, continuous, bounded functions on T, respectively. Let $(\mathcal{X}, \mathcal{M}, \mu)$ denote a measure space. Similarly to the previous definitions, we denote by $\ell^p(\mu)$ the space of all real valued functions on $\mathcal{X}$ that are p-integrable with respect to μ.

F U- and U-Quantile Processes

The connection of the statistic $\widehat{DoD}_{(\beta)}$ to empirical U- and empirical U-quantile processes has already been pointed out in Section 1. In the following, we aim to establish the distributional limit theorems that lay the foundation for the derivation of Theorem 2.6 and Theorem 2.7 (see Section B).

We start with the introduction of the basic notation and concepts required to develop the subsequent theory. Let $X, Y, X_1, \dots, X_n$ be an independent, identically distributed sample from a probability distribution P with values in a space S.

Definition F.1. [33] We call a measurable and bounded function $h : S \times S \times \mathbb{R} \to \mathbb{R}$, which is symmetric in the first two arguments and non-decreasing in the third, a *kernel function* if for all $x, y \in S$ it holds $\lim_{t\to\infty} h(x, y, t) = 1$ and $\lim_{t\to-\infty} h(x, y, t) = 0$. For $t \in \mathbb{R}$ we define the *U-distribution function* as

$$U(t) := \mathbb{E}\left[h(X, Y, t)\right]$$

and the *empirical U-distribution function* as

$$U_n(t) := \frac{2}{n(n-1)} \sum_{1\le i<j\le n} h(X_i, X_j, t).$$

Further, the *empirical U-process* is for $t \in \mathbb{R}$ given by

$$\mathbb{U}_n(t) = \sqrt{n}\left(U_n(t) - U(t)\right).$$

Remark F.2. For a measurable, symmetric function $g : S \times S \to \mathbb{R}$ one obtains the empirical cumulative distribution function of $\left(g\left(X_i, X_j\right)\right)_{1\le i<j\le n}$ as empirical U-distribution by choosing $h(x, y, t) := \mathbb{1}_{\{g(x,y)\le t\}}$. Although the definition is more general, we will focus on empirical U-processes of this type.

In a short example we introduce the, at least in this note, most important kernel functions.

Example F.3. Let d_S be a metric on the space S. Then, $d_S : S \times S \to \mathbb{R}$ is a symmetric function. Consequently, it follows by Remark F.2, that

$$h(x, y, t) = \mathbb{1}_{\{d_S(x,y)\le t\}}$$

is a kernel function,

$$U(t) = \mathbb{E}\left[\mathbb{1}_{\{d_S(X,Y)\le t\}}\right] = \mathbb{P}\left(d_S(X, Y) \le t\right)$$

is a U-distribution function and

$$U_n(t) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathbb{1}_{\{d_S(X_i,X_j)\le t\}}$$

is the corresponding empirical U-distribution function.

Just as for the classical empirical process we further define the *empirical U-quantile process*.

Definition F.4. [43, Sec. 6.1] Let $h : S \times S \times \mathbb{R} \to \mathbb{R}$ be a kernel function. Let U and U_n be the corresponding U- and empirical U-distribution functions. Then, for $t \in (0,1)$ the *t-U-quantile* is defined as

$$U^{-1}(t) = \inf\{s \in \mathbb{R} : U(s) \geq t\}$$

and the *empirical t-U-quantile* as

$$U_n^{-1}(t) = \inf\{s \in \mathbb{R} : U_n(s) \geq t\}.$$

The *empirical U-quantile process* is for $t \in (0,1)$ given as

$$\mathbb{U}_n^{-1}(t) = \sqrt{n}\left(U_n^{-1}(t) - U^{-1}(t)\right).$$

The definitions of the empirical U-process and of the empirical U-quantile process essentially correspond to the ones of the classical empirical process and the classical empirical quantile process.

The goal of this part of the supplement is to verify that as $n \to \infty$ it holds under certain assumptions

$$\mathbb{U}_n^{-1} = \sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G} \tag{F.1}$$

in $\ell^\infty[p,q]$, $0 < p < q < 1$, and under slightly different assumptions that

$$\mathbb{U}_n^{-1} = \sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G} \tag{F.2}$$

in $\ell^1(0,1)$. Here, $\mathbb{G}$ denotes a centered Gaussian process. In order to prove both statements, we rely on the distributional limits of the empirical U-process $\mathbb{U}_n$ [3, 33] as well as the Hadamard differentiability of the inversion functional $\phi_{inv} : F \to F^{-1}$ considered as a function into $\ell^\infty[p,q]$, $0 < p < q < 1$, and into $\ell^1(0,1)$, respectively. Combining these results with the delta-method [40, Thm. 3.9.4] directly yields (F.1) and (F.2) . We remark that, although this line of proof seems to be well known [43], the authors failed to find a full proof of either (F.1) or (F.2). Thus, we demonstrate both statements in the following sections.
In Section F.1 we briefly recall some results on the distributional convergence of the empirical U-process and afterwards, in Section F.2, we prove the stated convergences via the delta-method.

F.1 Distributional Limits for the Empirical U-Process

Let $(S, \mathcal{S}, P)$ be a probability space and let $X, Y, X_1, \ldots, X_n$ be independent, identically distributed random variables with law P. Similar as the classical empirical process the empirical U-process has been considered as as indexed by function classes. Among other things this has lead to very general convergence results. Before we can state an for our purposes suitable result we have to introduce some notation.

In what follows $\mathscr{F}$ denotes a *permissible* class (cf. Definition H.4) of symmetric functions $f : S \times S \to \mathbb{R}$ with envelope F, i.e., F is a measurable function with $F(x_1, x_2) \geq \sup_{f \in \mathscr{F}} |f(x_1, x_2)|$ for all $x_1, x_2 \in S$. The *U-distribution indexed by $f \in \mathscr{F}$* is given as

$$U^{\mathscr{F}}(f) = \mathbb{E}\left[f(X,Y)\right],$$

and the *empirical U-distribution indexed by $f \in \mathscr{F}$* as

$$U_n^{\mathscr{F}}(f) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} f(X_i, X_j).$$

Further, we denote by $\mathbb{U}_n^{\mathscr{F}} = \sqrt{n}\left(U_n^{\mathscr{F}} - U^{\mathscr{F}}\right)$ the *empirical U-process indexed by $\mathscr{F}$*.

Depending on $\mathscr{F}$ the process $\mathbb{U}_n^{\mathscr{F}}$ can be extremely general and hard to investigate. However, it turns out that, similar to the classical empirical process, also the empirical U-process is easy to handle if it is indexed by a so called VC-subgraph class.

Definition F.5. [36, Chap. II]

1. We say a collection of subsets $\mathscr{D}$ of the sample space S *picks out* a certain subset of the finite set $\{s_1, \ldots, s_n\} \subset S$ if it can be written as $\{s_1, \ldots, s_n\} \cap D$ for some $D \in \mathscr{D}$. The collection $\mathscr{D}$ is said to *shatter* $\{s_1, \ldots, s_n\}$ if $\mathscr{D}$ picks out each of its 2^n subsets. If there exists a finite n such that no set of size n is shattered by $\mathscr{D}$, $\mathscr{D}$ is a *VC-class*.
2. We denote $\mathscr{F}$ as *VC-subgraph class*, if the set $\mathscr{V} = \{\text{subgraph}(f) : f \in \mathscr{F}\}$, where

   $$\text{subgraph}(f) := \{(x, y, s) \in S \times S \times \mathbb{R} : 0 < s < f(x,y) \text{ or } 0 > s > f(x,y)\},$$

   forms a VC-class.

With the above notations and definitions we can finally state a basic convergence result for $\mathbb{U}_n^{\mathscr{F}}$ derived by Arcones and Giné [3].

Theorem F.6. [3, Thm. 4.9] *Let $\mathscr{F}$ be a permissible class of symmetric functions $f : S \times S \to \mathbb{R}$ satisfying the following conditions:*

(i) $\mathscr{F}$ is a VC-subgraph class with $\mathbb{E}\left[(F(X,Y))^2\right] < \infty$.

(ii) For all $x_1, x_2 \in S$ it holds

$$\sup_{f \in \mathscr{F}} \left| \int f(x_1, y)\, dP(y) - U^{\mathscr{F}}(f) \right| < \infty \text{ and } \sup_{f \in \mathscr{F}} \left| f(x_1, x_2) - U^{\mathscr{F}}(f) \right| < \infty.$$

Then, it follows for $n \to \infty$ that

$$\mathbb{U}_n^{\mathscr{F}} = \sqrt{n}\left(U_n^{\mathscr{F}} - U^{\mathscr{F}}\right) \rightsquigarrow \mathbb{K}^{\mathscr{F}}$$

in $\ell^\infty(\mathscr{F})$, where $\mathbb{K}^{\mathscr{F}}$ is a centered Gaussian process with covariance

$$\begin{aligned}\operatorname{Cov}\left(\mathbb{K}^{\mathscr{F}}(f_1), \mathbb{K}^{\mathscr{F}}(f_2)\right) = 4\Bigg(&\int\int f_1(x,y)\, dP(y) \int f_2(x,y)\, dP(y)dP(x)\\ &- \int\int f_1(x,y)\, dP(y)dP(x) \int\int f_2(x,y)\, dP(y)dP(x)\Bigg).\end{aligned}$$

Remark F.7.

1. The formulation of the above theorem differs slightly from the one in Arcones and Giné [3]. We demand $\mathscr{F}$ to be permissible instead of *image admissible Suslin*. Image admissibility Suslin is another regularity restriction on the class $\mathscr{F}$ that is slightly weaker than permissibility (cf. Section H.3).

2. For $\mathscr{F} = \left\{f_t(x,y) = \mathbb{1}_{\{g(x,y)\le t\}} \;:\; t \in [C_1, C_2] \subseteq \mathbb{R}\right\}$ we have a one-to-one correspondence between $t \in [C_1, C_2]$ and $f_t \in \mathscr{F}$, namely

   $$t \longleftrightarrow f_t(x,y) = \mathbb{1}_{\{g(x,y)\le t\}}.$$

   Thus, it is natural to identify the space $\ell^\infty(\mathscr{F})$ with the for our purposes more natural space $\ell^\infty[C_1, C_2]$. Further, it holds

   $$U_n^{\mathscr{F}}(f_t) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathbb{1}_{\{g(X_i,X_j)\le t\}} = U_n(t)$$

   and

   $$U^{\mathscr{F}}(f_t) = \mathbb{E}\left[\mathbb{1}_{\{g(X,Y)\le t\}}\right] = U(t).$$

3. Let $\sup_{x_1\in S, x_2\in S} |F(x_1,x_2)| \le C$ for a finite constant $C > 0$. Then, it obviously holds that

   $$\mathbb{E}\left[(F(X,Y))^2\right] \le \mathbb{E}\left[C^2\right] < \infty.$$

   Furthermore, the same argument yields that in this case the requirement (ii) of Theorem F.6 is trivially fulfilled.

For $\mathscr{F} = \left\{(x,y) \mapsto \mathbb{1}_{\{g(x,y)\le t\}} : t \in [C_1, C_2]\right\}$ a combination of Theorem F.6, Remark F.7 and Theorem 7.1 of Dudley [17] (to establish the continuity of the limiting Gaussian process) immediately yields the subsequent corollary.

Corollary F.8. *Let $h(x,y,t) \coloneqq \mathbb{1}_{\{g(x,y)\le t\}}$ be a kernel function as in Remark F.2 such that the function class $\mathscr{F} = \{(x,y) \mapsto h(x,y,t) : t \in [C_1, C_2]\}$ is a permissible VC-subgraph class and let the corresponding U-distribution function U be continuously differentiable on $[C_1, C_2]$. Then, it follows for $n \to \infty$ that*

$$\mathbb{U}_n = \sqrt{n}\,(U_n - U) \rightsquigarrow \mathbb{K}$$

in $\ell^\infty[C_1, C_2]$, where $\mathbb{K}$ is a centered, continuous Gaussian process with covariance

$$\begin{aligned}\mathrm{Cov}\left(\mathbb{K}_1(t)\mathbb{K}_1(t')\right) = 4\Gamma_{\mathbb{K}}(t,t') \coloneqq 4\Bigg(&\int\int h(x,y,t)\,dP(y)\int h(x,y,t')\,dP(y)\,dP(x)\\ &-\int\int h(x,y,t)\,dP(y)\,dP(x)\int\int h(x,y,t')\,dP(y)\,dP(x)\Bigg).\end{aligned}$$

F.2 Distributional Limits for the Empirical U-Quantile Processes

Next, we come to the distributional limits of the empirical U-quantile process. We consider the inversion functional $\phi_{inv} : F \mapsto F^{-1}$ as a map from the set of restricted distribution functions into the space $\ell^\infty[p,q]$, for given $0 < p < q < 1$, and as a map from the set of compactly supported distribution functions into the space $\ell^1(0,1)$, respectively. In both settings, we verify that ϕ_{inv} is Hadamard differentiable given certain (different) conditions. Then, we employ the delta-method for Hadamard differentiable functionals [40, Theorem 3.9.4] to derive (F.1) and (F.2). Before we come to this, we recall Hadamard differentiability and introduce some notation.

We recall: A map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$, where $\mathbb{D}$ and $\mathbb{E}$ are normed spaces, is *Hadamard-differentiable* at $\theta \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$, if there exists a continuous linear map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h)$$

as $n \to \infty$, for all converging sequences $t_n \to 0$ and $h_n \to h \in \mathbb{D}_0$, with $h_n \in \mathbb{D}$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all $n \ge 1$ sufficiently large [23, Sec. 2.2].

Given an interval $[a,b] \subset \mathbb{R}$, let $D[a,b]$ denote the space of cádlág functions on $[a,b]$ (equipped with the supremum norm), let $\mathbb{D}_1$ be the set of all restrictions of distribution functions on $\mathbb{R}$ to $[a,b]$ and let $\mathbb{D}_2$ be the subset of $\mathbb{D}_1$ of distribution functions of measures that concentrate on $(a,b]$. First, we consider the inversion functional ϕ_{inv} as a map from $\mathbb{D}_1 \subset D[a,b] \to \ell^\infty[p,q]$, $0 < p < q < 1$.

Lemma F.9. [40, Lemma 3.9.23] *Let $0 < p < q < 1$, and let F be continuously differentiable on the interval $[a,b] = [F^{-1}(p) - \epsilon, F^{-1}(q) + \epsilon]$ for some $\epsilon > 0$, with strictly positive derivative f. Then the inversion functional $\phi_{inv} : F \mapsto F^{-1}$ as a map $\mathbb{D}_1 \subset D[a,b] \to$*

$\ell^\infty[p,q]$ is Hadamard-differentiable at F tangentially to $C[a,b]$. The derivative is the map $\alpha \mapsto -(\alpha/f) \circ F^{-1}$.

Under slightly different assumptions, we have already verified in Section B.5.2 that ϕ_{inv} is Hadamard differentiable as a map from $\mathbb{D}_2 \subset D[a,b] \to \ell^1(0,1)$. For the ease of readability, we recall the corresponding lemma.

Lemma B.19. *Let F have compact support on $[a,b]$ and let F be continuously differentiable on its support with derivative f that is strictly positive on (a,b) (Possibly, $f(a) = 0$ and/or $f(b)) = 0$). Then the inversion functional $\phi_{inv} : F \mapsto F^{-1}$ as a map $\mathbb{D}_2 \subset D[a,b] \to \ell^1(0,1)$ is Hadamard-differentiable at F tangentially to $C[a,b]$. The derivative is the map $\alpha \mapsto -(\alpha/f) \circ F^{-1}$.*

Now, we are able to verify (F.1) and (F.2), which are essential for proving Theorem 2.6 and Theorem 2.7 (cf. Section B).

Theorem F.10. *Let $h(x,y,t) := \mathbb{1}_{\{g(x,y)\leq t\}}$ be a kernel function as in Remark F.2 such that the function class $\mathscr{F} = \{(x,y) \mapsto h(x,y,t) : t \in [C_1,C_2]\}$ is a permissible VC-subgraph class. Further, let the corresponding U-distribution function U be continuously differentiable on the interval $[C_1,C_2] = [U^{-1}(p)-\epsilon, U^{-1}(q)+\epsilon]$ for some $\epsilon > 0$ with strictly positive density u. Then, as the sample size n grows to infinity, it holds that*

$$\mathbb{U}_n^{-1} = \sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G}$$

in $\ell^\infty[p,q]$, where $\mathbb{G}$ is a centered Gaussian process with covariance

$$\operatorname{Cov}\left(\mathbb{G}(t), \mathbb{G}(t')\right) = \frac{4}{(u \circ U^{-1}(t))(u \circ U^{-1}(t'))} \Gamma_{\mathbb{K}}(U^{-1}(t), U^{-1}(t')).$$

Here, $\Gamma_{\mathbb{K}}$ is as defined in Corollary F.8

Proof. We realize that the statement follows by an application of Theorem 3.9.4 of van der Vaart and Wellner [40] with the inverse functional

$$\phi_{inv} : \mathbb{D}_1 \subset \ell^\infty[C_1,C_2] \to \ell^\infty[p,q], \;\; F \mapsto F^{-1}$$

to the empirical U-process $\mathbb{U}_n$. In order to apply it in this setting, we have to verify that

(*i*) the inverse functional ϕ_{inv} is Hadamard differentiable tangentially to $C[C_1,C_2]$ with derivative $\alpha \mapsto -(\alpha/f) \circ F^{-1}$;

(*ii*) the empirical U-distribution U_n takes values in $\mathbb{D}_1$;

(*iii*) it holds
$$\sqrt{n}\,(U_n - U) \rightsquigarrow \mathbb{K}$$
in $\ell^\infty[C_1,C_2]$, where $\mathbb{K}$ is the Gaussian process defined in Corollary F.8;

(*iv*) the Gaussian process $\mathbb{K}$ takes its values in the subspace $C[C_1, C_2]$ and is tight in $\ell^\infty[C_1, C_2]$.

In the following, we check that all the above requirements are fulfilled.

Concerning (*i*), we realize that the assumptions of Lemma F.9 (*i*) are given. Thus, we conclude that the inversion functional ϕ_{inv} is Hadamard-differentiable tangentially to $C[C_1, C_2]$ with the stated derivative.

The empirical U-distribution function is a distribution function by definition. Thus, U_n is in $\mathbb{D}_1$ when restricted to $[C_1, C_2]$, i.e., (*ii*) is fulfilled.

We obtain (*iii*) by Corollary F.8, whose requirements are given by assumption.

Finally, we come to (*iv*). The Gaussian process $\mathbb{K}$ is almost surely continuous (cf. Corollary F.8). This additionally guarantees that $\mathbb{K}$ is a tight random variable in $\ell^\infty[C_1, C_2]$, as it takes values in the Polish subspace $C[C_1, C_2]$ [40, Lemma 1.3.2]. Hence, also the fourth requirement is met.

As all its requirements are fulfilled, it follows by Theorem 3.9.4 of van der Vaart and Wellner [40] that

$$\sqrt{n}\left(U_n^{-1} - U^{-1}\right) = \sqrt{n}\left(\phi_{inv}\left(U_n\right) - \phi_{inv}\left(U\right)\right) \rightsquigarrow -\frac{\mathbb{K} \circ U^{-1}}{u \circ U^{-1}} \stackrel{\mathcal{D}}{=} \mathbb{G}$$

in $\ell^\infty[p, q]$. This yields the claim. □

It is easily verified that the Gaussian process $\mathbb{G}$ is continuous on $[p, q]$.

Lemma F.11. *Assume the setting of Theorem F.10. Then, the process $\mathbb{G}$ is continuous on $I := [p, q]$.*

Proof. Let $\mathbb{K}$ be a centered Gaussian process with covariance structure

$$\mathrm{Cov}\left(\mathbb{K}(s), \mathbb{K}(t)\right) = \Gamma_{\mathbb{K}}(s, t).$$

Then, the Gaussian process $\mathbb{K}$ is continuous almost surely (cf. Corollary F.8). By assumption the function $g(t) = \frac{1}{u(t)}$ is continuous and bounded on the interval $[C_1, C_2]$. Consequently, also the centered Gaussian proecess $g(t)\mathbb{K}(t)$, whose covariance is given as

$$\mathrm{Cov}\left(g(s)\mathbb{K}(s), g(t)\mathbb{K}(t)\right) = g(s)g(t)\mathrm{Cov}\left(\mathbb{K}(s), \mathbb{K}(s)\right) = \frac{1}{u(s)u(t)}\Gamma_{\mathbb{K}}(s, t),$$

is almost surely continuous. If we now time rescale $g(t)\mathbb{K}(t)$ with $t = U^{-1}(s)$ for $s \in I \subseteq [p, q]$, the process

$$\widetilde{\mathbb{K}}(s) := \frac{1}{(u \circ U^{-1}(s))}\mathbb{K}\left(U^{-1}(s)\right)$$

is a continuous Gaussian process, as U^{-1} is continuous and strictly increasing on I. To be precise, $\widetilde{\mathbb{K}}$ is a centered Gaussian process continuous on I with covariance structure

$$\mathrm{Cov}\left(\widetilde{\mathbb{K}}(s), \widetilde{\mathbb{K}}(s')\right) = \mathrm{Cov}(\mathbb{G}(s), \mathbb{G}(s')).$$

Hence, there exists a version of $\mathbb{G}$ that is continuous on I. □

Finally, we come to the proof of (F.2). By employing the analogous arguments as in the proof of Theorem F.10 and using Lemma B.19 instead of Lemma F.9, we directly obtain the subsequent result.

Theorem F.12. *Let $h(x, y, t) \coloneqq \mathbb{1}_{\{g(x,y) \le t\}}$ be a kernel function as in Remark F.2 such that the function class $\mathscr{F} = \{(x, y) \mapsto h(x, y, t) : t \in [C_1, C_2]\}$ is a permissible VC-subgraph class. Further, let the corresponding U-distribution function U be supported on $[C_1, C_2]$ and continuously differentiable with density u. Let u be strictly positive on (C_1, C_2). Then, as the sample size grows to infinity, it holds that*

$$\mathbb{U}_n^{-1} = \sqrt{n}\left(U_n^{-1} - U^{-1}\right) \rightsquigarrow \mathbb{G}$$

in $\ell^1(0, 1)$, where $\mathbb{G}$ is the Gaussian process defined in Theorem F.10.

G Introduction of the Bootstrap

In this section we gather various result about the bootstrap. In the first subsection, we introduce the general bootstrap and recap the for our purposes most important results. Afterwards, we concentrate on the specific resampling scheme for the empirical U-process that we use in Section 3 to approximate the quantiles required for testing.

G.1 The General Bootstrap

It is a common statistical problem that the limiting distribution of a statistic of interest is intractable. To carry out inference on the underlying quantity, one possibility consists of using a *bootstrap* or *resampling scheme*. The generic setup considered in this subsection is as follows. Let $(\mathbb{D}, d)$ be a metric space. Suppose that we are interested in approximating the limit distribution of the $\mathbb{D}$-valued statistic $\mathbb{X}_n$ that is based on the observations $\mathcal{X}_n = \{X_1, \dots, X_n\}$ (i.e., $\mathbb{X}_n = \mathbb{X}_n(\mathcal{X}_n)$). It is noteworthy that we only assume that $X_1, \dots, X_n$ are measurable maps from the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into some measurable space and that they may be dependent. A *bootstrap replicate* of $\mathbb{X}_n$ is denoted by $\mathbb{X}_n^* = \mathbb{X}_n^*(\mathcal{X}_n, M_n)$, where $M_n \in \mathbb{R}^n$ is a random weight vector living on the probability space $(\Pi, \mathcal{F}, \mathbb{Q})$ independent of $\mathcal{X}_n$. This set up is very general and it encompasses for example the resampling scheme for the empirical U-process presented in Example G.1 as well as several more general ones introduced in Kosorok [23, Chap. 10].

Example G.1. Let $(S, \mathcal{S}, P)$ be a probability space. Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} P$ and let P_n denote the corresponding empirical measure. Let $h : S \times S \times \mathbb{R} \to \mathbb{R}$ be a kernel function, denote by $U(t)$ the corresponding U-distribution function and by $U_n(t)$ the corresponding empirical one as defined in Definition F.1. Draw n_B times with replacement from the set of sample values $\mathcal{X}_n = \{X_1, \dots, X_n\}$ and denote the resampled (bootstrap) values by $X_1^*, \dots, X_{n_B}^*$. Then, conditionally on $\mathcal{X}_n$, the $X_1^*, \dots, X_{n_B}^*$ are independent and identically distributed with distribution P_n. The *bootstrap empirical U-distribution* $U_{n_B}^*$ is given for $t \in \mathbb{R}$ as

$$U_{n_B}^*(t) \coloneqq \frac{2}{n_B(n_B-1)} \sum_{1 \le i < j \le n_B} h\left(X_i^*, X_j^*, t\right), \tag{G.1}$$

where the X_i's in the definition of U_n are replaced by their bootstrap replicates. Let M_{ni} be the number of times that X_i is "redrawn" from the original sample such that the vector $M_n = (M_{n1}, \dots, M_{nn})$ is multinomial distributed on $\mathcal{X}_n$ with parameters n_B and (probabilities) $1/n, \dots, 1/n$. Then, we can write

$$U_{n_B}^*(t) = U_{n_B}^*(t; \mathcal{X}_n, M_n) = \frac{2}{n_B(n_B-1)} \Bigg(\sum_{1 \le i < j \le n} M_{ni} M_{nj} h(X_i, X_j, t) \\ + \sum_{1 \le i \le n} \frac{M_{ni}(M_{ni}-1)}{2} h(X_i, X_i, t) \Bigg). \tag{G.2}$$

The corresponding *bootstrap empirical U-process* is given as $\mathbb{U}_{n_B}^* = \sqrt{n_B}\big(U_{n_B}^* - U_n\big)$. Furthermore, the *bootstrap empirical t-U-quantile* is defined for $t \in (0,1)$ as

$$\left(U_{n_B}^*\right)^{-1}(t) = \left(U_{n_B}^*\right)^{-1}(t; \mathcal{X}_n, M_n) = \inf\left\{s \in \mathbb{R} \middle|\, U_{n_B}^*(s) \ge t\right\}$$

and the *bootstrap empirical U-quantile process* as $\left(\mathbb{U}_{n_B}^*\right)^{-1} = \sqrt{n_B}\Big(\left(U_{n_B}^*\right)^{-1} - U_n^{-1}\Big)$.

Obviously, prior to the use of a resampling scheme, its consistency should be demonstrated. In the introduced setting there are two sources of randomness, the observed data and the resampling done by the bootstrap. Because of these two sources of randomness, convergence of conditional laws is assessed in a slightly different manner than usual weak convergence. We note that $\mathbb{X}_n \rightsquigarrow \mathbb{X}$ in the metric space $(\mathbb{D}, d)$ if and only if

$$\sup_{f \in \mathrm{BL}_1} \left|\mathbb{E}^* f(\mathbb{X}_n) - \mathbb{E} f(\mathbb{X})\right| \to 0,$$

where BL_1 is the space of all real functions f on $\mathbb{D}$ with $\|f\|_\infty \le 1$ and $|f(x) - f(y)| \le d(x,y)$ for all $x, y \in \mathbb{D}$ [40, Sec. 1.12]. Based on this alternative definition of weak convergence one can define convergence of conditional limit laws for bootstrap replicates. Let $\mathbb{X}_n^*$

be a sequence of bootstrap replicates with values in $\mathbb{D}$ and random weights M_n. For some tight, $\mathbb{D}$-valued random variable $\mathbb{X}$, we use from now on the notation $\mathbb{X}_n^* \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{X}$ to express that

$$\sup_{h \in \mathrm{BL}_1} |\mathbb{E}_M^* h(\mathbb{X}_n^*) - \mathbb{E}h(\mathbb{X})| \overset{\mathbb{P}^*}{\to} 0 \text{ and } \mathbb{E}_M\left(h(\mathbb{X}_n^*)\right)^\star - \mathbb{E}_M\left(h(\mathbb{X}_n^*)\right)_\star \overset{\mathbb{P}^*}{\to} 0,$$

for all $h \in \mathrm{BL}_1$ as $n \to \infty$. Here, the subscript M indicates conditional expectation over the weights M_n given the data and BL_1 is defined as previously. Further, $(h(X_n^*))^\star$ and $(h(X_n^*))_\star$ denote minimal measurable majorants and maximal measurable minorants with respect to the joint data $(\mathcal{X}_n, M_n)$ (see Section H and Kosorok [23], van der Vaart and Wellner [40] for further details).

G.2 The Bootstrap for the Empirical U-Process

Let $(S, \mathcal{S}, P)$ be a probability space and let $X, Y, X_1, \ldots, X_n$ be independent, identically distributed random variables with law P. Let P_n be the empirical measure based on the sample $X_1, \ldots, X_n$. Conditionally on the sample, let $X_1^*, \ldots, X_{n_B}^*$ be an independent, identically distributed sample from P_n. Further, denote by $P_{n_B}^*$ the empirical distribution of $X_1^*, \ldots, X_{n_B}^*$ and let $\mathscr{G}$ be a class of measurable function on S. Then, the *bootstrap empirical process indexed by* $g \in \mathscr{G}$ is defined as

$$\mathbb{F}_{n_B}^*(g) = \sqrt{n_B}\left(\int g\, dP_{n_B}^* - \int g\, dP_n\right).$$

It is well known that $\mathbb{F}_{n_B}^*$ indexed by $\mathscr{G}$ converges in $\ell^\infty(\mathscr{G})$ to a tight Brownian bridge process for almost all $X_1, X_2, \ldots$, if $\mathscr{G}$ is a Donsker class and $\int \left(\left\|f - \int f\, dP\right\|_{\mathscr{G}}^2\right)^\star dP < \infty$ [40, Thm. 3.6.2]. Here, $\|H\|_{\mathscr{G}} = \sup_{g \in \mathscr{G}} |H(g)|$ for $H \in \ell^\infty(\mathscr{G})$ and "$\star$" denotes a maximal measurable majorant.

Indeed one can define the *bootstrap empirical U-process indexed by* (a function class) $\mathscr{F}$ in a similar manner and a comparable statement can be shown under stricter assumptions on the indexing function class. This result is the foundation of the proofs in Section D. However, before we can state it, we need to introduce some concepts and notation.

In what follows $\mathscr{F}$ denotes a *permissible* class (cf. Definition H.4) of symmetric functions $f : S \times S \to \mathbb{R}$ with envelope F, i.e., F is a measurable function with $F(x_1, x_2) \geq \sup_{f \in \mathscr{F}} |f(x_1, x_2)|$ for all $x_1, x_2 \in S$. Given a pseudometric space (T, d), the ϵ-covering number $N(\epsilon, T, d)$ is defined as

$$N(\epsilon, T, d) = \min\{n : \text{There esists a covering of } T \text{ by } n \text{ balls of } d\text{-radius} \leq \epsilon\}.$$

For a given probability measure μ on $(S \times S, \mathcal{S} \otimes \mathcal{S})$ and $f_1, f_2 \in \mathscr{F}$ we define the pseudometric

$$||f_1 - f_2||_{L^2(\mu)} = \sqrt{\int (f_1 - f_2)^2 \, d\mu}.$$

Based on this pseudometric and the previous definition of covering numbers we define the quantity

$$N_2(\epsilon, \mathscr{F}, \mu) = N\left(\epsilon, \mathscr{F}, ||\cdot||_{L^2(\mu)}\right).$$

The *U-distribution indexed by $f \in \mathscr{F}$* is given as

$$U^{\mathscr{F}}(f) = \mathbb{E}\left[f(X, Y)\right],$$

the *empirical U-distribution indexed by $f \in \mathscr{F}$* as

$$U_n^{\mathscr{F}}(f) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} f(X_i, X_j)$$

and the *empirical V-distribution indexed by $f \in \mathscr{F}$* as

$$V_n^{\mathscr{F}}(f) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} f(X_i, X_j).$$

We denote by $U_{n_B}^{*,\mathscr{F}}(f)$ the *bootstrap empirical U-distribution indexed by $f \in \mathscr{F}$* that is defined analogously to the bootstrap empirical U-distribution in (G.1). Moreover, we call $\mathbb{U}_n^{\mathscr{F}} = \sqrt{n}\left(U_n^{\mathscr{F}} - U^{\mathscr{F}}\right)$ the *empirical U-process* and $\mathbb{U}_{n_B}^{*,\mathscr{F},U} = \sqrt{n}\left(U_{n_B}^{*,\mathscr{F}} - U_n^{\mathscr{F}}\right)$ the *the bootstrap empirical U-process indexed by $\mathscr{F}$.*

Theorem G.2. [4, Thm. 2.1] *Let $\mathscr{F}$ be a permissible class of symmetric functions $f : S \times S \to \mathbb{R}$ satisfying the following conditions:*

(i) There is a function $\lambda : (0, \infty) \to [0, \infty)$ with $\int_0^\infty \lambda(v) dv < \infty$ such that for each probability measure μ with $\mu F^2 = \int_{S \times S} F^2 \, d\mu < \infty$ it follows

$$\left(\log N_2\left(v(\mu F^2)^{1/2}, \mathscr{F}, \mu\right)\right)^{1/2} \le \lambda(v), \quad v > 0.$$

(ii) For all $1 \le i_1, i_2 \le n$ we have

$$\mathbb{E}\left|F(X_{i_1}, X_{i_2})\right|^{\mathrm{Card}(\{i_1, i_2\})} < \infty,$$

where $\mathrm{Card}(A)$ denotes the cardinality of the finite set A.

(iii) It holds that

$$\mathbb{U}_n^{\mathscr{F}} = \sqrt{n}\left(U_n^{\mathscr{F}} - U^{\mathscr{F}}\right) \rightsquigarrow \mathbb{K}^{\mathscr{F}}$$

in $\ell^\infty(\mathscr{F})$, where $\mathbb{K}^{\mathscr{F}}$ is a centered Gaussian process indexed by $\mathscr{F}$ with covariance

$$\begin{aligned}\operatorname{Cov}\left(\mathbb{K}^{\mathscr{F}}(f_1), \mathbb{K}^{\mathscr{F}}(f_2)\right) = 4\Bigg(&\int\int f_1(x,y)\,dP(y)\int f_2(x,y)\,dP(y)dP(x)\\ &-\int\int f_1(x,y)\,dP(y)dP(x)\int\int f_2(x,y)\,dP(y)dP(x)\Bigg).\end{aligned}$$

Then, if $n \to \infty$ and $n_B \to \infty$,

$$\mathbb{U}_{n_B}^{*,\mathscr{F},V} = \sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}} - V_n^{\mathscr{F}}\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}^{\mathscr{F}} \tag{G.3}$$

in $\ell^\infty(\mathscr{F})$.

Remark G.3.

1. The formulation of the above theorem differs slightly from the one in Arcones and Giné [4]. We demand $\mathscr{F}$ to be permissible instead of *image admissible Suslin*. Image admissibility Suslin is another measurability restriction on the class $\mathscr{F}$ that is slightly weaker than permissibility (cf. Section H.3).

2. For $\mathscr{F} = \left\{f_t(x,y) = \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t\}} \,:\, t \in [C_1, C_2] \subseteq \mathbb{R}\right\}$ we have a one-to-one correspondence between $t \in [C_1, C_2]$ and $f_t \in \mathscr{F}$, namely

$$t \longleftrightarrow f_t(x,y) = \mathbb{1}_{\{d_{\mathcal{X}}(x,y)\le t\}}.$$

   Thus, it is natural to identify the space $\ell^\infty(\mathscr{F})$ with the for our purposes more natural space $\ell^\infty[C_1, C_2]$. Further, it follows

$$U_n^{\mathscr{F}}(f_t) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \mathbb{1}_{\{d_{\mathcal{X}}(X_i,X_j)\le t\}} = U_n(t).$$

   Obviously, the analogue equalities hold for $U^{\mathscr{F}}(f_t)$ and $U_{n_B}^{*,\mathscr{F}}(f_t)$.

3. The Assumption (i) is fulfilled if $\mathscr{F}$ is a *VC-subgraph class*, i.e., if $\{\text{subgraph}(f) : f \in \mathscr{F}\}$ is a VC-class (cf. Definition F.5). Here, $\text{subgraph}(f) \coloneqq \{\,(x,t) \in S \times \mathbb{R} : 0 < t < f(x) \text{ or } 0 > t > f(x)\}$.

   Proof. This statement requires a short proof. If $\mathscr{F}$ is a VC-subgraph class, then there are finite constants A and c such that for each probability measure μ with $\mu F^2 < \infty$,

$$N_2(\epsilon, \mathscr{F}, \mu) \le A((\mu F^2)^{1/2}/\epsilon)^{-c} \tag{G.4}$$

[36, Lemma II.36]. This implies that for $v \leq 1$ we have

$$\left(\log N_2\left(v(\mu F^2)^{1/2}, \mathscr{F}, \mu\right)\right)^{1/2} \leq \left(\log\left(\frac{A}{v^c}\right)\right)^{1/2}.$$

Since it holds that

$$\begin{aligned}\int_0^1 \left(\log \frac{A}{v^c}\right)^{1/2} dv &= \int_0^1 \left(\log A + \log \frac{1}{v^c}\right)^{1/2} dv \\ &\leq (\log A)^{1/2} + \int_0^1 \left(\log \frac{1}{v^c}\right)^{1/2} dv \\ &= (\log A)^{1/2} + \frac{\sqrt{\pi}\sqrt{c}}{2} < \infty,\end{aligned}$$

the bound (G.4) controls the covering numbers for $v \leq 1$ as required. Obviously, this is the difficult part, as for v increasing the covering number $N_2\left(v(\mu F^2)^{1/2}, \mathscr{F}, \mu\right)$ is monotonically decreasing. Thus, we find that the function $\lambda : (0, \infty) \to [0, \infty)$

$$\lambda(v) = \begin{cases} \left(\log\left(\frac{A}{v^c}\right)\right)^{1/2} & \text{if } v \leq 1, \\ 0 & \text{if } v > 1. \end{cases}$$

meets the requirements of Assumption (ii). □

4. Suppose that $\sup_{x_1 \in S, x_2 \in S} |F(x_1, x_2)| \leq C$ for a finite constant $C > 0$. Then, for all $1 \leq i_1, i_2 \leq n$ it holds

$$\mathbb{E}\left|F(X_{i_1}, X_{i_2})\right|^{\mathrm{Card}(\{i_1, i_2\})} \leq \mathbb{E}\left|C\right|^{\mathrm{Card}(\{i_1, i_2\})} < \infty.$$

Thus, Requirement (ii) is fulfilled.

Theorem G.2 guarantees that the resampled process $\mathbb{U}_{n_B}^{*,\mathscr{F},V} = \sqrt{n_B}\big(U_{n_B}^{*,\mathscr{F}} - V_n^{\mathscr{F}}\big)$ indeed converges to the limit, if it exists, of $\mathbb{U}_n^{\mathscr{F}}$. However, regarding, for example, the delta-method for the bootstrap [23, Thm. 12.1], we would rather have the same statement for $\mathbb{U}_{n_B}^{*,\mathscr{F},U} = \sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}} - U_n^{\mathscr{F}}\right)$. Fortunately, this result is a simple consequence of Theorem G.2.

Theorem G.4. *Let the assumptions of Theorem G.2 be met. Further, let $|F(x_1, x_2)| \leq C < \infty$ for all $x_1, x_2 \in S$ and let $\sqrt{n_B} = o(n)$. Then, we have that*

$$\mathbb{U}_{n_B}^{*,\mathscr{F},U} = \sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}} - U_n^{\mathscr{F}}\right) \underset{M}{\overset{P}{\rightsquigarrow}} \mathbb{K}^{\mathscr{F}}$$

in $\ell^\infty(\mathscr{F})$.

Proof. We need to show that the empirical V-distribution in (G.3) can be replaced by the empirical U-distribution. To this end, we demonstrate in a first step that for any given $X_1, \ldots, X_n$ we have that

$$\begin{aligned}&\sup_{f\in\mathscr{F}}\left|\mathbb{U}_{n_B}^{*,\mathscr{F},V}(f)-\mathbb{U}_{n_B}^{*,\mathscr{F},U}(f)\right|\\ =&\sup_{f\in\mathscr{F}}\left|\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}(f)-V_n^{\mathscr{F}}(f)\right)-\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}(f)-U_n^{\mathscr{F}}(f)\right)\right|\to 0\end{aligned}$$

for $n\to\infty$. Afterwards, we verify that this indeed yields the claim.

Regarding the definitions of $V_n^{\mathscr{F}}(f)$ and $U_n^{\mathscr{F}}(f)$, we realize that for each $X_1, \ldots, X_n$ given and $f\in\mathscr{F}$ it holds

$$\begin{aligned}V_n^{\mathscr{F}}(f)=&\frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n}f(X_i,X_j)+\frac{1}{n^2}\sum_{i=1}^{n}f(X_i,X_i)\\ =&\frac{n-1}{n}U_n^{\mathscr{F}}(f)+\frac{1}{n^2}\sum_{1\leq i\leq n}f(X_i,X_i).\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}\left|U_n^{\mathscr{F}}(f)-V_n^{\mathscr{F}}(f)\right|&=\left|U_n^{\mathscr{F}}(f)-\left(\frac{n-1}{n}U_n^{\mathscr{F}}(f)+\frac{1}{n^2}\sum_{1\leq i\leq n}f(X_i,X_i)\right)\right|\\ &=\left|\frac{1}{n}U_n^{\mathscr{F}}(f)-\frac{1}{n^2}\sum_{1\leq i\leq n}f(X_i,X_i)\right|\\ &=\left|\frac{2}{n^2(n-1)}\sum_{1\leq i<j\leq n}f(X_i,X_j)-\frac{1}{n^2}\sum_{1\leq i\leq n}f(X_i,X_i)\right|.\end{aligned}$$

With this we can prove that the distance between $\mathbb{U}_{n_B}^{*,\mathscr{F},V}$ and $\mathbb{U}_{n_B}^{*,\mathscr{F},U}$ in $\ell^\infty(\mathscr{F})$ can be bounded for any collection $X_1, \ldots, X_n$ given. Therefore, we realize that for any $X_1, \ldots, X_n$

it holds

$$\begin{aligned}
&\sup_{f\in\mathscr{F}}\left|\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-V^{\mathscr{F}}_n(f)\right)-\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-U^{\mathscr{F}}_n(f)\right)\right|\\
&=\sup_{f\in\mathscr{F}}\left|\sqrt{n_B}\left(U^{\mathscr{F}}_n(f)-V^{\mathscr{F}}_n(f)\right)\right|\\
&=\sup_{f\in\mathscr{F}}\sqrt{n_B}\left|\frac{2}{n^2(n-1)}\sum_{1\le i<j\le n} f(X_i,X_j)-\frac{1}{n^2}\sum_{1\le i\le n} f(X_i,X_i)\right|\\
&\le\sup_{f\in\mathscr{F}}\sqrt{n_B}\left|\frac{2}{n^2(n-1)}\sum_{1\le i<j\le n} f(X_i,X_j)\right|+\sup_{f\in\mathscr{F}}\sqrt{n_B}\left|\frac{1}{n^2}\sum_{1\le i\le n} f(X_i,X_i)\right|\\
&\le\sup_{f\in\mathscr{F}}\sqrt{n_B}\left|\frac{2}{n^2(n-1)}\sum_{1\le i<j\le n} C\right|+\sup_{f\in\mathscr{F}}\sqrt{n_B}\left|\frac{1}{n^2}\sum_{1\le i\le n} C\right|=\frac{2C\sqrt{n_B}}{n},
\end{aligned}$$

where we used in the last line that $|f(x_1,x_2)|\le|F(x_1,x_2)|\le C<\infty$ for all $x_1,x_2\in S$ and $f\in\mathscr{F}$ by assumption. Since we demand that $\sqrt{n_B}=o(n)$, it follows for any given $X_1,\ldots,X_n$ that

$$\sup_{f\in\mathscr{F}}\left|\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-V^{\mathscr{F}}_n(f)\right)-\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-U^{\mathscr{F}}_n(f)\right)\right|\le\frac{2C\sqrt{n_B}}{n}\to 0 \tag{G.5}$$

for $n\to\infty$. This concludes the first step.

Next, we demonstrate that this indeed yields $\sqrt{n_B}(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n)\underset{M}{\overset{P}{\rightsquigarrow}}\mathbb{K}^{\mathscr{F}}$. Therefore, we have to show that

$$\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)-\mathbb{E}h\left(\mathbb{K}^{\mathscr{F}}\right)\right|\xrightarrow{\mathbb{P}^*}0 \tag{G.6}$$

and

$$\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)\right)^{\star}-\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)\right)_{\star}\xrightarrow{\mathbb{P}^*}0 \tag{G.7}$$

for any $h\in\mathrm{BL}_1$ as $n\to\infty$. Recall that $(h(\cdot))^{\star}$ and $(h(\cdot))_{\star}$ denote minimal measurable majorants and maximal measurable minorants with respect to the joint data $(\mathcal{X}_n,M_n)$ (see Lemma H.10 and Lemma H.9) and BL_1 designates the set of all real functions h on $\ell^\infty(\mathscr{F})$ with $\|h\|_\infty\le 1$ and $|h(x)-h(y)|\le\|x-y\|_{\mathscr{F}}:=\sup_{f\in\mathscr{F}}|x(f)-y(f)|$ for all $x,y\in\ell^\infty(\mathscr{F})$.

Concerning (G.6), we realize

$$\begin{aligned}
&\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)-\mathbb{E}h\left(\mathbb{K}^{\mathscr{F}}\right)\right|\\
&=\sup_{h\in \mathrm{BL}_1}\Bigg|\mathbb{E}^*_M\Bigg[h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)-h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)\\
&\qquad\qquad+h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)\Bigg]-\mathbb{E}h\left(\mathbb{K}^{\mathscr{F}}\right)\Bigg|\\
&\le\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M\left[h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)-h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)\right]\right|\\
&+\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)-\mathbb{E}h\left(\mathbb{K}^{\mathscr{F}}\right)\right|.
\end{aligned}$$

Next, we consider both summands of the last inequality separately.

For the first summand it holds

$$\begin{aligned}
&\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M\left[h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)-h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)\right]\right|\\
&\overset{(i)}{\le}\mathbb{E}^*_M\left[\sup_{f\in\mathscr{F}}\left|\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-U^{\mathscr{F}}_n(f)\right)-\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}(f)-V^{\mathscr{F}}_n(f)\right)\right|\right]\\
&\overset{(ii)}{\le}\mathbb{E}^*_M\left[\frac{2C\sqrt{n_B}}{n}\right]=\frac{2C\sqrt{n_B}}{n}\to 0,
\end{aligned}$$

for $n\to\infty$. Here, (i) follows as all $h\in \mathrm{BL}_1$ are Lipschitz continuous with Lipschitz constant 1 and (ii) is induced by (G.5).
Next, we come to the second term. Since all its requirements are fulfilled by assumption, Theorem G.2 is applicable and we obtain

$$\sup_{h\in \mathrm{BL}_1}\left|\mathbb{E}^*_M h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-V^{\mathscr{F}}_n\right)\right)-\mathbb{E}h\left(\mathbb{K}^{\mathscr{F}}\right)\right|\overset{\mathbb{P}^*}{\to}0$$

for $n\to\infty$. The conjunction of our findings yields (G.6).

Finally, we come to the verification of (G.7). By definition it holds that

$$\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)\right)^{\star}-\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U^{*,\mathscr{F}}_{n_B}-U^{\mathscr{F}}_n\right)\right)\right)_{\star}\ge 0 \tag{G.8}$$

almost surely. Thus, if we can bound the difference in (G.8) almost surely from above with an expression that converges to zero in outer probability, then we can deduce (G.7).

Therefore, we recognize that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-U_n^{\mathscr{F}}\right)\right)\right)^\star\\
=&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)+h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)^\star\\
\leq&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)^\star+\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)^\star
\end{aligned}$$

almost surely, where we used in the last step that for any two real valued maps X and Y it holds $(X+Y)^\star \leq (X)^\star + (Y)^\star$ (see Lemma H.11). Further, we have almost surely that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)^\star+\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)^\star\\
\leq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star+\mathbb{E}_M\left(\left|h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right|\right)^\star\\
\leq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star+\mathbb{E}_M\left(\sup_{f\in\mathscr{F}}\left|\mathbb{U}_{n_B}^{*,\mathscr{F},U}(f)-\mathbb{U}_{n_B}^{*,\mathscr{F},V}(f)\right|\right)^\star,
\end{aligned}$$

since h is Lipschitz continuous with Lipschitz constant one. Using (G.5) we conclude that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star+\mathbb{E}_M\left(\sup_{f\in\mathscr{F}}\left|\mathbb{U}_{n_B}^{*,\mathscr{F},U}(f)-\mathbb{U}_{n_B}^{*,\mathscr{F},V}(f)\right|\right)^\star\\
\leq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star+\mathbb{E}_M\left(\frac{2C\sqrt{n_B}}{n}\right)^\star\\
=&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star+\frac{2C\sqrt{n_B}}{n}
\end{aligned}$$

almost surely. On the other hand it holds almost surely that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-U_n^{\mathscr{F}}\right)\right)\right)_\star\\
=&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)+h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)_\star\\
\geq&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)_\star+\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)_\star,
\end{aligned}$$

where the last inequality follows since for any two real valued maps X and Y we have $(X+Y)_\star \geq (X)_\star + (Y)_\star$ (c.f. Lemma H.11). Moreover, we find that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)_\star+\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-V_n^{\mathscr{F}}\right)\right)\right)_\star\\
\geq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star-\mathbb{E}_M\left(\left|h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},U}\right)-h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right|\right)_\star\\
\geq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star-\mathbb{E}_M\left(\sup_{f\in\mathscr{F}}\left|\mathbb{U}_{n_B}^{*,\mathscr{F},U}(f)-\mathbb{U}_{n_B}^{*,\mathscr{F},V}(f)\right|\right)_\star
\end{aligned}$$

almost surely, where we used that h is Lipschitz continuous with Lipschitz constant one. With (G.5) we deduce that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star - \mathbb{E}_M\left(\sup_{f\in\mathscr{F}}\left|\mathbb{U}_{n_B}^{*,\mathscr{F},U}(f)-\mathbb{U}_{n_B}^{*,\mathscr{F},V}(f)\right|\right)_\star\\
\geq&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star - \mathbb{E}_M\left(\frac{2C\sqrt{n_B}}{n}\right)_\star\\
=&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star - \frac{2C\sqrt{n_B}}{n}
\end{aligned}$$

almost surely.

In conjunction, the previous results yield that

$$\begin{aligned}
&\mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-U_n^{\mathscr{F}}\right)\right)\right)^\star - \mathbb{E}_M\left(h\left(\sqrt{n_B}\left(U_{n_B}^{*,\mathscr{F}}-U_n^{\mathscr{F}}\right)\right)\right)_\star\\
\overset{(i)}{\leq}&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star + \frac{2C\sqrt{n_B}}{n} - \left(\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star - \frac{2C\sqrt{n_B}}{n}\right)\\
=&\mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)^\star - \mathbb{E}_M\left(h\left(\mathbb{U}_{n_B}^{*,\mathscr{F},V}\right)\right)_\star + \frac{4C\sqrt{n_B}}{n} \xrightarrow{\mathbb{P}^*} 0,
\end{aligned}$$

as $n\to\infty$. Here, (i) holds almost surely and the last line follows by Theorem G.2 as well as the observation we made in (G.5). Thus, we can conclude (G.7) and, as previously argued, this yields the claim. □

H Technical Results

In this final section we collect various results that find implicit or explicit usage in the course of this part.

H.1 Integrated Difference of Quantile Functions

Here, we state a result on the integrated difference of quantile functions that has been used multiple times in Section B.7.

Lemma H.1. *Let X and Y be non-negative, real valued, compactly supported random variables with distribution functions F and G. Let F^{-1} and G^{-1} denote the corresponding quantile functions. Then, it holds for $\kappa\in[0,1)$ that*

$$\int_0^\kappa F^{-1}(t)-G^{-1}(t)\,dt = \int_0^{F^{-1}(\kappa)} G(x)-F(x)\,dx + \int_{F^{-1}(\kappa)}^{G^{-1}(\kappa)} G(x)-\kappa\,dt. \tag{H.1}$$

In particular, for $\kappa = 1$, it follows that

$$\int_0^1 F^{-1}(t) - G^{-1}(t)\,dt = \int_0^\infty G(x) - F(x)\,dx.$$

Remark H.2. It is noteworthy that the roles of $G^{-1}(\kappa)$ and $F^{-1}(\kappa)$ in (H.1) are interchangeable, i.e., by the same line of proof we could also obtain that

$$\int_0^\kappa F^{-1}(t) - G^{-1}(t)\,dt = \int_0^{G^{-1}(\kappa)} G(t) - F(t)\,dt + \int_{G^{-1}(\kappa)}^{F^{-1}(\kappa)} \kappa - F(t)\,dt. \tag{H.2}$$

Proof. We start by showing the first part of the statement. We have that

$$\begin{aligned}&\int_0^\kappa F^{-1}(t) - G^{-1}(t)\,dt = \int_0^\kappa \int_0^\infty \mathbb{1}_{\{x\le F^{-1}(t)\}} - \mathbb{1}_{\{x\le G^{-1}(t)\}}\,dx\,dt\\ &= \int_0^\kappa \int_0^\infty \mathbb{1}_{\{F(x)\le t\}} - \mathbb{1}_{\{G(x)\le t\}}\,dx\,dt = \int_0^\infty \int_0^\kappa \mathbb{1}_{\{F(x)\le t\}} - \mathbb{1}_{\{G(x)\le t\}}\,dt\,dx,\end{aligned}$$

where the last step follows by the Theorem of Tonelli/Fubini [6, Thm. 18.3]. This yields that

$$\begin{aligned}&\int_0^\kappa F^{-1}(t) - G^{-1}(t)\,dt = \int_0^\infty (G(x)\wedge\kappa) - (F(x)\wedge\kappa)\,dx\\ &= \int_0^{F^{-1}(\kappa)} G(t) - F(t)\,dt + \int_{F^{-1}(\kappa)}^{G^{-1}(\kappa)} G(t) - \kappa\,dt.\end{aligned}$$

Here, the last equality is obvious if $G^{-1}(\kappa) > F^{-1}(\kappa)$. However, it also holds for $G^{-1}(\kappa) < F^{-1}(\kappa)$, since we take $\int_a^b f(x)\,dx := -\int_b^a f(x)\,dx$ when $a > b$.

Next, we come to the second part of the statement. By similar arguments as previously (using that X and Y are compactly supported), we obtain

$$\begin{aligned}&\int_0^1 F^{-1}(t) - G^{-1}(t)\,dt = \int_0^1 \int_0^\infty \mathbb{1}_{\{F(x)\le t\}} - \mathbb{1}_{\{G(x)\le t\}}\,dx\,dt\\ &= \int_0^\infty \int_0^1 \mathbb{1}_{\{F(x)\le t\}} - \mathbb{1}_{\{G(x)\le t\}}\,dt\,dx = \int_0^\infty G(x) - F(x)\,dx,\end{aligned}$$

which gives the second part of the claim. $\square$

H.2 Differentiation and Integration

The following lemma allows to pointwise interchange integration and differentiation.

Lemma H.3. *Given a measure space $(X, \mathcal{M}, \mu)$ and a real valued function f on $X \times (a, b)$ such that $f(\cdot, t) \in \ell^1(\mu)$ for each $t \in (a, b)$, let $F(t) = \int_X f(x, t)\, d\mu(x)$. Suppose that the partial derivative $\partial_t f = \partial f / \partial t$ exists and assume that there is a set $X' \subseteq X$ with $\mu(X') = \mu(X)$ such that there exits $g \in \ell^1(\mu)$ with $|\partial_s f(x, s)| \leq g(x)$ for all $x \in X'$ and $s \in [t - \epsilon, t + \epsilon]$ for some $\epsilon > 0$. Then, F is differentiable at t with*

$$F'(t) = \int_X \partial_t f(x, t)\, d\mu(x).$$

Proof. We observe that

$$\lim_{s \to t} \frac{F(t) - F(s)}{t - s} = \lim_{s \to t} \int_X \frac{f(x, t) - f(x, s)}{t - s}\, d\mu(x) = \lim_{s \to t} \int_{X'} \frac{f(x, t) - f(x, s)}{t - s}\, d\mu(x).$$

Since $f(x, \cdot)$ is differentiable at t for all $x \in X'$, we obtain for $|s - t| < \epsilon$ that

$$\frac{f(x, t) - f(x, s)}{t - s} = \partial_t f(x, \zeta)$$

for some $\zeta \in [t - \epsilon, t + \epsilon]$. By assumption, we have that there exists $g \in \ell^1(\mu)$ with $|\partial_s f(x, s)| \leq g(x)$ for all $x \in X'$ and $s \in [t - \epsilon, t + \epsilon]$. In consequence, the Dominated Convergence Theorem yields that

$$\lim_{s \to t} \frac{F(t) - F(s)}{t - s} = \int_{X'} \lim_{s \to t} \frac{f(x, t) - f(x, s)}{t - s}\, d\mu(x) = \int_{X'} \partial_t f(x, t)\, d\mu(x) = \int_X \partial_t f(x, t)\, d\mu(x).$$

□

H.3 Measurability of Function Classes

Depending on the complexity of the indexing function class $\mathscr{F}$, stochastic processes indexed by $\mathscr{F}$ can be extremely general and difficult to handle. In the following we define two possible regularity assumptions on $\mathscr{F}$.

We begin with the introduction of the concept of *permissible* function classes.

Definition H.4. [36, Appendix C, Def. 1] Let $(S, \mathcal{S})$ be a measurable space and let $\mathscr{F}$ be a real valued function class on S that is indexed by a parameter $t \in T$, i.e., $\mathscr{F} = \left\{ \tilde{f}(\cdot, t) : t \in T \right\}$. Further, let T be a separable metric space and $\mathscr{B}(T)$ the Borel σ-field on T. We call the class $\mathscr{F}$ *permissible* if it is indexed by T in such a way that

(i) the function $\tilde{f}(\cdot, \cdot)$ is $\mathcal{S} \otimes \mathscr{B}(T)$-measurable as a function from $S \times T$ into the real line;

(ii) T is an analytic subset of a compact metric space $\bar{T}$ (from which it inherits its metric and Borel σ-field).

Another possibility is to assume that the indexing function $\mathscr{F}$ class is a *image admissible Suslin* class.

Definition H.5. [18, Sec. 5.3] A separable, measurable space $(\Omega, \mathcal{A})$ is called a *Suslin space* if and only if there exists a Polish space Ψ and a Borel measurable map from Ψ to Ω.
If $(S, \mathcal{S})$ is a measurable space and F a set, then a real-valued function $\Phi : S \times F \to \mathbb{R}$, $(s, f) \mapsto \Phi(s, f)$ is called image admissible Suslin via $(\Omega, \mathcal{A}, V)$ if $(\Omega, \mathcal{A})$ is a Suslin measurable space, V is a function from Ω onto F, and $(s, \omega) \mapsto \Phi(s, V(\omega))$ is jointly measurable on $S \times \Omega$.
A class of real-valued functions $\mathscr{F}$ on S is denoted as *image admissible Suslin* if for $F = \mathscr{F}$ the map $\Phi(s, f) = f(s)$ is image admissible Suslin via some $(\Omega, \mathcal{A}, V)$ as defined above.

At a first glance, both introduced concepts seem similar. However, as shown by the next lemma, if a function class is permissible in the sense of Definition H.4, then it is automatically image admissible Suslin.

Lemma H.6. *Let $(S, \mathcal{S})$ be a measurable space and let $\mathscr{F}$ be a permissible function class on $(S, \mathcal{S})$. Then, $\mathscr{F}$ is image admissible Suslin.*

Proof. Since $\mathscr{F}$ is permissible it admits a representation as

$$\mathscr{F} = \left\{ \tilde{f}(\cdot, t) : t \in T \right\},$$

where $\tilde{f}$ is $\mathcal{S} \otimes \mathscr{B}(T)/\mathscr{B}(\mathbb{R})$-measurable as a function from $S \times T$ to $\mathbb{R}$, and T is an analytic subset of a compact metric space $\bar{T}$. By the theorem of Heine-Borel for metric spaces [10, Thm. 9.58] it follows that $\bar{T}$ is separable and complete, i.e., a Polish space. As $T \subseteq \bar{T}$ is an analytic set it is the image of a Polish space under a continuous, i.e., Borel measurable, mapping [15, Sec. 8.2]. Thus, $(T, \mathscr{B}(T))$ is a Suslin space.
By assumption, the map $\tilde{f}(\cdot, \cdot)$ is $\mathcal{S} \otimes \mathscr{B}(T)$-measurable. Consequently, the map $\Phi : S \times \mathscr{F} \to \mathbb{R}$, $\Phi(s, f) = \tilde{f}(s, t)$ is image admissible Suslin via $(T, \mathscr{B}(T), V)$, where

$$V : T \to \mathscr{F}, \quad t \mapsto \tilde{f}(\cdot, t).$$

□

In order to verify joint measurability of a function (as required to show permissibility of a function class) the subsequent theorem is often helpful. This theorem is well known in the theory of stochastic processes (see e.g. Capasso and Bakstein [13, Sec. 2]) and a proof is only added for the sake of completeness.

Theorem H.7. *Let $([C_1, C_2], \mathscr{B}([C_1, C_2]))$ be a real, bounded interval endowed with the Borel σ-algebra and let $(S, \mathcal{S})$ be an arbitrary measurable space. Let $f : S \times [C_1, C_2] \to \mathbb{R}$ be a function such that*

1. *$t \mapsto f(x,t)$ is right-/left-continuous for all $x \in S$;*
2. *$x \mapsto f(x,t)$ is measurable for all $t \in [C_1, C_2]$.*

Then, f is $\mathcal{S} \otimes \mathscr{B}([C_1, C_2])$-measurable.

Proof. Since both cases can be treated analogously, we only show the claim under the assumption $t \mapsto f(x,t)$ is right-continuous for all $x \in S$. Therefore, we define for $t \in [C_1 + k(C_2 - C_1)/2^n, C_1 + (k+1)(C_2 - C_1)/2^n)$, where $k = 0, \dots, 2^n - 1$, the function $f_n : S \times [C_1, C_2) \to \mathbb{R}$ as

$$f_n(x,t) = f\left(x, C_1 + \frac{(k+1)(C_2 - C_1)}{2^n}\right).$$

Further, we set for $t = C_2$

$$f_n(x,t) = f(x, C_2).$$

Next, we verify that f_n is for $n \in \mathbb{N}$ a $\mathcal{S} \otimes \mathscr{B}([C_1, C_2])$-measurable function. Therefore, let $A \in \mathscr{B}(\mathbb{R})$ and $n \in \mathbb{N}$ be fix. Then, we have

$$\begin{aligned} f_n^{-1}(A) &= \{(x,t) : f_n(x,t) \in A,\ C_1 \le t \le C_2\} \\ &= \bigcup_{k=0}^{2^n-1} \left[C_1 + \frac{k(C_2 - C_1)}{2^n}, C_1 + \frac{(k+1)(C_2 - C_1)}{2^n}\right) \\ &\qquad \times \left\{f\left(x, C_1 + \frac{(k+1)(C_2 - C_1)}{2^n}\right) \in A\right\} \\ &\quad \cup \{C_2\} \times \{f(x, C_2) \in A\}. \end{aligned}$$

We realize, that for $0 \le k \le 2^n - 1$ the set

$$F_k := \left[C_1 + \frac{k(C_2 - C_1)}{2^n}, C_1 + \frac{(k+1)(C_2 - C_1)}{2^n}\right)$$

belongs to $\mathscr{B}([C_1, C_2])$ and that the same holds for $F_{2^n} = \{C_2\}$. Furthermore, the sets

$$E_k := \left\{f\left(x, C_1 + \frac{(k+1)(C_2 - C_1)}{2^n}\right) \in A\right\},\ 0 \le k \le 2^n - 1$$

and $E_{2^n} = \{f(x, C_2) \in A\}$ are included in $\mathcal{S}$ as $f(\cdot, t)$ is measurable for $t \in [C_1, C_2]$ by assumption. This means that $E_k \times F_k \in \mathcal{S} \otimes \mathscr{B}([C_1, C_2])$, $0 \leq k \leq 2^n$ [6, Sec. 18]. Thus, for any $A \in \mathscr{B}(\mathbb{R})$ it holds that $f_n^{-1}(A)$ is a finite union of measurable sets, i.e., $f_n^{-1}(A) \in \mathcal{S} \otimes \mathscr{B}([C_1, C_2])$. Consequently, f_n^{-1} is a measurable function for $n \in \mathbb{N}$.
We have assumed that $f(x, \cdot)$ is right continuous for all $x \in S$. Thus, by construction it holds

$$\lim_{n \to \infty} f_n(x, t) = f(x, t)$$

for all $t \in [C_1, C_2]$. We conclude that f is the pointwise limit of a sequence of measurable functions and thus measurable itself [6, Thm. 13.4]. This yields the claim. □

Remark H.8. The theorem's claim can also be shown for $[C_1, C_2] = \mathbb{R}$.

H.4 Further Measurability Issues

To overcome potential measurability issues we work in Section G.2 with *minimal measurable majorants* and *maximal measurable minorants*, which we define next.

Lemma H.9. [23, Lemma 6.3] *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any map $X : \Omega \to \bar{\mathbb{R}}$ there exists a measurable map $(X)^\star : \Omega \to \bar{\mathbb{R}}$ with*

(i) $(X)^\star \geq X$*;*

(ii) For every measurable $U : \Omega \to \bar{\mathbb{R}}$ with $U \geq X$ almost surely, it holds $(X)^\star \leq U$ almost surely.

The map $(X)^\star$ is called *minimal measurable majorant.* Furthermore, the above lemma suggests that if X and Y are two arbitrary maps from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}$ and $X \leq Y$ almost surely, then it holds $(X)^\star \leq (Y)^\star$ almost surely. Analogously the following lemma holds.

Lemma H.10. [23, Lemma 6.4] *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any map $X : \Omega \to \bar{\mathbb{R}}$ there exists a measurable map $(X)_\star : \Omega \to \bar{\mathbb{R}}$ with*

(i) $(X)_\star \leq X$*;*

(ii) For every measurable $U : \Omega \to \bar{\mathbb{R}}$ with $U \leq X$ almost surely, it holds $(X)_\star \geq U$ almost surely.

The map $(X)_\star$ is called *maximal measurable minorant.* It can be equivalently defined as $(X)_\star = (-(-X))^\star$ [23, Lemma 6.3]. Moreover, also the previous lemma implies that for two arbitrary maps from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}$, it holds $(X)_\star \leq (Y)_\star$ almost surely, if $X \leq Y$ almost surely. Some further, useful properties of these functions are collected in the following lemma.

Lemma H.11. [23, Lemma 6.6] *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space. The following statements are true almost surely for arbitrary maps* $X, Y : \Omega \to \bar{\mathbb{R}}$*, provided that the statement is well defined:*

(i) $(X + Y)^\star \leq (X)^\star + (Y)^\star$*;*

(ii) $(X)_\star + (Y)_\star \leq (X + Y)_\star$.

References

[1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer Monographs in Mathematics. Springer-Verlag, 2007.

[2] H. Amann. *Ordinary differential equations: An introduction to nonlinear analysis*, volume 13. Walter de gruyter, 2011.

[3] M. A. Arcones and E. Giné. Limit theorems for U-processes. *The Annals of Probability*, pages 1494–1542, 1993.

[4] M. A. Arcones and E. Giné. U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Processes and their Applications*, 52(1):17–38, 1994.

[5] J. R. Berrendero, A. Cuevas, and B. Pateiro-López. Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146:237–247, 2016.

[6] P. Billingsley. *Probability and Measure.* John Wiley & Sons, 1979.

[7] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, 2013.

[8] S. Bobkov and M. Ledoux. *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*, volume 261. American Mathematical Society, 2019.

[9] C. Brécheteau. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature. *Electronic Journal of Statistics*, 13(1):795–849, 2019.

[10] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson. *Real Analysis.* ClassicalRealAnalysis. com, 1997.

[11] A. Bücher and I. Kojadinovic. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, pages 1–21, 2017.

[12] A. J. Cabo and A. J. Baddeley. Line transects, covariance functions and set convergence. *Advances in applied probability*, 27(3):585–605, 1995.

[13] V. Capasso and D. Bakstein. *An Introduction to Continuous-Time Stochastic Processes.* Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston, 2012.

[14] S. Chowdhury and F. Mémoli. The Gromov-Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.

[15] D. L. Cohn. *Measure Theory*, volume 165. Springer, 1980.

[16] G. David and S. Semmes. Uniform rectifiability and singular sets. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 13, pages 383–443. Elsevier, 1996.

[17] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, 1(3), 1967.

[18] R. M. Dudley. *Uniform Central Limit Theorems*, volume 142. Cambridge university press, 2014.

[19] L. Dümbgen. Empirische Prozesse. Lecture Notes, 2017.

[20] H. Federer. *Geometric measure theory.* Springer, 2014.

[21] B. Galerne. Computation of the perimeter of measurable sets via their covariogram. applications to random sets. *Image Analysis & Stereology*, 30(1):39–51, 2011.

[22] T. Kaji. Switching to the new norm: From heuristics to formal tests using integrable empirical processes. Technical report, Massachusetts Institute of Technology, 2018.

[23] M. R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference.* Springer, 2008.

[24] A. M. Mathai. *An introduction to geometrical probability: Distributional aspects with applications*, volume 1. CRC Press, 1999.

[25] G. Matheron. *Les Variables Régionalisées et leur Estimation: Une Application de la Théorie des Fonctions Aléatoires aux Sciences de la Nature.* Masson et CIE, 1965.

[26] P. Mattila and P. Saaranen. Ahlfors-David regular sets and bilipschitz maps. In *Annales Academiae Scientiarum Fennicae. Mathematica*, volume 34, pages 487–502. Suomalainen Tiedeakatemia, 2009.

[27] F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011.

[28] F. Mémoli and T. Needham. Distance distributions and inverse problems for metric measure spaces. *arXiv preprint arXiv:arXiv:1810.09646v3*, 2021.

[29] Q. Merigot and B. Thibert. Optimal transport: Discretization and algorithms. In *Handbook of Numerical Analysis*, volume 22, pages 133–212. Elsevier, 2021.

[30] D. Moltchanov. Distance distributions in random networks. *Ad Hoc Networks*, 10(6):1146–1166, 2012.

[31] F. Morgan. *Geometric measure theory: A beginner's guide.* Academic press, 2016.

[32] A. Munk and C. Czado. Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), 1998.

[33] D. Nolan and D. Pollard. Functional limit theorems for U-processes. *The Annals of Probability*, pages 1291–1298, 1988.

[34] J. Philip. The probability distribution of the distance between two random points in a box, 2007.

[35] D. Pokornỳ and M. Rauch. On the asymptotics of counting functions for Ahlfors regular sets. *arXiv preprint arXiv:1906.09600*, 2019.

[36] D. Pollard. *Convergence of Stochastic Processes.* Springer New York, softcover reprint of the original 1st ed. 1984 edition, 1984.

[37] P. Sidiropoulos. N-sphere chord length distribution. *arXiv preprint arXiv:1411.5639*, 2014.

[38] S. Smale and M. W. Hirsch. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic press, 1974.

[39] S. Troscheit. Fractal geometry. *Course notes*, 2020.

[40] A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag, 1996.

[41] C. Villani. *Topics in optimal transportation.* American Mathematical Soc., 2003.

[42] C. Villani. *Optimal Transport: Old and New.* Springer Science & Business Media, 2008.

[43] M. Wendler. U-quantile processes and generalized linear statistics of dependent data. *Dissertation*, 2011.

CHAPTER B

The Ultrametric Gromov-Wasserstein Distance

The Ultrametric Gromov-Wasserstein Distance

Facundo Mémoli* Axel Munk † ‡ Zhengchao Wan *
Christoph Alexander Weitkamp †

May 23, 2022

Abstract

In this paper, we investigate compact ultrametric measure spaces which form a subset $\mathcal{U}^w$ of the collection of all metric measure spaces $\mathcal{M}^w$. In analogy with the notion of the ultrametric Gromov-Hausdorff distance on the collection of ultrametric spaces $\mathcal{U}$, we define ultrametric versions of two metrics on $\mathcal{U}^w$, namely of Sturm's Gromov-Wasserstein distance of order p and of the Gromov-Wasserstein distance of order p. We study the basic topological and geometric properties of these distances as well as their relation and derive for $p = \infty$ a polynomial time algorithm for their calculation. Further, several lower bounds for both distances are derived and some of our results are generalized to the case of finite ultra-dissimilarity spaces. Finally, we study the relation between the Gromov-Wasserstein distance and its ultrametric version (as well as the relation between the corresponding lower bounds) in simulations and apply our findings for phylogenetic tree shape comparisons.

Keywords Ultrametric space, Gromov-Hausdorff distance, Gromov-Wasserstein distance, Optimal transport

1 Introduction

Over the last decade the acquisition of ever more complex data, structures and shapes has increased dramatically. Consequently, the need to develop meaningful methods for comparing general objects has become more and more apparent. In numerous applications, e.g. in molecular biology [17, 43, 54], computer vision [45, 61] and electrical engineering [55, 77], it is important to distinguish between different objects in a pose invariant manner:

*Department of Mathematics and Department of Computer Science and Engineering, Ohio State University, 100 Math Tower, 231 W 18th Ave Columbus, OH 43210

†Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen

‡Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen

two instances of the a given object in *different* spatial orientations are deemed to be equal. Furthermore, also the comparisons of graphs, trees, ultrametric spaces and networks, where mainly the underlying connectivity structure matters, have grown in importance [21, 29]. One possibility to compare two general objects in a pose invariant manner is to model them as metric spaces (X, d_X) and (Y, d_Y) and regard them as elements of the collection of isometry classes of compact metric spaces denoted by $\mathcal{M}$ (i.e. two compact metric spaces (X, d_X) and (Y, d_Y) are in the same class if and only if they are isometric to each other which we denote by $X \cong Y$). It is possible to compare (X, d_X) and (Y, d_Y) via the *Gromov-Hausdorff distance* [32, 41], which is a metric on $\mathcal{M}$. It is defined as

$$d_{\mathrm{GH}}(X, Y) := \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{(Z,d_Z)}(\phi(X), \psi(Y)), \tag{1}$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into a metric space (Z, d_Z) and $d_{\mathrm{H}}^{(Z,d_Z)}$ denotes the *Hausdorff distance in* Z. The Hausdorff distance is a metric on the collection of compact subsets of a metric space (Z, d_Z), which is denoted by $\mathcal{S}(Z)$, and for $A, B \in \mathcal{S}(Z)$ defined as follows

$$d_{\mathrm{H}}^{(Z,d_Z)}(A, B) := \max\left(\sup_{a\in A}\inf_{b\in B} d_Z(a, b),\ \sup_{b\in B}\inf_{a\in A} d_Z(a, b)\right). \tag{2}$$

While the Gromov-Hausdorff distance has been applied successfully for various shape and data analysis tasks (see e.g. [12–16, 19, 20, 69]), it turns out that it is generally convenient to equip the modelled objects with more structure and to model them as *metric measure spaces* [66, 67]. A metric measure space $\mathcal{X} = (X, d_X, \mu_X)$ is a triple, where (X, d_X) denotes a metric space and μ_X stands for a Borel probability measure on X with full support. This additional probability measure can be thought of as signalling the importance of different regions in the modelled object. Moreover, two metric measure spaces $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ are considered as isomorphic (denoted by $\mathcal{X} \cong_w \mathcal{Y}$) if and only if there exists an isometry $\varphi : (X, d_X) \to (Y, d_Y)$ such that $\varphi_\# \mu_X = \mu_Y$. Here, $\varphi_\#$ denotes the pushforward map induced by φ. From now on, $\mathcal{M}^w$ denotes the collection of all (isomorphism classes of) compact metric measure spaces.

The additional structure of the metric measure spaces allows to regard the modelled objects as probability measures instead of compact sets. Hence, it is possible to substitute the Hausdorff component in Equation (1) by a relaxed notion of proximity, namely the *Wasserstein distance*. This distance is fundamental to a variety of mathematical developments and is also known as Kantorovich distance [47], Kantorovich-Rubinstein distance [48], Mallows distance [63] or as the Earth Mover's distance [85]. Given a compact metric space (Z, d_Z), let $\mathcal{P}(Z)$ denote the space of probability measures on Z and let $\alpha, \beta \in \mathcal{P}(Z)$. Then, the Wasserstein distance of order p, for $1 \le p < \infty$, between α and β is defined as

$$d_{\mathrm{W},p}^{(Z,d_Z)}(\alpha, \beta) := \left(\inf_{\mu\in\mathcal{C}(\alpha,\beta)} \int_{Z\times Z} d_Z^p(x, y)\, \mu(dx \times dy)\right)^{\frac{1}{p}}, \tag{3}$$

and for $p = \infty$ as

$$d_{\mathrm{W},\infty}^{(Z,d_Z)}(\alpha,\beta) := \inf_{\mu\in\mathcal{C}(\alpha,\beta)} \sup_{(x,y)\in\mathrm{supp}(\mu)} d_Z(x,y), \tag{4}$$

where supp (μ) stands for the support of μ and $\mathcal{C}(\alpha,\beta)$ denotes the set of all couplings of α and β, i.e., the set of all probability measures μ on the product space $Z \times Z$ such that

$$\mu(A \times Z) = \alpha(A) \quad \text{and} \quad \mu(Z \times B) = \beta(B)$$

for all Borel measurable sets A and B of Z. It is worth noting that the Wasserstein distance between probability measures on the real line admits a closed form solution (see [99] and Remark 2.12).

Sturm [92] has shown that replacing the Hausdorff distance in Equation (1) with the Wasserstein distance indeed yields a meaningful metric on $\mathcal{M}^w$. Let $\mathcal{X} = (X, d_X, \mu_X)$ and $\mathcal{Y} = (Y, d_Y, \mu_Y)$ be two metric measure spaces. Then, *Sturm's Gromov-Wasserstein distance* of order p, $1 \leq p \leq \infty$, is defined as

$$d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) := \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,d_Z)}(\phi_\#\mu_X, \psi_\#\mu_Y), \tag{5}$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into the metric space (Z, d_Z).

Based on similar ideas but starting from a different representation of the Gromov-Hausdorff distance, Mémoli [66, 67] derived a computationally more tractable and topologically equivalent metric on $\mathcal{M}^w$, namely the *Gromov-Wasserstein* distance: For $1 \leq p < \infty$, the *p-distortion* of a coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ is defined as

$$\mathrm{dis}_p(\mu) := \left(\iint_{X\times Y\times X\times Y} \left| d_X(x,x') - d_Y(y,y') \right|^p \mu(dx \times dy)\, \mu(dx' \times dy') \right)^{1/p} \tag{6}$$

and for $p = \infty$ it is given as

$$\mathrm{dis}_\infty(\mu) := \sup_{\substack{x,x'\in\mathcal{X},\, y,y'\in\mathcal{Y} \\ s.t.\ (x,y),(x',y')\in\mathrm{supp}(\mu)}} \left| d_X(x,x') - d_Y(y,y') \right|.$$

The *Gromov-Wasserstein distance* of order p, $1 \leq p \leq \infty$, is defined as

$$d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) := \frac{1}{2} \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \mathrm{dis}_p(\mu). \tag{7}$$

It is known that in general $d_{\mathrm{GW},p} \leq d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and that the inequality can be strict [67]. Although both $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, are in general NP-hard to compute [67], it is possible to efficiently approximate $d_{\mathrm{GW},p}$ via conditional gradient descent [67, 79]. This has led to numerous applications and extensions of this distance [4, 18, 24, 87, 95].

In many cases, since the direct computation of either of these distances can be onerous, the determination of the degree of similarity between two datasets is performed via firstly computing *invariant features* out of each dataset (e.g. global distance distributions [75]) and secondly by suitably comparing these features. This point of view has motivated the exploration of inverse problems arising from the study of such features [11, 67, 68, 93].

Clearly, $\mathcal{M}^w$ contains various, extremely general spaces. However, in many applications it is possible to have prior knowledge about the metric measure spaces under consideration and it is often reasonable to restrict oneself to work on a specific sub-collections $\mathcal{O}^w \subseteq \mathcal{M}^w$. For instance, it could be known that the metrics of the spaces considered are induced by the shortest path metric on some underlying trees and hence it is unnecessary to consider the calculation of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, for all of $\mathcal{M}^w$. The potential advantages of focusing on a specific sub-collection $\mathcal{O}^w$ are twofold. On the one hand, it might be possible to use the features of $\mathcal{O}^w$ to gain computational benefits. On the other hand, it might be possible to refine the definition $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, to obtain more informative comparisons on $\mathcal{O}^w$. Naturally, it is of interest to identify and study these subclasses and the corresponding refinements. This approach has been pursued to study (variants of) the Gromov-Hausdorff distance on compact *ultrametric spaces* by Zarichnyi [105] and Qiu [80], and on compact *p-metric spaces* by Mémoli and Wan [70]. Here, the metric space (X, d_X) is called a p-metric space ($1 \leq p < \infty$), if for all $x, x', x'' \in X$ it holds

$$d_X(x, x'') \leq \left(d_X(x, x')^p + d_X(x', x'')^p\right)^{1/p}.$$

Further, the metric space (X, u_X) is called an ultrametric space, if u_X fulfills for all $x, x', x'' \in X$ that

$$u_X(x', x'') \leq \max(u_X(x, x'), u_X(x', x'')). \tag{8}$$

In particular, note that ultrametrics can be considered as the limiting case of p-metrics as $p \to \infty$. In particular, Mémoli and Wan [70] derived a polynomial time algorithm for the calculation of the *ultrametric Gromov-Hausdorff* distance u_{GH} between two compact ultrametric spaces (X, u_X) and (Y, u_Y) (see Section 2.2), which is defined as

$$u_{\mathrm{GH}}(X, Y) := \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{(Z,u_Z)}(\phi(X), \psi(Y)), \tag{9}$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into a common *ultrametric* space (Z, u_Z) and $d_{\mathrm{H}}^{(Z,u_Z)}$ denotes the Hausdorff distance on Z.

A further motivation to study (surrogates of) the distances $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$ restricted on a subset $\mathcal{O}^w$ comes from the idea of *slicing* which originated as a method to efficiently estimate the Wasserstein distance $d_{\mathrm{W},p}^{\mathbb{R}^d}(\alpha, \beta)$ between probability measures α and β supported in a high dimensional euclidean space $\mathbb{R}^d$ [85]. The original idea is that given any line ℓ in $\mathbb{R}^d$ one first obtains α_ℓ and β_ℓ, the respective pushforwards of α and β under the orthogonal projection map $\pi_\ell : \mathbb{R}^d \to \ell$, and then one invokes the explicit formula for the

Wasserstein distance for probability measures on $\mathbb{R}$ (see Remark 2.12) to obtain a lower bound to $d_{\mathrm{W},p}^{\mathbb{R}^d}(\alpha,\beta)$ without incurring the possibly high computational cost associated to solving an optimal transportation problem. This lower bound is improved via repeated (often random) selections of the line ℓ [9, 53, 85].

Recently, Le et al. [58] pointed out that, thanks to the fact that the 1-Wasserstein distance also admits an explicit formula when the underlying metric space is a tree [28, 34, 65], one can also devise *tree slicing* estimates of the distance between two given probability measures by suitably projecting them onto tree-like structures. Most likely, the same strategy is successful for suitable projections on random ultrametric spaces, as on these there is also an explicit formula for the Wasserstein distance [50]. The same line of of work has also recently been explored in the Gromov-Wasserstein scenario [57, 98] and could be extended based on efficiently computable restrictions (or surrogates of) $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$. Inspired by the results of Mémoli and Wan [70] on the ultrametric Gromov-Hausdorff distance and the results of Kloeckner [50], who derived an explicit representation of the Wasserstein distance on ultrametric spaces, we study the collection of compact *ultrametric measure spaces* $\mathcal{U}^w \subseteq \mathcal{M}^w$, where $\mathcal{X} = (X, u_X, \mu_X) \in \mathcal{U}^w$, whenever the underlying metric space (X, u_X) is a compact ultrametric space.

In terms of applications, ultrametric spaces (and thus also ultrametric *measure* spaces) arise naturally in statistics as metric encodings of dendrograms [19, 46] which is a graph theoretical representations of ultrametric spaces, in the context of phylogenetic trees [90], in theoretical computer science in the probabilistic approximation of finite metric spaces [5, 35], and in physics in the context of a mean-field theory of spin glasses [71, 81].

Especially for phylogenetic trees (and dendrograms), where one tries to characterize the structure of an underlying evolutionary process or the difference between two such processes, it is important to have a meaningful method of comparison, i.e., to have a meaningful metric on $\mathcal{U}^w$. However, it is evident from the definition of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and the relationship between $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$ (see [67]), that the ultrametric structure of $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ is not taken into account in the computation of either $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ or $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$, $1 \leq p \leq \infty$. Hence, we suggest, just as for the ultrametric Gromov-Hausdorff distance, to adapt the definition of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Equation (5)) as well as the one of $d_{\mathrm{GW},p}$ (see Equation (7)) and verify in the following that this makes the comparisons of ultrametric measure spaces more sensitive and leads for $p = \infty$ to a *polynomial time* algorithm for the derivation of the proposed metrics.

1.1 The proposed approach

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be ultrametric measure spaces. Reconsidering the definition of Sturm's Gromov-Wasserstein distance in Equation (5), we propose to only infimize over ultrametric spaces (Z, u_Z) in Equation (5). Thus, we define for $p \in [1, \infty]$ *Sturm's ultrametric Gromov-Wasserstein distance* of order p as

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,u_Z)}(\phi_\# \mu_X, \psi_\# \mu_Y), \tag{10}$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into an ultrametric space (Z, u_Z).

In the subsequent sections of this paper, we will establish many theoretically appealing properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$. Unfortunately, we will verify that, although an explicit formula for the Wasserstein distance of order p on ultrametric spaces exists [50], for $p \in [1, \infty)$ the calculation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ yields a highly non-trivial combinatorial optimization problem (see Section 3.1.1). Therefore, we demonstrate that an adaption of the Gromov-Wasserstein distance defined in Equation (7) yields a topologically equivalent and easily approximable distance on $\mathcal{U}^w$. In order to define this adaption, we need to introduce some notation. For $a, b \geq 0$ and $1 \leq q < \infty$ let

$$\Lambda_q(a, b) := |a^q - b^q|^{1/q}.$$

Further define $\Lambda_\infty(a, b) := \max(a, b)$ whenever $a \neq b$ and $\Lambda_\infty(a, b) = 0$ if $a = b$.

Now, we can rewrite the p-distortion for $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ in the definition of $d_{\mathrm{GW},p}$, $1 \leq p < \infty$, (see (6) and (7)) as follows

$$\mathrm{dis}_p(\mu) = \left(\iint_{X \times Y \times X \times Y} \left(\Lambda_1(d_X(x, x'), d_Y(y, y'))\right)^p \mu(dx \times dy)\, \mu(dx' \times dy') \right)^{1/p}. \tag{11}$$

Considering the derivation of $d_{\mathrm{GW},p}$ in [67] and the results on the closely related ultrametric Gromov-Hausdorff distance studied in [70], this suggests to replace Λ_1 in Equation (11) with Λ_∞ in order to incorporate the ultrametric structures of (X, u_X, μ_X) and (Y, u_Y, μ_Y) into the comparison. Hence, we define the *p-ultra-distortion* of a coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ for $1 \leq p < \infty$ as

$$\mathrm{dis}_p^{\mathrm{ult}}(\mu) := \left(\iint_{X \times Y \times X \times Y} \left(\Lambda_\infty(u_X(x, x'), u_Y(y, y'))\right)^p \mu(dx \times dy)\, \mu(dx' \times dy') \right)^{1/p}. \tag{12}$$

and for $p = \infty$ as

$$\mathrm{dis}_\infty^{\mathrm{ult}}(\mu) := \sup_{\substack{x,x' \in \mathcal{X},\, y,y' \in \mathcal{Y} \\ s.t.\ (x,y),(x',y') \in \mathrm{supp}(\mu)}} \Lambda_\infty(u_X(x, x'), u_Y(y, y')).$$

The *ultrametric Gromov-Wasserstein distance* of order $p \in [1, \infty]$, is given as

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) := \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \mathrm{dis}_p^{\mathrm{ult}}(\mu). \tag{13}$$

Due to the structural similarity between $d_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}$, we can expect (and later verify) that many properties of $d_{\mathrm{GW},p}$ extend to $u_{\mathrm{GW},p}$. In particular, we will establish that also $u_{\mathrm{GW},p}$ can be approximated[1] via conditional gradient descent and admits several polynomial time computable lower bounds which are useful in applications.

It is worth mentioning that Sturm [93] studied the family of so-called $L^{p,q}$-distortion distances similar to our construction of $u_{\mathrm{GW},p}$. In our language, for any $p, q \in [1, \infty)$, the $L^{p,q}$-distortion distance is constructed by infimizing over the (p, q)-distortion defined by replacing Λ_∞ with $(\Lambda_q)^q$ in Equation (12). This distance shares many properties with $d_{\mathrm{GW},p}$.

1.2 Overview of our results

We give a brief overview of our results.

Section 2. We generalize the results of Carlsson and Mémoli [19] on the relation between ultrametric spaces and dendrograms and establish a bijection between compact ultrametric spaces and *proper dendrograms* (see Definition 2.1). After recalling some results on the ultrametric Gromov-Hausdorff distance (see Equation (9)), we use the connection between compact ultrametric spaces and dendrograms to reformulate the explicit formula for the p-Wasserstein distance ($1 \leq p < \infty$) on ultrametric spaces derived by Kloeckner [50] in terms of proper dendrograms. This allows us to derive a formulation of the ∞-Wasserstein distance on ultrametric spaces and to study the Wasserstein distance on compact subspaces of the ultrametric space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$, which will be relevant when studying lower bounds of $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$.

Section 3. We demonstrate that $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p \leq \infty$, are p-metrics on the collection of ultrametric measure spaces $\mathcal{U}^w$. We derive several alternative representations for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and study the relation between the metrics $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$. In particular, we show that, while for $1 \leq p < \infty$ it holds in general that $u_{\mathrm{GW},p} \leq 2^{\frac{1}{p}}\, u_{\mathrm{GW},p}^{\mathrm{sturm}}$, both metrics coincide for $p = \infty$, i.e., $u_{\mathrm{GW},\infty} = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$. Furthermore, we show how this equality in combination with an alternative representation of $u_{\mathrm{GW},\infty}$ leads to a *polynomial time algorithm* for the calculation of $u_{\mathrm{GW},\infty}^{\mathrm{sturm}} = u_{\mathrm{GW},\infty}$. Moreover, we study the topological properties of $(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ and $(\mathcal{U}^w, u_{\mathrm{GW},p})$, $1 \leq p \leq \infty$. Most importantly, we show that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ induce the same topology on $\mathcal{U}^w$ which is also different from the one induced by $d_{\mathrm{GW},p}^{\mathrm{sturm}}/d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. While we further prove that the metric spaces $(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ and $(\mathcal{U}^w, u_{\mathrm{GW},p})$, $1 \leq p < \infty$, are neither complete nor separable metric space, we demonstrate

[1] Here "approximation" is meant in the sense that one can write code which will locally minimize the functional. There are in general no theoretical guarantees that these algorithms will converge to a global minimum.

that the ultrametric space $(\mathcal{U}^w, u_{\mathrm{GW},\infty}^{\mathrm{sturm}})$, which coincides with $(\mathcal{U}^w, u_{\mathrm{GW},\infty})$, is complete. Finally, we establish that $(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ is a geodesic space.

Section 4. Unfortunately, it does not seem to be possible to derive a polynomial time algorithm for the calculation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$, $1 \leq p < \infty$. Consequently, based on easily computable invariant features, in Section 4 we derive several polynomial time computable lower bounds for $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$. Due to the structural similarity between $d_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}$, these are in a certain sense analogue to those derived in [66, 67] for $d_{\mathrm{GW},p}$. Among other things, we show that

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) := \inf_{\gamma \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \|\Lambda_\infty(u_X, u_Y)\|_{L^p(\gamma)}. \tag{14}$$

We verify that the lower bound $\mathbf{SLB}_p^{\mathrm{ult}}$ can be reformulated in terms of the Wasserstein distance on the ultrametric space $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ (we derive an explicit formula for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}$ in Section 2.3). This allows us to efficiently calculate $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y})$ in $O((m \vee n)^2)$, where m stands for the cardinality of X and n for the one of Y.

Section 5. As the ultrametric space assumption is somewhat restrictive (especially in the context of phylogenetic trees, see [90]), we prove in Section 5 that the results on $u_{\mathrm{GW},p}$ can be extended to the more general *ultra-dissimilarity spaces* (see Definition 5.1). In particular, we prove that $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, is a metric on the *isomorphism classes* of ultra-dissimilarity spaces (see Definition 5.5).

Section 6. We illustrate the behaviour and relation between $u_{\mathrm{GW},1}$ (which can be approximated via conditional gradient descent) and $\mathbf{SLB}_1^{\mathrm{ult}}$ in a set of illustrative examples. Additionally, we carefully illustrate the differences between $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$, and $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ (see Section 4 for a definition), respectively.

Section 7. Finally, we apply our ideas to *phylogenetic tree shape comparison*. To this end, we compare two sets of phylogenetic tree shapes based on the HA protein sequences from human influenza collected in different regions with the lower bound $\mathbf{SLB}_1^{\mathrm{ult}}$. In particular, we contrast our results in both settings to the ones obtained with the tree shape metric introduced in Equation (4) of Colijn and Plazzotta [25].

1.3 Related work

In order to better contextualize our contribution, we now describe related work, both in applied and computational geometry, and in phylogenetics (where notions of distance between trees have arisen naturally).

Metrics between trees: the phylogenetics perspective

In phylogenetics, where one chief objective is to infer the evolutionary relationship between species via methods that evaluate observable traits, such as DNA sequences, the need to be able to measure dissimilarity between different trees arises from the fact that the process of reconstruction of a phylogenetic tree may depend on the set of genes being considered. At the same time, even for the same set of genes, different reconstruction methods could be applied which would result in different trees. As such, this has led to the development of many different metrics for measuring distance between phylogenetic trees. Examples include the Robinson-Foulds metric [84], the subtree-prune and regraft distance [42], and the nearest-neighbor interchange distance [83].

As pointed out in [76], many of these distances tend to quantify differences between tree topologies and often do not take into account edge lengths. A certain phylogenetic tree metric space which encodes for edge lengths was proposed in [6] and studied algorithmically in [76]. This tree space assumes that the all trees have the same set of taxa. An extension to the case of trees over different underlying sets is given in [40]. Lafond et al. [56] considered one type of metrics on possibly *muiltilabeled* phylogenetic trees with a fixed number of leafs. As the authors pointed out, a multilabeled phylogenetic tree in which no leafs are repeated is just a standard phylogenetic tree, whereas a multilabeled phylogenetic tree in which all labels are equal defines a *tree shape*. The authors then proceeded to study the computational complexity associated to generalizations of some of the usual metrics for phylogenetic trees (such as the Robinson-Foulds distance) to the multilabeled case. Colijn and Plazzotta [25] studied a metric between (binary) phylogenetic tree shapes based on a bottom to top enumeration of specific connectivity structures. The authors applied their metric to compare evolutionary trees based on the HA protein sequences from human influenza collected in different regions.

Metrics between trees: the applied geometry perspective

From a different perspective, ideas from applied geometry and applied and computational topology have been applied to the comparison of tree shapes in applications in probability, clustering and applied and computational topology.

Metric trees are also considered in probability theory in the study of models for random trees together with the need to quantify their distance; Evans [33] described some variants of the Gromov-Hausdorff distance between metric trees. See also [39] for the case of metric measure space representations of trees and a certain Gromov-Prokhorov type of metric on the collection thereof.

Trees, in the form of dendrograms, are abundant in the realm of hierarhical clustering methods. In their study of the *stability* of hierarchical clustering methods, Carlsson and

Mémoli [19] utilized the Gromov-Hausdorff distance between the ultrametric representation of dendrograms. Schmiedl [88] proved that computing the Gromov-Hausdorff distance between tree metric spaces is NP-hard. Liebscher [59] suggested some variants of the Gromov-Hausdorff distance which are applicable in the context of phylogenetic trees. As mentioned before, Zarichnyi [105] introduced the ultrametric Gromov-Hausdorff distance u_{GH} between compact ultrametric spaces (a special type of tree metric spaces). Certain theoretical properties such as precompactness of u_{GH} has been studied in [80]. In contrast with the NP-hardness of computing d_{GH}, Mémoli and Wan [70] devised an polynomial time algorithm for computing u_{GH}.

In computational topology *merge trees* arise through the study of the sublevel sets of a given function [1, 82] with the goal of shape simplification. Morozov et al. [74] developed the notion of *interleaving distance* between merge trees which is related to the Gromov-Hausdorff distance between trees through bi-Lipschitz bounds. In [2], exploiting the connection between the interleaving distance and the Gromov-Hausdorff between metric trees, the authors approached the computation of the Gromov-Hausdorff distance between metric trees in general and provide certain approximation algorithms. Touli and Wang [96] devised fixed-parameter tractable (FPT) algorithms for computing the interleaving distance between metric trees. One can imply from their methods an FPT algorithm to compute a 2-approximation of the Gromov-Hausdorff distance between ultrametric spaces. Mémoli and Wan [70] devised an FPT algorithm for computing the exact value of the Gromov-Hausdorff distances between ultrametric spaces.

2 Preliminaries

In this section we briefly summarize the basic notions and concepts required throughout the paper.

2.1 Ultrametric spaces and dendrograms

We begin by describing compact ultrametric spaces in terms of *proper dendrograms*. To this end, we introduce some definitions and some notation. Given a set X, a *partition* of X is a set $P_X = \{X_i\}_{i\in I}$ where I is any index set, $\emptyset \neq X_i \subseteq X$, $X_i \cap X_j = \emptyset$ for all $i \neq j \in I$ and $\bigcup_{i\in I} X_i = X$. We call each element X_i a *block* of the given partition P_X and denote by $\mathbf{Part}(X)$ the collection of all partitions of X. For two partitions P_X and P'_X we say that P_X is *finer* than P'_X, if for every block $X_i \in P_X$ there exists a block $X'_j \in P'_X$ such that $X_i \subseteq X'_j$.

Definition 2.1 (Proper dendrogram)**.** Given a set X (not necessarily finite), a *proper dendrogram* $\theta_X : [0,\infty) \to \mathbf{Part}(X)$ is a map satisfying the following conditions:

1. $\theta_X(s)$ is finer than $\theta_X(t)$ for any $0 \leq s < t < \infty$;
2. $\theta_X(0)$ is the finest partition consisting only singleton sets;
3. There exists $T > 0$ such that for any $t \geq T$, $\theta_X(t) = \{X\}$ is the trivial partition;
4. For each $t > 0$, there exists $\varepsilon > 0$ such that $\theta_X(t) = \theta_X(t')$ for all $t' \in [t, t+\varepsilon]$.
5. For any distinct points $x, x' \in X$, there exists $T_{xx'} > 0$ such that x and x' belong to different blocks in $\theta_X(T_{xx'})$.
6. For each $t > 0$, $\theta_X(t)$ consists of only finitely many blocks.
7. Let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence such that $\lim_{n\to\infty} t_n = 0$ and let $X_n \in \theta_X(t_n)$. If for any $1 \leq n < m$, $X_m \subseteq X_n$, then $\bigcap_{n\in\mathbb{N}} X_n \neq \emptyset$.

When X is finite, a function $\theta_X : [0,\infty) \to \mathbf{Part}(X)$ satifying conditions (1) to (4) will satisfy conditions (5), (6) and (7) automatically, and thus a proper dendrogram reduces to the usual dendrogram (see [19, Sec. 3.1] for a formal definition). Let θ_X be a proper dendrogram over a set X. For any $x \in X$ and $t \geq 0$, we denote by $[x]_t^X$ the block in $\theta(t)$ that contains $x \in X$ and abbreviate $[x]_t^X$ to $[x]_t$ when the underlying set X is clear from the context. Similar to [19], who considered the relation between finite ultrametric spaces and dendrograms, we will prove that there is a bijection between compact ultrametric spaces and proper dendrograms. In particular, one can show that the subsequent theorem generalizes [19, Theorem 9]. Since its proof depends on several concepts not yet introduced, we postpone it to Appendix A.1.1.

Theorem 2.2. *Given a set X, denote by $\mathcal{U}(X)$ the collection of all compact ultrametrics on X and $\mathcal{D}(X)$ the collection of all proper dendrograms over X. For any $\theta \in \mathcal{D}(X)$, consider u_θ defined as follows:*

$$\forall x, x' \in X, \;\; u_\theta(x, x') := \inf\{t \geq 0 \,|\, x, x' \text{ belong to the same block of } \theta(t)\}.$$

Then, $u_\theta \in \mathcal{U}(X)$ and the map $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ sending θ to u_θ is a bijection.

Remark 2.3. From now on, we denote by θ_X the proper dendrogram corresponding to a given compact ultrametric u_X on X under the bijection given above. Note that a block $[x]_t$ in $\theta_X(t)$ is actually the closed ball $B_t(x)$ in X centered at x with radius t. So for each $t \geq 0$, $\theta_X(t)$ partitions X into a union of several closed balls in X with respect to u_X.

2.2 The ultrametric Gromov-Hausdorff distance

Both $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, are by construction closely related to the Gromov-Hausdorff distance. In a recent paper, Mémoli and Wan [70] studied an ultrametric version of this distance, namely the *ultrametric Gromov-Hausdorff distance* (denoted as u_{GH}).

Since we will demonstrate several connections between $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, and this distance, we briefly summarize some of the results in [70]. We start by recalling the formal definition of u_{GH}.

Definition 2.4. Let (X, u_X) and (Y, u_Y) be two compact ultrametric spaces. Then, the *ultrametric Gromov-Hausdorff* between X and Y is defined as

$$u_{\mathrm{GH}}(X, Y) = \inf_{Z,\phi,\psi} d_{\mathrm{H}}^{Z}\left(\phi(X), \psi(Y)\right),$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings (distance preserving transformations) into the ultrametric space (Z, u_Z).

Zarichnyi [105] has shown that u_{GH} is an ultrametric on the isometry classes of compact ultrametric spaces, which are denoted by $\mathcal{U}$, and Mémoli and Wan [70] identified a structural theorem (cf. Theorem 2.5) that gives rise to a polynomial time algorithm for the calculation of u_{GH}. More precisely, it was proven in [70] that u_{GH} can be calculated via so-called *quotient ultrametric spaces*, which we define next. Let (X, u_X) be an ultrametric space and let $t \geq 0$. We define an equivalence relation $\sim_t$ on X as follows: $x \sim_t x'$ if and only if $u_X(x, x') \leq t$. We denote by $[x]_t^X$ (resp. $[x]_t$) the equivalence class of x under $\sim_t$ and by X_t the set of all such equivalence classes. In fact, $[x]_t^X = \{x' \in X \,|\, u(x, x') \leq t\}$ is exactly the closed ball centered at x with radius t and corresponds to a block in the corresponding proper dendrogram $\theta_X(t)$ (see Remark 2.3). Thus, one can think of X_t as a "set representation" of $\theta_X(t)$. We define an ultrametric u_{X_t} on X_t as follows:

$$u_{X_t}([x]_t, [x']_t) := \begin{cases} u_X(x, x'), & [x]_t \neq [x']_t \\ 0, & [x]_t = [x']_t. \end{cases}$$

Then, (X_t, u_{X_t}) is an ultrametric space and we call (X_t, u_{X_t}) the *quotient* of (X, u_X) at level t (see Figure 1 for an illustration). It is straightforward to prove that the quotient of a compact ultrametric space at level $t > 0$ is a finite ultrametric space (cf. [102, Lemma 2.3]). Furthermore, the quotient spaces characterize u_{GH} as follows.

Theorem 2.5 (Structural theorem for u_{GH}, [70, Theorem 5.7])**.** *Let (X, u_X) and (Y, u_Y) be two compact ultrametric spaces. Then,*

$$u_{\mathrm{GH}}(X, Y) = \inf\left\{t \geq 0 \,|\, X_t \cong Y_t\right\}.$$

Remark 2.6. Let (X, u_X) and (Y, u_Y) denote two finite ultrametric spaces and let $t \geq 0$. The quotient spaces X_t and Y_t can be considered as vertex weighted, rooted trees [70]. Hence, it is possible to check whether $X_t \cong Y_t$ in polynomial time [3]. Consequently, Theorem 2.5 induces a simple, polynomial time algorithm to calculate u_{GH} between two finite ultrametric spaces.

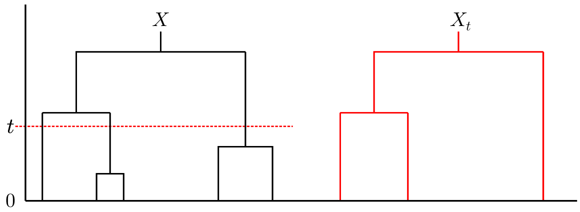

Fig. 1: Metric quotient: An ultrametric space (black) and its quotient at level t (red).

2.3 Wasserstein distance on ultrametric spaces

Kloeckner [50] uses the representation of ultrametric spaces as so called *synchronized rooted trees* to derive an explicit formula for the Wasserstein distance on ultrametric spaces. By the constructions of the dendrograms and of the synchronized rooted trees (see Appendix A.2.1), it is immediately clear how to reformulate the results of Kloeckner [50] on compact ultrametric spaces in terms of proper dendrograms. To this end, we need to introduce some notation. For a compact ultrametric space X, let θ_X be the associated proper dendrogram and let $V(X) := \bigcup_{t>0} \theta_X(t) = \{[x]_t | \, x \in X, t > 0\}$. It can be shown that $V(X)$ is the collection of all closed balls in X except for singletons $\{x\}$ such that x is a cluster point[2] (see Lemma A.8). For $B \in V(X)$, we denote by B^* the smallest (under inclusion) element in $V(X)$ such that $B \subsetneqq B^*$ (for the existence and uniqueness of B^* see Lemma A.1).

Theorem 2.7 (The Wasserstein distance on ultrametric spaces, [50, Theorem 3.1])**.** *Let (X, u_X) be a compact ultrametric space. For all $\alpha, \beta \in \mathcal{P}(X)$ and $1 \leq p < \infty$, we have*

$$\left(d_{\mathrm{W},p}^X\right)^p(\alpha,\beta) = 2^{-1} \sum_{B \in V(X)\setminus\{X\}} \left(\operatorname{diam}(B^*)^p - \operatorname{diam}(B)^p\right) |\alpha(B) - \beta(B)| \,. \tag{15}$$

While Theorem 2.7 is only valid for $p < \infty$, it can be extended to the case $p = \infty$.

Lemma 2.8. *Let X be a compact ultrametric space. Then, for any $\alpha, \beta \in P(X)$, we have*

$$d_{\mathrm{W},\infty}^X(\alpha,\beta) = \max_{B \in V(X)\setminus\{X\} \text{ and } \alpha(B)\neq\beta(B)} \operatorname{diam}(B^*) \,. \tag{16}$$

The proof of Lemma 2.8 is technical and we postpone it to Appendix A.1.2.

[2] A cluster point x in a topological space X is such that any neighborhood of x contains countably many points in X.

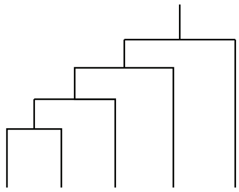

Fig. 2: Illustration of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$**:** This is the dendrogram for a subspace of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ consisting of 5 arbitrary distinct points of $\mathbb{R}_+$.

2.3.1 Wasserstein distance on $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$

The non-negative half real line $\mathbb{R}_{\geq 0}$ endowed with Λ_∞ turns out to be an ultrametric space (cf. [70, Remark 1.14]). Finite subspaces of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ are of particular interest in this paper. These spaces possess a particular structure (see Figure 2) and the computation of the Wasserstein distance on them can be further simplified.

Theorem 2.9 ($d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ between finitely supported measures)**.** *Suppose α, β are two probability measures supported on a finite subset $\{x_0, \ldots, x_n\}$ of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ such that $0 \leq x_0 < x_1 < \cdots < x_n$. Denote $\alpha_i \coloneqq \alpha(\{x_i\})$ and $\beta_i \coloneqq \beta(\{x_i\})$. Then, we have for $p \in [1, \infty)$ that*

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha, \beta) = 2^{-\frac{1}{p}} \left(\sum_{i=0}^{n-1} \left| \sum_{j=0}^{i} (\alpha_j - \beta_j) \right| \cdot |x_{i+1}^p - x_i^p| + \sum_{i=0}^{n} |\alpha_i - \beta_i| \cdot x_i^p \right)^{\frac{1}{p}}. \quad (17)$$

Let F_α and F_β denote the cumulative distribution functions of α and β, respectively. Then, for the case $p = \infty$ we obtain

$$d_{\mathrm{W},\infty}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha, \beta) = \max \left(\max_{0 \leq i \leq n-1, F_\alpha(x_i) \neq F_\beta(x_i)} x_{i+1}, \max_{0 \leq i \leq n, \alpha_i \neq \beta_i} x_i \right).$$

Proof. Clearly, $V(X) = \{\{x_0, x_1, \ldots, x_i\} | \, i = 1, \ldots, n\} \cup \{\{x_i\} | \, i = 1, \ldots, n\}$ (recall that each set corresponds to a closed ball). Thus, we conclude the proof by applying Theorem 2.7 and Lemma 2.8. □

Remark 2.10 (The case $p = 1$)**.** Note that when $p = 1$, for any finitely supported probability measures $\alpha, \beta \in \mathcal{P}(\mathbb{R}_{\geq 0})$,

$$d_{\mathrm{W},1}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha, \beta) = \frac{1}{2} \left(d_{\mathrm{W},1}^{(\mathbb{R},\Lambda_1)}(\alpha, \beta) + \int_{\mathbb{R}} x \, |\alpha - \beta|(dx) \right).$$

The formula indicates that the 1-Wasserstein distance on $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ is the average of the usual 1-Wasserstein distance on $(\mathbb{R}_{\geq 0}, \Lambda_1)$ and a "weighted total variation distance". The weighted total variation like distance term is sensitive to difference of supports. For example, let $\alpha = \delta_{x_1}$ and $\beta = \delta_{x_2}$, then $\int_{\mathbb{R}} x \, |\alpha - \beta|(dx) = x_1 + x_2$ if $x_1 \neq x_2$.

Remark 2.11 (Extension to compactly supported measures)**.** In fact, $X \subseteq (\mathbb{R}_{\geq 0}, \Lambda_\infty)$ is compact if and only if it is either a finite set or countable with 0 being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1) (see Lemma A.2). Hence, it is straightforward to extend Theorem 2.9 to compactly supported measures and we refer to Appendix A.3 for the missing details.

Remark 2.12 (Closed-form solution for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$)**.** We know that there is a closed-form solution for Wasserstein distance on $\mathbb{R}$ with the usual Euclidean distance Λ_1:

$$d_{\mathrm{W},p}^{(\mathbb{R},\Lambda_1)}(\alpha,\beta) = \left(\int_0^1 |F_\alpha^{-1}(t) - F_\beta^{-1}(t)|^p dt\right)^{\frac{1}{p}},$$

where F_α and F_β are cumulative distribution functions of α and β, respectively. We have also obtained a closed-form solution for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ in Theorem 2.9. We generalize these formulas to the case $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$ when $q \in (1,\infty)$ and $q \leq p$ in Appendix A.3.1.

3 Ultrametric Gromov-Wasserstein distances

In this section we investigate the properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ as well as $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, and study the relation between them.

3.1 Sturm's ultrametric Gromov-Wasserstein distance

We begin by establishing several basic properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p \leq \infty$, including a proof that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is indeed a metric (or more precisely a p-metric) on the collection of compact ultrametric measure spaces $\mathcal{U}^w$.

The definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ given in Equation (10) is clunky, technical and in general not easy to work with. Hence, the first observation to make is the fact that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p \leq \infty$, shares a further property with $d_{\mathrm{GW},p}^{\mathrm{sturm}}$: $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ can be calculated by minimizing over pseudo-ultrametrics instead of isometric embeddings.

Lemma 3.1. *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two ultrametric measure spaces. Let $\mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ denote the collection of all pseudo-ultrametrics u on the disjoint union $X \sqcup Y$ such that $u|_{X\times X} = u_X$ and $u|_{Y\times Y} = u_Y$. Let $p \in [1,\infty]$. Then, it holds that*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = \inf_{u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)} d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X,\mu_Y), \tag{18}$$

where $d_{\mathrm{W},p}^{(X\sqcup Y,u)}$ denotes the Wasserstein pseudometric *of order p defined in Equation* (34) *(resp. in Equation* (35) *for $p=\infty$) in Appendix B.5.1 of the supplement.*

Proof. The above lemma follows by the same arguments as Lemma 3.3 (*iii*) in [92]. □

Remark 3.2 (Wasserstein pseudometric)**.** The *Wasserstein pseudometric* is a natural extension of the Wasserstein distance to pseudometric spaces and has for example been studied in Thorsley and Klavins [94]. In Appendix B.5.1 we carefully show that it is closely related to the Wasserstein distance on a canonically induced metric space. We further establish that the Wasserstein distance and the Wasserstein pseudometric share many relevant properties. Hence, we do not notationally distinguish between these two concepts.

The representation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \le p \le \infty$, given by the above lemma is much more accessible and we first use it to establish the subsequent basic properties of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Appendix B.1.1 for a full proof).

Proposition 3.3. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, the following holds:*

1. *For any $p \in [1,\infty]$, we always have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \ge d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.*

2. *For any $1 \le p \le q \le \infty$, we have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \le u_{\mathrm{GW},q}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.*

3. *It holds that $\lim_{p\to\infty} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.*

Moreover, we use Lemma 3.1 to prove that $(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}})$ is indeed a metric space.

Theorem 3.4. *$u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is a p-metric on the collection $\mathcal{U}^w$ of compact ultrametric measure spaces. In particular, when $p = \infty$, $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$ is an ultrametric.*

In order to increase the readability of this section we postpone the proof of Theorem 3.4 to Appendix B.1.2. In the course of the proof, we will, among other things, verify the existence of optimal metrics and optimal couplings in Equation (18) (see Proposition B.1). Furthermore, it is important to note that the topology induced on $\mathcal{U}^w$ by $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \le p \le \infty$, is different from the one induced by $d_{\mathrm{GW},p}^{\mathrm{sturm}}$. This is well illustrated in the following example.

Example 3.5 ($u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ induce different topologies)**.** This example is an adaptation from Mémoli and Wan [70, Example 3.14]. For each $a > 0$, denote by $\Delta_2(a)$ the two-point metric space with interpoint distance a. Endow with $\Delta_2(a)$ the uniform probability measure μ_a and denote the corresponding ultrametric measure space $\hat{\Delta}_2(a)$. Now, let $\mathcal{X} := \hat{\Delta}_2(1)$ and let $\mathcal{X}_n := \hat{\Delta}_2\left(1 + \frac{1}{n}\right)$ for $n \in \mathbb{N}$. It is easy to check that for any $1 \le p \le \infty$, $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{X}_n) = \frac{1}{2n}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{X}_n) = 2^{-\frac{1}{p}}(1+\frac{1}{n})$ where we adopt the convention that $1/\infty = 0$. Hence, as n goes to infinity $\mathcal{X}_n$ will converge to $\mathcal{X}$ in the sense of $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, but not in the sense of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, for any $1 \le p \le \infty$.

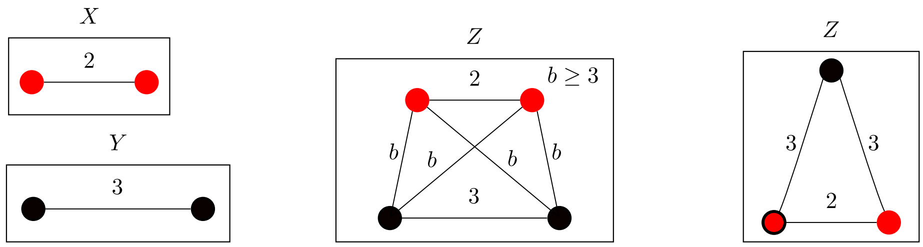

Fig. 3: Common ultrametric spaces: Representation of the two kinds of ultrametric spaces Z (middle and right) into which we can isometrically embed the spaces X and Y (left).

3.1.1 Alternative representations of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In this subsection, we derive an alternative representation for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ defined in Equation (10). We mainly focus on the case $p < \infty$, however it turns out that the results also hold for $p = \infty$ (see Section 3.3).

Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ and recall the original definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $p \in [1, \infty]$, given in Equation (10), i.e.,

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \inf_{Z,\phi,\psi} d_{\mathrm{W},p}^{(Z,u_Z)}(\varphi_{\#}\mu_Y, \psi_{\#}\mu_Y),$$

where $\phi : X \to Z$ and $\psi : Y \to Z$ are isometric embeddings into an ultrametric space (Z, u_Z). It turns out that we only need to consider relatively few possibilities of mapping two ultrametric spaces into a common ultrametric space. Exemplarily, this is shown in Figure 3, where we see two finite ultrametric spaces and two possibilities for a common ultrametric space Z. Indeed, it is straightforward to write down all reasonable embeddings and target spaces. We define the set

$$\mathcal{A} := \{(A, \varphi) \,|\, \emptyset \neq A \subseteq X \text{ is closed and } \varphi : A \hookrightarrow Y \text{ is an isometric embedding }\}. \tag{19}$$

Clearly, $\mathcal{A} \neq \emptyset$, as it holds for each $x \in X$ that $\{(\{x\}, \varphi_y)\}_{y \in Y} \subseteq \mathcal{A}$, where φ_y is the map sending x to $y \in Y$. Another possibility to construct elements in $\mathcal{A}$ is illustrated in the subsequent example.

Example 3.6. Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ be finite spaces and let $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. If $u^{-1}(0) \neq \emptyset$, we define $A := \pi_X(u^{-1}(0)) \subseteq X$, where $\pi_X : X \times Y \to X$ is the canonical projection. Then, the map $\varphi : A \to Y$ defined by sending $x \in A$ to $y \in Y$ such that $u(x, y) = 0$ is an isometric embedding and in particular, $(A, \varphi) \in \mathcal{A}$.

Now, fix two compact spaces $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Let $(A, \varphi) \in \mathcal{A}$ and let $Z_A = X \sqcup (Y \setminus \varphi(A)) \subseteq X \sqcup Y$. Furthermore, define $u_{Z_A} : Z_A \times Z_A \to \mathbb{R}_{\geq 0}$ as follows:

1. $u_{Z_A}|_{X \times X} := u_X$ and $u_{Z_A}|_{Y \setminus \varphi(A) \times Y \setminus \varphi(A)} := u_Y|_{Y \setminus \varphi(A) \times Y \setminus \varphi(A)}$;
2. For any $x \in A$ and $y \in Y \setminus \varphi(A)$ define $u_{Z_A}(x, y) := u_Y(y, \varphi(x))$;

3. For $x \in X \setminus A$ and $y \in Y \setminus \varphi(A)$ let $u_{Z_A}(x,y) \coloneqq \inf\{\max(u_X(x,a), u_Y(\varphi(a),y)) \,|\, a \in A\}$;
4. For any $x \in X$ and $y \in Y \setminus \varphi(A)$, $u_{Z_A}(y,x) \coloneqq u_{Z_A}(x,y)$.

Then, (Z_A, u_{Z_A}) is an ultrametric space such that X and Y can be mapped isometrically into Z_A (see [105, Lemma 1.1]). Let $\phi^X_{(A,\varphi)}$ and $\psi^Y_{(A,\varphi)}$ denote the corresponding isometric embeddings of X and Y, respectively. This allows us to derive the following statement, whose proof is postponed to Appendix B.1.3.

Theorem 3.7. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, we have for each $p \in [1,\infty)$ that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}} d^{Z_A}_{\mathrm{W},p}\left(\left(\phi^X_{(A,\varphi)}\right)_{\#}\mu_X, \left(\psi^Y_{(A,\varphi)}\right)_{\#}\mu_Y\right). \tag{20}$$

Remark 3.8. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite ultrametric measure spaces. The representation of $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$, $1 \le p \le \infty$ given by Theorem 3.7 is very explicit and recasts the computation of $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$, $1 \le p \le \infty$, as a combinatorial problem. In fact, as $\mathcal{X}$ and $\mathcal{Y}$ are finite, the set $\mathcal{A}$ in Equation (20) can be further reduced. More precisely, we demonstrate in Appendix B.1.3 (see Corollary B.7) that it is sufficient to infimize over the set of all *maximal pairs*, denoted by $\mathcal{A}^*$. Here, a pair $(A,\varphi_1) \in \mathcal{A}$ is denoted as *maximal*, if for all pairs $(B,\varphi_2) \in \mathcal{A}$ with $A \subseteq B$ and $\varphi_2|_A = \varphi_1$ it holds $A = B$. Using the ultrametric Gromov-Hausdorff distance (see Equation (9)) it is possible to determine if two ultrametric spaces are isometric in polynomial time [70, Theorem 5.7]. However, this is clearly not sufficient to identify all $(A,\varphi) \in \mathcal{A}^*$ in polynomial time. Especially, for a given, viable $A \subseteq X$, there are usually multiple ways to define the corresponding map φ. Furthermore, we have for $1 \le p < \infty$ neither been able to further restrict the set $\mathcal{A}^*$ nor to identify the optimal (A^*, φ^*). This just leaves a brute force approach which is computationally not feasible. On the other hand, for $p = \infty$ we are able to explicitly construct the optimal pair (A^*, φ^*) (see Theorem 3.22).

3.2 The ultrametric Gromov-Wasserstein distance

In the following, we consider basic properties of $u_{\mathrm{GW},p}$ and prove the analogue of Theorem 3.4, i.e., we verify that also $u_{\mathrm{GW},p}$ is a p-metric, $1 \le p \le \infty$, on the collection of ultrametric measure spaces.

The subsequent proposition collects three basic properties of $u_{\mathrm{GW},p}$ which are also shared by $u^{\mathrm{sturm}}_{\mathrm{GW},p}$ (cf. Proposition 3.3). We refer to Appendix B.2.1 for its proof.

Proposition 3.9. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, the following holds:*

1. *For any $p \in [1,\infty]$, we always have that $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \ge d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$.*

2. *For any* $1 \leq p \leq q \leq \infty$*, it holds* $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \leq u_{\mathrm{GW},q}(\mathcal{X},\mathcal{Y})$*;*
3. *We have that* $\lim_{p\to\infty} u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

Next, we verify that $u_{\mathrm{GW},p}$ is indeed a metric on the collection of ultrametric measure spaces.

Theorem 3.10. *The ultrametric Gromov-Wasserstein distance* $u_{\mathrm{GW},p}$ *is a* p*-metric on the collection* $\mathcal{U}^w$ *of compact ultrametric measure spaces. In particular, when* $p = \infty$*,* $u_{\mathrm{GW},\infty}$ *is an ultrametric.*

The full proof of Theorem 3.10, which is based on the existence of optimal couplings in Equation (13) (see Proposition B.10), is postponed to Appendix B.2.2.

Remark 3.11 ($u_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}$ induce different topologies)**.** Reconsidering Example 3.5, it is easy to verify that in this setting $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{X}_n) = 2^{-\frac{1}{p}}\left(1+\frac{1}{n}\right)$ while $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{X}_n) = \frac{1}{2^{1/p}n}$, $1 \leq p \leq \infty$. Hence, just like $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$, $u_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}$ do not induce the same topology on $\mathcal{U}^w$. This result can also be obtained from Section 3.4 where we derive that $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ give rise to the same topology.

Remark 3.12. By the same arguments as for $d_{\mathrm{GW},p}$, $1 \leq p < \infty$, [67, Sec. 7], it follows that for two finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ the computation of $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$, $1 \leq p < \infty$, boils down to solving a (non-convex) quadratic program. This is in general NP-hard [78]. On the other hand, for $p = \infty$, we will derive a polynomial time algorithm to determine $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$ (cf. Section 3.2.1).

3.2.1 Alternative representations of $u_{\mathrm{GW},\infty}$

In the following, we will derive an alternative representation of $u_{\mathrm{GW},\infty}$ that resembles the one of u_{GH} derived in [70, Theorem 5.7]. It also leads to a polynomial time algorithm for the computation of $u_{\mathrm{GW},\infty}$. For this purpose, we define the *weighted quotient* of an ultrametric measure space. Let $\mathcal{X} = (X, u_X, \mu_X) \in \mathcal{U}^w$ and let $t \geq 0$. Then, the *weighted quotient* of $\mathcal{X}$ at level t, is given as $\mathcal{X}_t = (X_t, u_{X_t}, \mu_{X_t})$, where (X_t, u_{X_t}) is the quotient of the ultrametric space (X, u_X) at level t (see Section 2.2) and $\mu_{X_t} \in \mathcal{P}(X_t)$ is the push forward of μ_X under the canonical quotient map $Q_t : (X, u_X) \to (X_t, u_{X_t})$ sending x to $[x]_t$ for $x \in X$. Figure 4 illustrates the weighted quotient in a simple example. Based on this definition, we show the following theorem, whose proof is postponed to Appendix B.2.3.

Theorem 3.13. *Let* $\mathcal{X} = (X, u_X, \mu_X)$ *and* $\mathcal{Y} = (Y, u_Y, \mu_Y)$ *be two compact ultrametric measure spaces. Then, it holds that*

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \min\left\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\right\}.$$

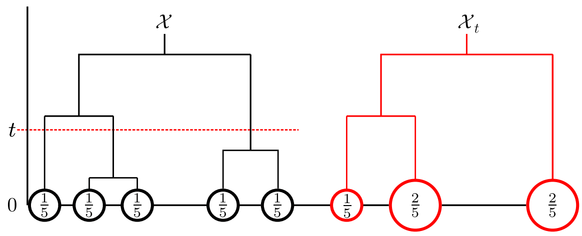

Fig. 4: Weighted Quotient: An ultrametric measure space (black) and its weighted quotient at level t (red).

Remark 3.14. The weighted quotients $\mathcal{X}_t$ and $\mathcal{Y}_t$ can be considered as vertex weighted, rooted trees and thus it is possible to verify whether $\mathcal{X}_t \cong_w \mathcal{Y}_t$ in polynomial time [3]. In consequence, we obtain an polynomial time algorithm for the calculation of $u_{\mathrm{GW},\infty}$. See Section 6.1.2 for details.

The representations of u_{GH} in Theorem 2.5 and $u_{\mathrm{GW},\infty}$ in Theorem 3.13 strongly resemble themselves. As a direct consequence of both Theorem 2.5 and Theorem 3.13, we obtain the following comparison between the two metrics

Corollary 3.15. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, it holds that*

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) \geq u_{\mathrm{GH}}(X, Y). \tag{21}$$

The inequality in Equation (21) is sharp and we illustrate this as follows. By Mémoli and Wan [70, Corollary 5.8] we know that if the considered ultrametric spaces (X, u_X) and (Y, u_Y) have different diameters (w.l.o.g. $\mathrm{diam}\,(X) < \mathrm{diam}\,(Y)$), then $u_{\mathrm{GH}}(X, Y) = \mathrm{diam}\,(Y)$. The same statement also holds for $u_{\mathrm{GW},\infty}$

Corollary 3.16. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ be such that* $\mathrm{diam}\,(X) < \mathrm{diam}\,(Y)$. *Then,*

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = \mathrm{diam}\,(Y) = u_{\mathrm{GH}}(X, Y).$$

Proof. The rightmost equality follows directly from Corollary 5.8 of Mémoli and Wan [70]. As for the leftmost equality, let $t := \mathrm{diam}\,(Y)$, then it is obvious that $\mathcal{X}_t \cong_w * \cong_w \mathcal{Y}_t$, where $*$ denotes the one point ultrametric measure space. Let $s \in (\mathrm{diam}\,(X), \mathrm{diam}\,(Y))$, then $\mathcal{X}_t \cong_w *$ whereas $\mathcal{Y} \ncong_w *$. By Theorem 3.13, $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = t = \mathrm{diam}\,(Y)$. □

3.3 The relation between $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In this section, we study the relation of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$, $1 \leq p \leq, \infty$ and establish the topological equivalence between the two metrics.

3.3.1 Lipschitz relation

We first study the Lipschitz relation between $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$. For this purpose, we have to distinguish the cases $p < \infty$ and $p = \infty$.

The case $p < \infty$. We start the consideration of this case by proving that it is essentially enough to consider the case $p = 1$ (see Theorem 3.17). To this end, we need to introduce some notation. For each $\alpha > 0$, we define a function $S_\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ by $x \mapsto x^\alpha$. Given an ultrametric space (X, u_X) and $\alpha > 0$, we abuse the notation and denote by $S_\alpha(X)$ the new space $(X, S_\alpha \circ u_X)$. It is obvious that $S_\alpha(X)$ is still an ultrametric space. This transformation of metric spaces is also known as the *snowflake transform* [26]. Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ denote two ultrametric measure spaces. Let $1 \leq p < \infty$. We denote by $S_p(\mathcal{X})$ the ultrametric measure space $(X, S_p \circ u_X, \mu_X)$. The snowflake transform can be used to relate $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ as well as $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ with $u_{\mathrm{GW},1}(S_p(\mathcal{X}), S_p(\mathcal{Y}))$ and $u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}), S_p(\mathcal{Y}))$, respectively.

Theorem 3.17. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ and let $p \in [1, \infty)$. Then, we obtain*

$$\left(u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})\right)^p = u_{\mathrm{GW},1}(S_p(\mathcal{X}), S_p(\mathcal{Y})) \quad \textit{and} \quad \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})\right)^p = u_{\mathrm{GW},1}^{\mathrm{sturm}}(S_p(\mathcal{X}), S_p(\mathcal{Y})).$$

We give full proof of Theorem 3.17 in Appendix B.2.4. Based on this result, we can directly relate the metrics $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ by only considering the case $p = 1$ and prove the following Theorem 3.18 (see Appendix B.3.1 for a detailed proof).

Theorem 3.18. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, we have for $p \in [1, \infty)$ that*

$$u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) \leq 2^{\frac{1}{p}}\, u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}).$$

The subsequent example verifies that the coefficient in Theorem 3.18 is tight.

Example 3.19. For each $n \in \mathbb{N}$, let $\mathcal{X}_n$ be the three-point space $\Delta_3(1)$ (i.e. the 3-point metric labeled by $\{x_1, x_2, x_3\}$ where all distances are 1) with a probability measure μ_X^n such that $\mu_X^n(x_1) = \mu_X^n(x_2) = \frac{1}{2n}$ and $\mu_X^n(x_3) = 1 - \frac{1}{n}$. Let $Y = *$ and μ_Y be the only probability measure on Y. Then, it is routine (using Proposition B.23 from Appendix B.5.3) to check that $u_{\mathrm{GW},1}(\mathcal{X}_n, \mathcal{Y}) = \frac{2}{n}\left(1 - \frac{3}{4n}\right)$ and $u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}_n, \mathcal{Y}) = \frac{1}{n}$. Therefore, we have

$$\lim_{n \to \infty} \frac{u_{\mathrm{GW},1}(\mathcal{X}_n, \mathcal{Y})}{u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}_n, \mathcal{Y})} = 2.$$

Example 3.20 ($u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ are not bi-Lipschitz equivalent)**.** Following [67, Remark 5.17], we verify in Appendix B.3.2 that for any positive integer n

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}\left(\hat{\Delta}_n(1), \hat{\Delta}_{2n}(1)\right) \geq \frac{1}{4} \text{ and } u_{\mathrm{GW},p}\left(\hat{\Delta}_n(1), \hat{\Delta}_{2n}(1)\right) \leq \left(\frac{3}{2n}\right)^{\frac{1}{p}}.$$

Here, $\hat{\Delta}_n(1)$ denotes the n-point metric measure space with interpoint distance 1 and the uniform probability measure. Thus, there exists no constant $C > 0$ such that the inequality $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \leq C \cdot u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ holds for every input spaces $\mathcal{X}$ and $\mathcal{Y}$. Hence, $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ are not bi-Lipschitz equivalent.

The case $p = \infty$. Next, we consider the relation between $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},\infty}$. By taking the limit $p \to \infty$ in Theorem 3.18, one might expect that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}} \geq u_{\mathrm{GW},\infty}$. In fact, we prove that the equality holds (for the full proof see Appendix B.3.3).

Theorem 3.21. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Then, it holds that*

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}).$$

One application of Theorem 3.21 is to explicitly derive the minimizing pair $(A, \phi) \in \mathcal{A}^*$ in Equation (31) for $p = \infty$ (see Appendix B.3.4 for an explicit construction):

Theorem 3.22. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Let $s := u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$ and assume that $s > 0$. Then, there exists $(A, \phi) \in \mathcal{A}$ defined in Equation (19) such that*

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y),$$

where Z_A denotes the ultrametric space defined in Section 3.1.1.

3.3.2 Topological equivalence between $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

Mémoli [67] proved the topological equivalence between $d_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$. We establish an analogous result for $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$. To this end, we recall the *modulus of mass distribution.*

Definition 3.23 (Greven et al. [39, Def. 2.9])**.** Given $\delta > 0$ we define *the modulus of mass distribution* of $\mathcal{X} \in \mathcal{U}^w$ as

$$v_\delta(\mathcal{X}) := \inf \left\{ \varepsilon > 0 \middle| \, \mu_X \left(\{ x : \, \mu_X \left(B_\varepsilon^\circ(x) \right) \leq \delta \} \right) \leq \varepsilon \right\}, \tag{22}$$

where $B_\varepsilon^\circ(x)$ denotes the *open* ball centered at x with radius ε.

We note that $v_\delta(\mathcal{X})$ is non-decreasing, right-continuous and bounded above by 1. Furthermore, it holds that $\lim_{\delta \searrow 0} v_\delta(\mathcal{X}) = 0$ [39, Lemma 6.5]. With Definition 3.23 at hand, we derive the following theorem.

Theorem 3.24. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$, $p \in [1, \infty)$ and $\delta \in \left(0, \frac{1}{2}\right)$. Then, whenever $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) < \delta^5$ we have*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \leq (4 \cdot \min(v_\delta(\mathcal{X}), v_\delta(\mathcal{Y})) + \delta)^{\frac{1}{p}} \cdot M,$$

where $M := 2 \cdot \max(\mathrm{diam}\,(X), \mathrm{diam}\,(Y)) + 54$.

Remark 3.25. Since it holds that $\lim_{\delta \searrow 0} v_\delta(\mathcal{X}) = 0$ and that $2^{-1/p} u_{\mathrm{GW},p}^{\mathrm{sturm}} \geq u_{\mathrm{GW},p}$ (see Theorem 3.18), the above theorem gives the topological equivalence between $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p < \infty$ (the topological equivalence between $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},\infty}$ holds trivially thanks to Theorem 3.21).

The proof of the Theorem 3.24 follows the same strategy used for proving Proposition 5.3 in [67] and we refer to Appendix B.3.5 for the details.

3.4 Topological and geodesic properties

In this section, we consider the topology induced by $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ on $\mathcal{U}^w$ and discuss the geodesic properties of both $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ for $1 \leq p \leq \infty$.

3.4.1 Completeness and separability

We study completeness and separability of the two metrics $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$, $1 \leq p \leq \infty$, on $\mathcal{U}^w$. To this end, we derive the subsequent theorem whose proof is postponed to Appendix B.4.1.

Theorem 3.26.

1. *For $p \in [1, \infty)$, the metric space $(\mathcal{U}^w, u_{\mathrm{GW},p})$ is neither complete nor separable.*
2. *For $p \in [1, \infty)$, the metric space $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is neither complete nor separable.*
3. *$(\mathcal{U}^w, u_{\mathrm{GW},\infty}) = (\mathcal{U}^w, u_{\mathrm{GW},\infty}^{\mathrm{sturm}})$ is complete but not separable.*

3.4.2 Geodesic property

A *geodesic* in a metric space (X, d_X) is a continuous function $\gamma : [0, 1] \to X$ such that for each $s, t \in [0, 1]$, $d_X(\gamma(s), \gamma(t)) = |s - t| \cdot d_X(\gamma(0), \gamma(1))$. We say a metric space is geodesic if for any two distinct points $x, x' \in X$, there exists a geodesic $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. For any $p \in [1, \infty)$, the notion of p-geodesic is introduced in [70]: A p-geodesic in a metric space (X, d_X) is a continuous function $\gamma : [0, 1] \to X$ such that for each $s, t \in [0, 1]$, $d_X(\gamma(s), \gamma(t)) = |s - t|^{1/p} \cdot d_X(\gamma(0), \gamma(1))$. Similarly, we say a metric space is p-geodesic if for any two distinct points $x, x' \in X$, there exists a p-geodesic $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = x'$. Note that a 1-geodesic is a usual geodesic and a 1-geodesic space is a usual geodesic space. The subsequent theorem establishes (p-)geodesic properties of $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ for $p \in [1, \infty)$. A full proof is given in Appendix B.4.2.

Theorem 3.27. *For any $p \in [1,\infty)$, the space $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is p-geodesic.*

Remark 3.28. Due to the fact that a p-geodesic space cannot be geodesic when $p > 1$ (cf. Lemma B.15), $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is not geodesic for all $p > 1$.

Remark 3.29. Though the geodesic properties of $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$, $1 \le p < \infty$ are clear, we remark that geodesic properties of $(\mathcal{U}^w, u_{\mathrm{GW},p})$, $1 \le p < \infty$, still remain unknown to us.

Remark 3.30 (The case $p = \infty$)**.** Being an ultrametric space itself (cf. Theorem 3.10), $(\mathcal{U}^w, u_{\mathrm{GW},\infty})$ $\left(= \left(\mathcal{U}^w, u_{\mathrm{GW},\infty}^{\mathrm{sturm}}\right)\right)$ is *totally disconnected*, i.e., any subspace with at least two elements is disconnected [89]. This in turn implies that each continuous curve in $(\mathcal{U}^w, u_{\mathrm{GW},\infty})$ is constant. Therefore, $(\mathcal{U}^w, u_{\mathrm{GW},\infty})$ is not a p-geodesic space for any $p \in [1,\infty)$.

4 Lower bounds for $u_{\mathrm{GW},p}$

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two ultrametric measure spaces. The metrics $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ respect the ultrametric structure of the spaces $\mathcal{X}$ and $\mathcal{Y}$. Thus, one would hope that comparing ultrametric measure spaces with $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ or $u_{\mathrm{GW},p}$ is more meaningful than doing it with the usual Gromov-Wasserstein distance or Sturm's distance. Unfortunately, for $p < \infty$, the computation of both $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $u_{\mathrm{GW},p}$ is complicated and for $p = \infty$ both metrics are extremely sensitive to differences in the diameters of the considered spaces (see Corollary 3.16). Thus, it is not feasible to use these metrics in many applications. However, we can derive meaningful lower bounds for $u_{\mathrm{GW},p}$ (and hence also for $u_{\mathrm{GW},p}^{\mathrm{sturm}}$) that resemble those of the Gromov-Wasserstein distance. Naturally, the question arises whether these lower bounds are better/sharper than the ones of the usual Gromov-Wasserstein distance in this setting. This question is addressed throughout this section and will be readdressed in Section 6 as well as Section 7.

In [67], the author introduced three lower bounds for $d_{\mathrm{GW},p}$ that are computationally less expensive than the calculation of $d_{\mathrm{GW},p}$. We will briefly review these three lower bounds and then define candidates for the corresponding lower bounds for $u_{\mathrm{GW},p}$. In the following, we always assume $p \in [1,\infty]$.

First lower bound Let $s_{X,p} : X \to \mathbb{R}_{\ge 0}$, $x \mapsto \|u_X(x,\cdot)\|_{L^p(\mu_X)}$. Then, the first lower bound $\mathbf{FLB}_p(\mathcal{X},\mathcal{Y})$ for $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$ is defined as follows

$$\mathbf{FLB}_p(\mathcal{X},\mathcal{Y}) \coloneqq \frac{1}{2} \inf_{\mu \in \mathcal{C}(\mu_X,\mu_Y)} \left\| \Lambda_1(s_{X,p}(\cdot), s_{Y,p}(\cdot)) \right\|_{L^p(\mu)}.$$

Following our intuition of replacing Λ_1 with Λ_∞, we define the ultrametric version of **FLB** as

$$\mathbf{FLB}_p^{\text{ult}}(\mathcal{X},\mathcal{Y}) \coloneqq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \|\Lambda_\infty(s_{X,p}(\cdot), s_{Y,p}(\cdot))\|_{L^p(\mu)}.$$

Second lower bound The second lower bound $\mathbf{SLB}_p(\mathcal{X},\mathcal{Y})$ for $d_{\text{GW},p}(\mathcal{X},\mathcal{Y})$ is given as

$$\mathbf{SLB}_p(\mathcal{X},\mathcal{Y}) \coloneqq \frac{1}{2}\inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|\Lambda_1(u_X,u_Y)\|_{L^p(\gamma)}.$$

Thus, we define the ultrametric second lower bound between two ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ as follows:

$$\mathbf{SLB}_p^{\text{ult}}(\mathcal{X},\mathcal{Y}) \coloneqq \inf_{\gamma\in\mathcal{C}(\mu_X\otimes\mu_X,\mu_Y\otimes\mu_Y)} \|\Lambda_\infty(u_X,u_Y)\|_{L^p(\gamma)}.$$

Third lower bound Before we introduce the final lower bound, we have to define several functions. First, let $\Gamma^1_{X,Y} : X\times Y\times X\times Y \to \mathbb{R}_{\geq 0}$, $(x,y,x',y') \mapsto \Lambda_1(u_X(x,x'),u_Y(y,y'))$ and let $\Omega^1_p : X\times Y \to \mathbb{R}_{\geq 0}$, $p\in[1,\infty]$, be given by

$$\Omega^1_p(x,y) \coloneqq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \left\|\Gamma^1_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^p(\mu)}.$$

Then, the third lower bound $\mathbf{TLB}_p$ is given as

$$\mathbf{TLB}_p(\mathcal{X},\mathcal{Y}) \coloneqq \frac{1}{2}\inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \left\|\Omega^1_p(\cdot,\cdot)\right\|_{L^p(\mu)}.$$

Analogously to the definition of previous ultrametric versions, we define $\Gamma^\infty_{X,Y} : X\times Y\times X\times Y \to \mathbb{R}_{\geq 0}$, $(x,y,x',y') \mapsto \Lambda_\infty(u_X(x,x'),u_Y(y,y'))$. Further, for $p\in[1,\infty]$, let $\Omega^\infty_p : X\times Y \to \mathbb{R}_{\geq 0}$ be given by

$$\Omega^\infty_p(x,y) \coloneqq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \left\|\Gamma^\infty_{X,Y}(x,y,\cdot,\cdot)\right\|_{L^p(\mu)}.$$

Then, the ultrametric third lower bound between two ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ is defined as

$$\mathbf{TLB}_p^{\text{ult}}(\mathcal{X},\mathcal{Y}) \coloneqq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)} \left\|\Omega^\infty_p(\cdot,\cdot)\right\|_{L^p(\mu)}.$$

4.1 Properties and computation of the lower bounds

Next, we examine the quantities $\mathbf{FLB}^{\text{ult}}, \mathbf{SLB}^{\text{ult}}$ and $\mathbf{TLB}^{\text{ult}}$ more closely. Since we have $\Lambda_\infty(a,b) \geq \Lambda_1(a,b) = |a-b|$ for any $a, b \geq 0$, it is easy to conclude that $\mathbf{FLB}_p^{\text{ult}} \geq \mathbf{FLB}_p$, $\mathbf{SLB}_p^{\text{ult}} \geq \mathbf{SLB}_p$ and $\mathbf{TLB}_p^{\text{ult}} \geq \mathbf{TLB}_p$. Moreover, the three ultrametric lower bounds satisfy the following theorem (for a complete proof see Appendix C.1.1).

Theorem 4.1. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ and let $p \in [1, \infty]$.*

1. $u_{\text{GW},\infty}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{FLB}_\infty^{\text{ult}}(\mathcal{X}, \mathcal{Y})$.
2. $u_{\text{GW},p}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{TLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) \geq \mathbf{SLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y})$.

Remark 4.2. Interestingly, it turns out that $\mathbf{FLB}_p^{\text{ult}}$ is not a lower bound of $u_{\text{GW},p}$ in general when $p < \infty$. For example, let $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ and define u_X such that $u_X(x_1, x_2) = 1$ and $u_X(x_i, x_j) = 2\delta_{i \neq j}$ for $(i,j) \neq (1,2)$, $(i,j) \neq (2,1)$ and $i, j = 1, \ldots, n$. Let $u_Y(y_i, y_j) = 2\delta_{i \neq j}$, $i, j = 1, \ldots, n$, and let μ_X and μ_Y be uniform measures on X and Y, respectively. Then, $u_{\text{GW},1}(\mathcal{X}, \mathcal{Y}) \leq \frac{4}{n^2}$ whereas $\mathbf{FLB}_1^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = \frac{4n-4}{n^2}$ which is greater than $u_{\text{GW},1}(\mathcal{X}, \mathcal{Y})$ as long as $n > 2$. Moreover, we have in this case that $\mathbf{FLB}_1^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = O\left(\frac{1}{n}\right)$ whereas $u_{\text{GW},1}(\mathcal{X}, \mathcal{Y}) = O\left(\frac{1}{n^2}\right)$. Hence, there exists no constant $C > 0$ such that $\mathbf{FLB}_1^{\text{ult}} \leq C \cdot u_{\text{GW},1}$ in general.

Remark 4.3. There exist ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathbf{TLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = 0$ whereas $u_{\text{GW},p}(\mathcal{X}, \mathcal{Y}) > 0$ (examples described in [67, Figure 8] will serve the purpose). Furthermore, there are spaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\mathbf{SLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = 0$ whereas $\mathbf{TLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) > 0$ (see Appendix C.1.3). The analogous statement holds true for $\mathbf{TLB}_p$ and $\mathbf{SLB}_p$, which are nevertheless useful in various applications (see e.g. [37]).

From the structure of $\mathbf{SLB}_p^{\text{ult}}$ and $\mathbf{TLB}_p^{\text{ult}}$ it is obvious that their computations leads to different optimal transport problems (see e.g. [99]). However, in analogy to Chowdhury and Mémoli [23, Theorem 3.1] we can rewrite $\mathbf{SLB}_p^{\text{ult}}$ and $\mathbf{TLB}_p^{\text{ult}}$ in order to further simplify their computation. The full proof of the subsequent proposition is given in Appendix C.1.2.

Proposition 4.4. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ and let $p \in [1, \infty]$. Then, we find that*

1. $\mathbf{SLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = d_{\text{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}\left((u_X)_\#(\mu_X \otimes \mu_X), (u_Y)_\#(\mu_Y \otimes \mu_Y)\right)$;
2. *For each* $x, y \in X \times Y$, $\Omega_p^\infty(x,y) = d_{\text{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}\left(u_X(x,\cdot)_\#\mu_X, u_Y(y,\cdot)_\#\mu_Y\right)$.

Remark 4.5. Since we have by Theorem 2.9 an explicit formula for the Wasserstein distance on $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ between finitely supported probability measures, these alternative representations of the lower bound $\mathbf{SLB}_p^{\text{ult}}$ and the cost functional Ω_p^∞ drastically reduce the computation time of $\mathbf{SLB}_p^{\text{ult}}$ and $\mathbf{TLB}_p^{\text{ult}}$, respectively. In particular, we note that this allows us to compute $\mathbf{SLB}_p^{\text{ult}}$, $1 \le p \le \infty$, between finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$ with $|X| = m$ and $|Y| = n$ in $O((m \vee n)^2)$ steps.

Proposition 4.4 allows us to direclty compare the two lower bounds $\mathbf{SLB}_1^{\text{ult}}$ and $\mathbf{SLB}_1$.

Corollary 4.6. *For any finite ultrametric measure spaces $\mathcal{X}$ and $\mathcal{Y}$, we have that*

$$\mathbf{SLB}_1^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1(\mathcal{X}, \mathcal{Y}) + \frac{1}{2} \int_{\mathbb{R}} t \, |(u_X)_\#(\mu_X \otimes \mu_X) - (u_Y)_\#(\mu_Y \otimes \mu_Y)| \, (dt). \quad (23)$$

Proof. The claim follows directly from Proposition 4.4 and Remark 2.10. □

This corollary implies that $\mathbf{SLB}_p^{\text{ult}}$ is more rigid than $\mathbf{SLB}_p$, since the second summand on the right hand side of Equation (23) is sensitive to distance perturbations. This is also illustrated very well in the subsequent example.

Example 4.7. Recall notations from Example 3.5. For any $d, d' > 0$, we let $X := \Delta_2(d)$ and let $Y := \Delta_2(d')$. Assume that X and Y have underlying sets $\{x_1, x_2\}$ and $\{y_1, y_2\}$, respectively. Define $\mu_X \in \mathcal{P}(X)$ and $\mu_Y \in \mathcal{P}(Y)$ as follows. Let $\alpha_1, \alpha_2 \ge 0$ be such that $\alpha_1 + \alpha_2 = 1$. Let $\mu_X(x_1) = \mu_Y(y_1) := \alpha_1$ and let $\mu_X(x_2) = \mu_Y(y_2) := \alpha_2$. Then, it is easy to verify that

1. $u_{\text{GW},1}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = 2\alpha_1\alpha_2\Lambda_\infty(d, d')$.
2. $d_{\text{GW},1}(\mathcal{X}, \mathcal{Y}) = \mathbf{SLB}_1(\mathcal{X}, \mathcal{Y}) = \alpha_1\alpha_2\Lambda_1(d, d') = \alpha_1\alpha_2|d - d'|$.
3. $\frac{1}{2} \int_{\mathbb{R}} t \, |(u_X)_\#(\mu_X \otimes \mu_X) - (u_Y)_\#(\mu_Y \otimes \mu_Y)| \, (dt) = \alpha_1\alpha_2(d + d')\delta_{d \ne d'}$.

From 1 and 2 we observe that both second lower bounds are tight. Moreover, since we obviously have that $(d+d')\delta_{d\ne d'} + |d-d'| = 2\Lambda_\infty(d, d')$, we have also verified Equation (23) through this example. Unlike $\mathbf{SLB}_1(\mathcal{X}, \mathcal{Y})$ being proportional to $|d - d'|$, as long as $d \ne d'$, even if $|d-d'|$ is small, $\Lambda_\infty(d, d') = \max(d, d')$ which results in a large value of $\mathbf{SLB}_1^{\text{ult}}(\mathcal{X}, \mathcal{Y})$ when d and d' are large numbers. This example illustrates that $\mathbf{SLB}_1^{\text{ult}}$ (and hence $u_{\text{GW},1}$) is rigid with respect to distance perturbation.

5 $u_{\text{GW},p}$ on ultra-dissimilarity spaces

A natural generalization of ultrametric spaces is provided by *ultra-dissimilarity spaces*. These spaces naturally occur when working with symmetric ultranetworks (see [91]) or phylogenetic tree data (see [90]). In this section, we will introduce these spaces and briefly illustrate to what extend the results for $u_{\text{GW},p}$ can be adapted for ultra-dissimilarity measure spaces. We start by formally introducing *ultra-dissimilarity spaces*.

Definition 5.1 (Ultra-dissimilarity spaces)**.** An *ultra-dissimilarity* space (X, u_X) is a couple that consists of a set X and a function $u_X : X \times X \to \mathbb{R}_{\geq 0}$ satisfying the following conditions for any $x, y, z \in X$:

1. $u_X(x, y) = u_X(y, x)$;
2. $u_X(x, y) \leq \max(u_X(x, z), u_X(z, y))$;
3. $\max(u_X(x, x), u_X(y, y)) \leq u_X(x, y)$ and the equality holds if and only if $x = y$.

Remark 5.2. Note that when (X, u_X) is an ultrametric space the third condition is trivially satisfied.

In the following, we restrict ourselves to *finite* ultra-dissimilarity spaces to avoid technical issues in topology (see [22, 23] for a more complete treatment of infinite spaces). One important aspect of ultra-dissimilarity spaces is the connection with the so-called *treegrams* [70, 91], which can be regarded as generalized dendrograms. For a finite set X, let $\textbf{SubPart}(X)$ denote the collection of all *subpartitions* of X: Any partition P' of a non-empty subset $X' \subseteq X$ is called a subpartition of X. Given two subpartitions P_1, P_2, we say P_1 is coarser than P_2 if each block in P_2 is contained in some block in P_1.

Definition 5.3 (Treegrams)**.** A *treegram* $T_X : [0, \infty) \to \textbf{SubPart}(X)$ is a map parametrizing a nested family of subpartitions over the same set X and satisfying the following conditions:

1. For any $0 \leq s < t < \infty$, $T_X(t)$ is coarser than $T_X(s)$;
2. There exists $t_X > 0$ such that for any $t \geq t_X$, $T_X(t) = \{X\}$;
3. For each $t \geq 0$, there exists $\varepsilon > 0$ such that $T_X(t) = T_X(t')$ for all $t' \in [t, t + \varepsilon]$;
4. For each $x \in X$, there exists $t_x \geq 0$ such that $\{x\}$ is a block in $T_X(t_x)$.

Similar to Theorem 2.2, which correlates ultrametrics to dendrograms, there exists an equivalence relation between ultra-dissimilarity functions and treegrams on a finite set (see Figure 5 for an illustration).

Proposition 5.4 (Smith et al. [91])**.** *Given a finite set X, denote by $\mathcal{U}_{\text{dis}}(X)$ the collection of all ultrametric dissimilarity functions on X and by $\mathcal{T}(X)$ the collection of all treegrams over X. Then, there exists a bijection $\Upsilon_X : \mathcal{T}(X) \to \mathcal{U}_{\text{dis}}(X)$.*

An *ultra-dissimilarity measure space* is a triple $\mathcal{X} = (X, u_X, \mu_X)$ where (X, u_X) is an ultra-dissimilarity space and μ_X is a probability measure fully supported on X. Just as for metric spaces or metric measure spaces, it is important to have a notion of isomorphism between ultra-dissimilarity spaces.

Definition 5.5 (Isomorphism)**.** Given two ultra-dissimilarity measure spaces $\mathcal{X}$ and $\mathcal{Y}$, we say they are *isomorphic*, denoted $\mathcal{X} \cong_w \mathcal{Y}$, if there is a bijective function $f : X \to Y$ such

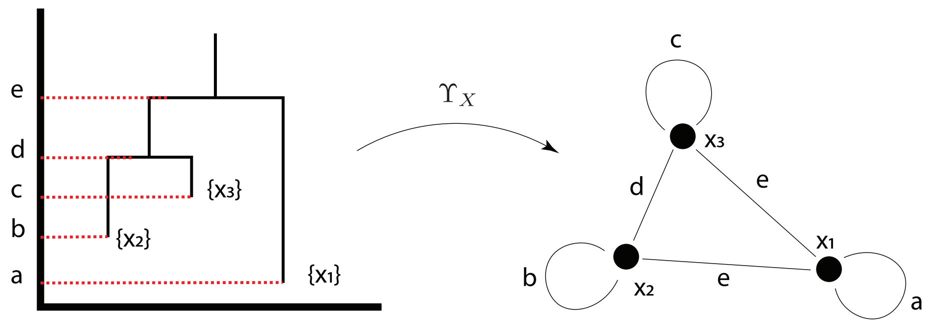

Fig. 5: Treegrams: Relation between ultra-dissimilarity functions and treegrams

that $f_{\#}\mu_X = \mu_Y$ and for any $x, x' \in X$ it holds $u_Y(f(x), f(x')) = u_X(x, x')$. The collection of all isomorphism classes of ultra-dissimilarity spaces is denoted by $\mathcal{U}^w_{\text{dis}}$.

Given the previous results it is straightforward to show that $u_{\text{GW},p}$, $1 \leq p \leq \infty$, is a metric on the isomorphism classes of $\mathcal{U}^w_{\text{dis}}$. For the complete proof of the subsequent statement, we refer to Appendix D.1.1.

Theorem 5.6. *The ultrametric Gromov-Wasserstein distance $u_{\text{GW},p}$ is a p-metric on $\mathcal{U}^w_{\text{dis}}$.*

Remark 5.7. Since $u_{\text{GW},p}$ translates to a metric on $\mathcal{U}^w_{\text{dis}}$, it is clear that it admits the lower bounds introduced in Section 4.

6 Computational aspects

In this section, we investigate algorithms for approximating/calculating $u_{\text{GW},p}$, $1 \leq p \leq \infty$. Furthermore, we evaluate for $p < \infty$ the performance of the computationally efficient lower bound $\mathbf{SLB}_p^{\text{ult}}$ introduced in Section 4 and compare our findings to the results of the classical Gromov-Wasserstein distance $d_{\text{GW},p}$ (see Equation (7)). Matlab implementations of the presented algorithms and comparisons are available at `https://github.com/ndag/uGW`.

6.1 Algorithms

Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be two finite ultrametric measure spaces with cardinalities m and n, respectively.

6.1.1 The case $p < \infty$

We have already noted in Remark 3.12 that calculating $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ for $p < \infty$ yields a non-convex quadratic program (which is an NP-hard problem in general [78]). Solving this is not feasible in practice. However, in many practical applications it is sufficient to work with good approximations. Therefore, we propose to approximate $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y})$ for $p < \infty$ via conditional gradient descent. To this end, we note that the gradient G that arises from Equation (12) can in the present setting be expressed with the following partial derivative with respect to $\mu \in \mathcal{C}(\mu_X, \mu_Y)$

$$G_{i,j} = 2\sum_{k=1}^{m}\sum_{l=1}^{n}(\Lambda_\infty(u_X(x_i,x_k),u_Y(y_j,y_l)))^p\mu_{kl}, \quad \forall 1 \le i \le m, 1 \le j \le n. \tag{24}$$

As we deal with a non-convex minimization problem, the performance of the gradient descent strongly depends on the starting coupling $\mu^{(0)}$. Therefore, we follow the suggestion of Chowdhury and Needham [24] and employ a Markov Chain Monte Carlo Hit-And-Run sampler to obtain multiple random start couplings. Running the gradient descent from each point in this ensemble greatly improves the approximation in many cases. For a precise description of the proposed procedure, we refer to Algorithm 1.

Algorithm 1 $u_{\mathrm{GW},p}(X, Y, p, N, L)$

//Create a list of random couplings
couplings =CreateRandomCouplings(N)
stat_points = cell(N)
for i=1:N **do**
  $\mu^{(0)}$ =couplings$\{i\}$
  for j=1:L **do**
    G = Gradient from Equation (24) w.r.t. $\mu^{(j-1)}$
    $\tilde{\mu}^{(j)}$ = Solve OT with ground loss G
    $\gamma^{(j)} = \frac{2}{j+2}$
    //Alt. find $\gamma \in [0,1]$ *that minimizes* $\mathrm{dis}_p^{\mathrm{ult}}\left(\mu^{(j-1)} + \gamma\big(\tilde{\mu}^{(j)} - \mu^{(j-1)}\big)\right)$
    $\mu^{(j)} = (1-\gamma^{(j)})\mu^{(j-1)} + \gamma^{(j)}\tilde{\mu}^{(j)}$
  end for
  stat_points$\{i\}$= $\mu^{(L)}$
end for
Find μ^* in stat_points that minimizes $\mathrm{dis}_p^{\mathrm{ult}}(\mu)$
result =$\mathrm{dis}_p^{\mathrm{ult}}(\mu^*)$

6.1.2 The case $p = \infty$

For $p = \infty$, it follows by Theorem 3.13 that

$$u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y}) = \inf \{ t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t \} \,. \tag{25}$$

This identity allows us to construct a polynomial time algorithm for $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$ based on the ideas of Mémoli and Wan [70, Sec. 8.2.2]. More precisely, let

$$\mathrm{spec}\,(X) \coloneqq \{ u_X(x, x') \,|\, x, x' \in X \}$$

denote the spectrum of X. Then, it is evident that in order to find the infimum in Equation (25), we only have to check $\mathcal{X}_t \cong_w \mathcal{Y}_t$ for each $t \in \mathrm{spec}\,(X) \cup \mathrm{spec}\,(Y)$, starting from the largest to the smallest and $u_{\mathrm{GW},\infty}$ is given as the smallest t such that $\mathcal{X}_t \cong_w \mathcal{Y}_t$. This can be done in polynomial time by considering $\mathcal{X}_t$ and $\mathcal{Y}_t$ as labeled, weighted trees (e.g. by using a slight modification of the algorithm in Example 3.2 of [3]). This gives rise to a simple algorithm (see Algorithm 2) to calculate $u_{\mathrm{GW},\infty}$.

Algorithm 2 $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{Y})$

```
spec = sort(spec (X) ∪ spec (Y), 'descent')
for i = 1 : length(spec) do
  t = spec(i)
  if X_t ≇_w Y_t then
    return spec(i − 1)
  end if
end for
return 0
```

6.2 The relation between $u_{\mathrm{GW},1}$, $u_{\mathrm{GW},\infty}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$

In order to understand how $u_{\mathrm{GW},p}$ (or at least its approximation), $u_{\mathrm{GW},\infty}$ and $\mathbf{SLB}_p^{\mathrm{ult}}$ are influenced by small changes in the structure of the considered ultrametric measure spaces, we exemplarily consider the ultrametric measure spaces $\mathcal{X}_i = (X_i, d_{X_i}, \mu_{X_i})$, $1 \leq i \leq 4$, displayed in Figure 6. These ultrametric measure spaces differ only by one characteristic (e.g. one side length or the equipped measure). Exemplarily, we calculate $u_{\mathrm{GW},1}(\mathcal{X}_i, \mathcal{X}_j)$ (approximated with Algorithm 1, where $L = 5000$ and $N = 40$), $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}_i, \mathcal{X}_j)$ and $u_{\mathrm{GW},\infty}(\mathcal{X}_i, \mathcal{X}_j)$, $1 \leq i, j \leq 4$. The results suggest that $\mathbf{SLB}_1^{\mathrm{ult}}$ and $u_{\mathrm{GW},1}$ are influenced by the change in the diameter of the spaces the most (see Table 2 and Table 3 in Appendix E.1 for the complete results). Changes in the metric influence $\mathbf{SLB}_1^{\mathrm{ult}}$ in a similar fashion as $u_{\mathrm{GW},1}$, while changes in the measure have less impact on $\mathbf{SLB}_1^{\mathrm{ult}}$. Further, we observe that

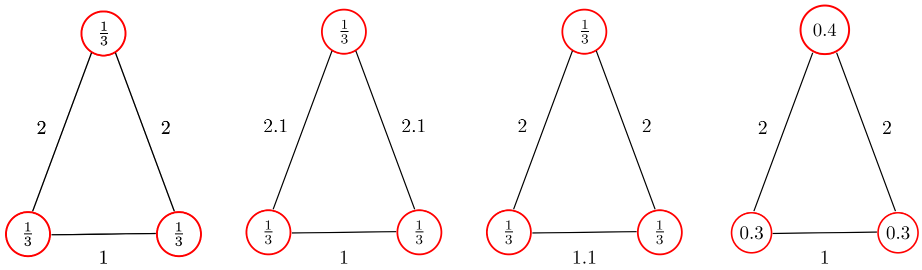

Fig. 6: Ultrametric measure spaces: Four non-isomorphic ultrametric measure spaces denoted (from left to right) as $\mathcal{X}_i = (X_i, d_{X_i}, \mu_{X_i})$, $1 \leq i \leq 4$.

$u_{\mathrm{GW},\infty}$ attains for almost all comparisons the maximal possible value. Only the comparison of $\mathcal{X}_1$ with $\mathcal{X}_3$, where the only small scale structure of the space was changed, yields a value that is smaller than the maximum of the diameters of the considered spaces.

6.3 Comparison of $u_{\mathrm{GW},1}$, $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$

In the remainder of this section, we will demonstrate the differences between $u_{\mathrm{GW},1}$, $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$. To this end, we first compare the metric measure spaces in Figure 6 based on $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$. We observe that $d_{\mathrm{GW},1}$ (approximated in the same manner as $u_{\mathrm{GW},1}$) and $\mathbf{SLB}_1$ are hardly influenced by the differences between the ultrametric measure spaces $\mathcal{X}_i$, $1 \leq i \leq 4$. In particular, it is remarkable that $d_{\mathrm{GW},1}$ is affected the most by the changes made to the measure and not the metric structure (see Table 4 in Appendix E.2 for the complete results).

Next, we consider the differences between the aforementioned quantities more generally. For this purpose, we generate 4 ultrametric spaces Z_k, $1 \leq k \leq 4$, with totally different dendrogram structures, whose diameters are between 0.5 and 0.6 (for the precise construction of these spaces see Appendix E.2). For each $t = 0, 0.2, 0.4, 0.6$, we perturb each Z_k independently to generate 15 ultrametric spaces $Z_{k,t}^i$, $1 \leq i \leq 15$, such that $(Z_{k,t}^i)_t \equiv (Z_k)_t$ for all i. The spaces $Z_{k,t}^i$ are called *pertubations of Z_k at level t* (see Figure 7 for an illustration and see Appendix E.2 for more details). The spaces $Z_{k,t}^i$ are endowed with the uniform probability measure and we obtain a collection of ultrametric measure spaces $\mathcal{Z}_{k,t}^i$. Naturally, we refer to k as the class of the ultrametric measure space $\mathcal{Z}_{k,t}^i$. We compute for each t the quantities $u_{\mathrm{GW},1}$, $\mathbf{SLB}_1^{\mathrm{ult}}$, $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ among the resulting 60 ultrametric measure spaces. The results, where the spaces have been ordered lexicographically by (k, i), are visualized in Figure 8. As previously, we observe that $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ as well as $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ behave in a similar manner. More precisely, we see that both $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ discriminate well between the different classes and that their behavior does not change too much for an increasing level of perturbation. On the other hand, $u_{\mathrm{GW},1}$ and

$\mathbf{SLB}_1^{\mathrm{ult}}$ are very sensitive to the level of perturbation. For small t they discriminate better than $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$ between the different classes and pick up clearly that the perturbed spaces differ. However, if the level of perturbation becomes too large both quantities start to discriminate between spaces from the same class (see Figure 8).

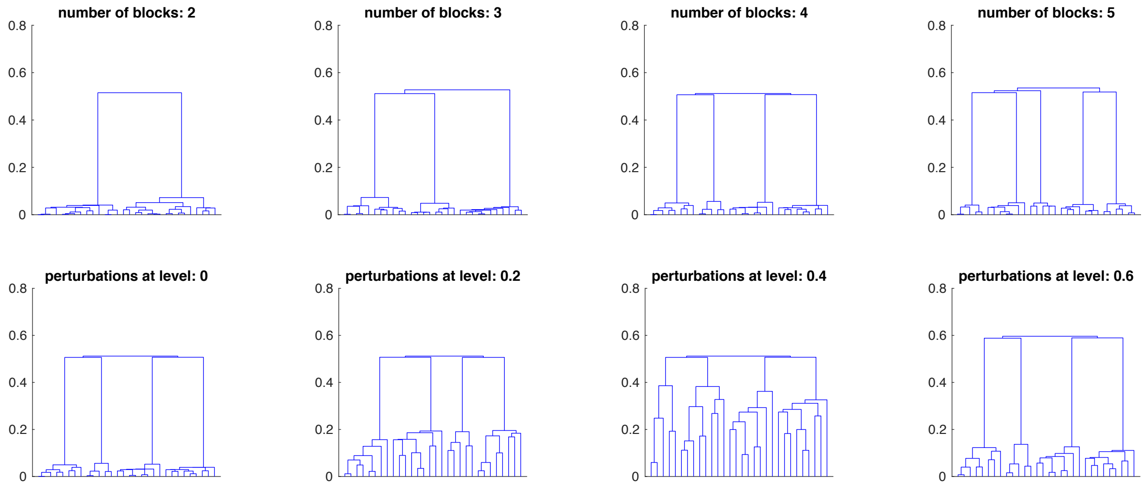

Fig. 7: Randomly sampled ultrametric measure spaces: Illustration of Z_k for $k = 2, 3, 4, 5$ (top row) and instances for perturbations of Z_4 with respect to perturbation level $t \in \{0, 0.2, 0.4, 0.6\}$ (bottom row).

In conclusion, $u_{\mathrm{GW},1}$ and $\mathbf{SLB}_1^{\mathrm{ult}}$ are sensitive to differences in the large scales of the considered ultrametric measure spaces. While this leads (from small t) to good discrimination in the above example, it also highlights that they are (different from $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$) susceptible to large scale noise.

7 Phylogenetic tree shapes

Rooted phylogenetic trees (for a formal definition see e.g., [90]) are a common tool to visualize and analyze the evolutionary relationship between different organisms. In combination with DNA sequencing, they are an important tool to study the rapid evolution of different pathogens. It is well known that the (unweighted) shape of a phylogenetic tree, i.e., the tree's connectivity structure without referring to its labels or the length of its branches, carries important information about macroevolutionary processes (see e.g., [8, 27, 72, 104]). In order to study the evolution of and the relation between different pathogens, it is of great interest to compare the shapes of phylogenetic trees created on the basis of different data sets. Currently, the number of tools for performing phylogenetic tree shape comparison is quite limited and the development of new methods for this is an

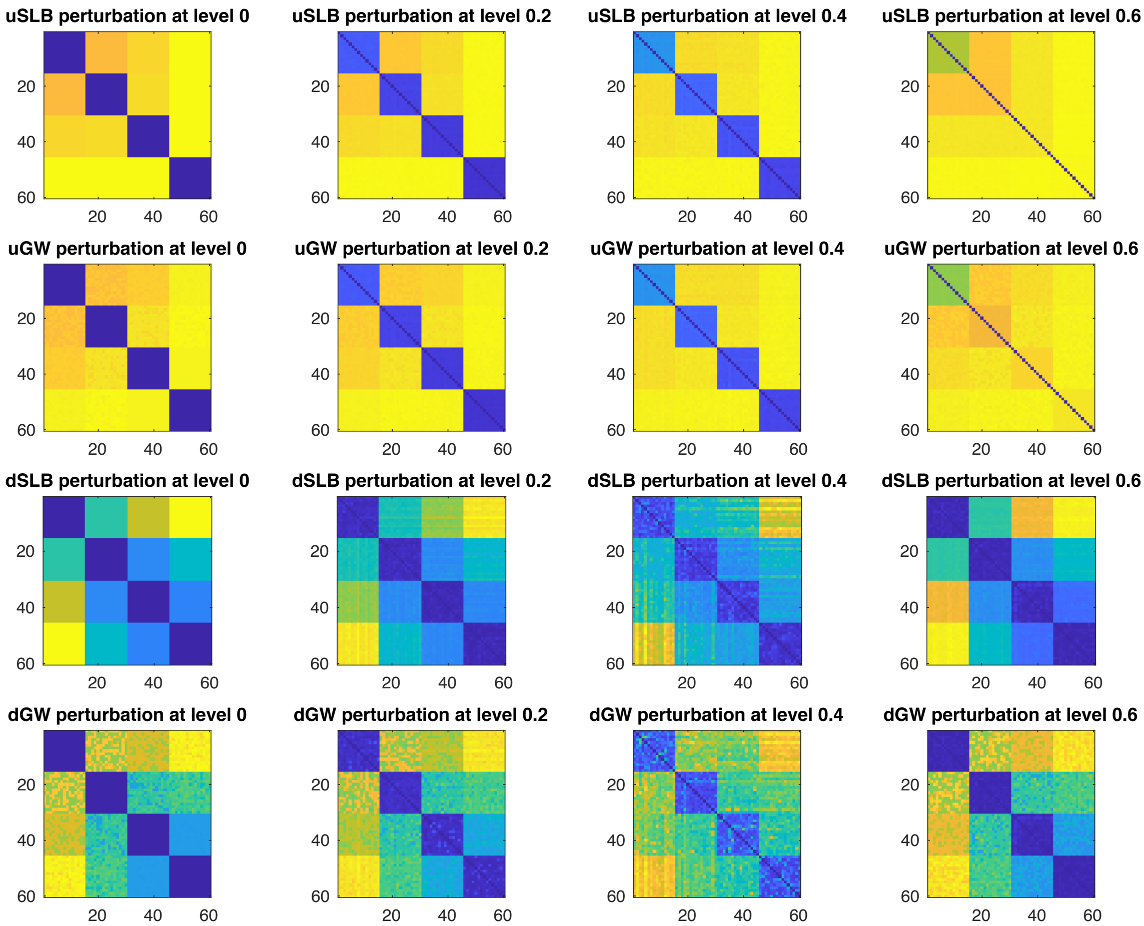

Fig. 8: $u_{\mathrm{GW},1}/\mathbf{SLB}_1^{\mathrm{ult}}$ **and** $d_{\mathrm{GW},1}/\mathbf{SLB}_1$ **among randomly generated ultrametric measure spaces:** Heatmap representations of $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{Z}_{n,t}^i,\mathcal{Z}_{n',t}^{i'})$ (top row), $u_{\mathrm{GW},1}(\mathcal{Z}_{n,t}^i,\mathcal{Z}_{n',t}^{i'})$ (second row), $\mathbf{SLB}_1(\mathcal{Z}_{n,t}^i,\mathcal{Z}_{n',t}^{i'})$ (third row) and $d_{\mathrm{GW},1}(\mathcal{Z}_{n,t}^i,\mathcal{Z}_{n',t}^{i'})$ (bottom row), $k,k'\in\{2,\ldots,5\}$ and $i,i'\in\{1,\ldots,15\}$.

active field of research [25, 49, 60, 73]. It is well known that certain classes of phylogenetic trees (as well as their respective tree shapes) can be identified as ultrametric spaces [90, Sec. 7]. On the other hand, general phylogenetic trees are closely related to treegrams (see Definition 5.3). In the following, we will use this connection and demonstrate exemplarily that the computationally efficient lower bound $\mathbf{SLB}_1^{\mathrm{ult}}$ has some potential for comparing phylogenetic tree shapes. In particular, we contrast it to the metric defined for this application in Equation (4) of Colijn and Plazzotta [25], in the following denoted as $d_{\mathrm{CP},2}$, and study the behavior of $\mathbf{SLB}_1$ in this framework.

In this section, we reconsider phylogenetic tree shape comparisons from Colijn and Plazzotta [25] and thereby study HA protein sequences from human influenza A (H3N2) (data

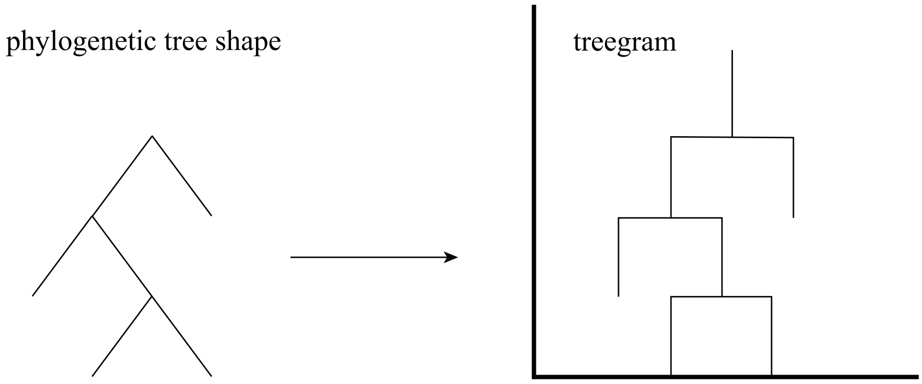

Fig. 9: Transforming a phylogenetic tree shape into an ultra-dissimilarity space: In this figure, we illustrate the treegram corresponding to the ultra-dissimilarity space generated by Equation (26) with respect to the phylogenetic tree shape on the left. Note that the treegram preserves the tree structure and the smallest birth time of points is exactly 0.

downloaded from NCBI on 22 January 2016). More precisely, we investigate the relation between two samples of size 200 of phylogenetic tree shapes with 500 tips. Phylogenetic trees from the first sample are based on a random subsample of size 500 of 2168 HA-sequences that were collected in the USA between March 2010 and September 2015, while trees from the second sample are based on a random subsample of size 500 of 1388 HA-sequences gathered in the tropics between January 2000 and October 2015 (for the exact construction of the trees see [25]). Although both samples of phylogenetic trees are based on HA protein sequences from human influenza A, we expect them to be quite different. On the one hand, influenza A is highly seasonal outside the tropics (where this seasonal variation is absent) with the majority of cases occurring in the winter [86]. On the other hand, it is well known that the undergoing evolution of the HA protein causes a 'ladder-like' shape of long-term influenza phylogenetic trees [51, 62, 101, 103] that is typically less developed in short term data sets. Thus, also the different collection period of the two data sets will most likely influence the respective phylogenetic tree shapes.

In order to compare the phylogenetic tree shapes of the resulting 400 trees, we have to transform the phylogenetic tree shapes into ultra-dissimilarity measure spaces $\mathcal{X}_i = (X_i, u_{X_i}, \mu_{X_i})$, $1 \le i \le 400$. To this end, we discard all the lables, denote by X_i the tips of the i'th phylogenetic tree and refer to the corresponding tree shape as $\mathcal{T}_i$. Next, we define the ultra-dissimilarities u_{X_i} on X_i, $1 \le i \le 400$. For this purpose, we set all edge length in the considered phylogenetic trees to one and construct u_{X_i} as follows: let $x_1^i, x_2^i \in X_i$ and let $a_{1,2}^i$ be the most recent common ancestor of x_1^i and x_2^i. Let $d_{a_{1,2}}^i$ be the length of the shortest path from $a_{1,2}^i$ to the root, let d_1^i be the length of the shortest path from x_1^i to the root and let d^i be the length of the longest shortest path from any tip to the root. Then, we define for any $x_1^i, x_2^i \in X_i$

$$u_{X_i}(x_1^i, x_2^i) = \begin{cases} d^i - d^i_{a_{1,2}} & \text{if } x_1^i \neq x_2^i \\ d^i - d^i_1 & \text{if } x_1^i = x_2^i, \end{cases} \tag{26}$$

and weight all tips in X_i equally (i.e. μ_{X_i} is the uniform measure on X_i). This naturally transforms the collection of phylogenetic tree shapes $\mathcal{T}_i$, $1 \leq i \leq 400$, into a collection of ultra-dissimilarity spaces (see Figure 9 for an illustration), which allows us to directly apply $\mathbf{SLB}_1^{\text{ult}}$ to compare them (once again we exemplarily choose $p = 1$).

In Figure 10 we contrast our findings for the comparisons of the shapes $\mathcal{T}_i$, $1 \leq i \leq 400$, to those obtained by computing the metric $d_{\text{CP},2}$ described in [25]. The top row of Figure 10 visualizes the dissimilarity matrix for the comparisons of all 400 phylogenetic tree shapes (the first 200 entries correspond to the tree shapes from the US-influenza and the second 200 correspond to the ones from the tropic influenza) obtained by applying $\mathbf{SLB}_1^{\text{ult}}$ as heat map (left) and as multidimensional scaling plot (right). The heat map shows that the collection of US trees is divided into a large group $\mathcal{G}_1 := (\mathcal{T}_i)_{1\leq i\leq 161}$, that is well separated from the phylogenetic tree shapes based on tropical data $\mathcal{G}_3 := (\mathcal{T}_i)_{201\leq i\leq 400}$, and a smaller subgroup $\mathcal{G}_2 := (\mathcal{T}_i)_{162\leq i\leq 200}$, that seems to be more similar (in the sense of $\mathbf{SLB}_1^{\text{ult}}$) to the tropical phylogenetic tree shapes. In the following $\mathcal{G}_1$ and $\mathcal{G}_2$ are referred to as *US main* and *US secondary group*, respectively. This division is even more evident in the MDS-plot on the right (black points represent trees shapes from the US main group, blue points trees shapes from the US secondary group and red points trees shapes based on the tropical data).

> We remark that in order to highlight the subgroups the US tree shapes have been reordered according to the output permutation of a single linkage dendrogram (w.r.t. $\mathbf{SLB}_1^{\text{ult}}$) based on the US tree submatrix created by MATLAB [64] and that the tropical tree shapes have been reordered analogously.

The second row of Figure 10 displays the analogous plots for $d_{\text{CP},2}$. It is noteworthy, that the coloring in the MDS-plot of the left is the same, i.e., $T_1 \in \mathcal{G}_1$ is represented by a black point, $T_2 \in \mathcal{G}_2$ by a blue one and $T_3 \in \mathcal{G}_3$ by a red one. Interestingly, the analysis based on these plots differs from the previous one. Using $d_{\text{CP},2}$ to compare the phylogenetic tree shapes at hand, we can split the data into two clusters, where one corresponds to the US data and the other one to the tropical data, with only a small overlap (see the MDS-plot in the second row of Figure 10 on the right). In particular, we notice that $d_{\text{CP},2}$ does not clearly distinguish between the US groups $\mathcal{G}_1$ and $\mathcal{G}_2$.

In order to analyze the different findings of $\mathbf{SLB}_1^{\text{ult}}$ and $d_{\text{CP},2}$, we collect and compare different characteristics of the tree shapes in the groups $\mathcal{G}_i$, $1 \leq i \leq 3$. More precisely, we concentrate on various "metric" properties of the considered ultra-dissimilarity spaces like

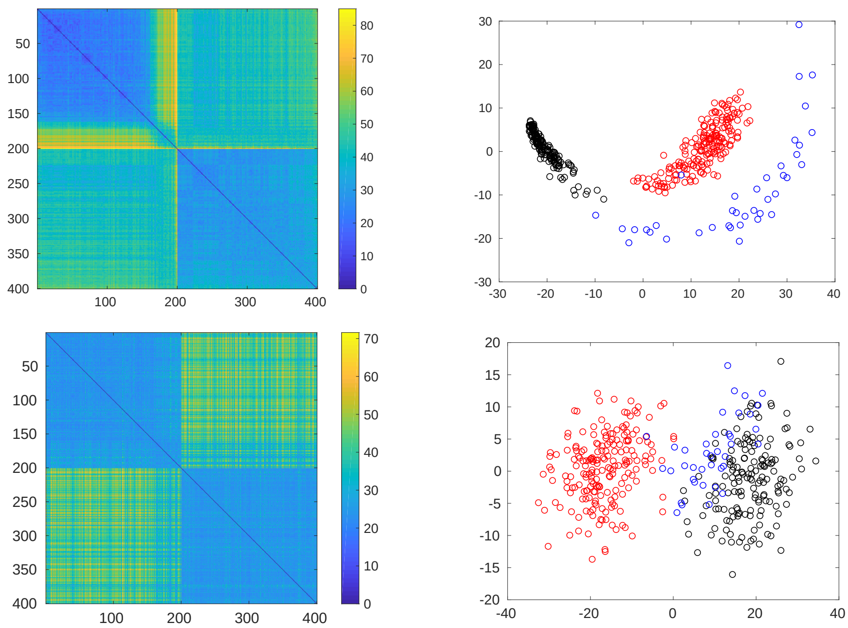

Fig. 10: Phylogenetic tree shape comparison: Visualization of the dissimilarity matrices for the comparison of the phylogenetic tree shapes $\mathcal{T}_i$, $1 \leq i \leq 400$, based on $\mathbf{SLB}_1^{\mathrm{ult}}$ (top row) and $d_{\mathrm{CP},2}$ (bottom row) as heat maps (left) and MDS-plots (right).

$\frac{1}{500^2|\mathcal{G}_i|}\sum_{\mathcal{T}_i\in\mathcal{G}_i}\sum_{x,x'\in X_i} u_{X_i}(x,x')$ ("mean average distance") or $\frac{1}{|\mathcal{G}_i|}\sum_{\mathcal{T}_i\in\mathcal{G}_i}\max\{u_{X_i}(x,x')|\, x,x'\in X_i\}$ ("mean maximal distance"), $1 \leq i \leq 3$, (these influence $\mathbf{SLB}_1^{\mathrm{ult}}$ strongly) as well as the mean numbers of certain connectivity structures, like the 4- and 5-structures (these influence $d_{\mathrm{CP},2}$, for a formal definition see [25]). Theses values (see Table 1) show that the mean average distance and the mean maximal distance differ drastically between the two groups of the US tree shapes. The tree shapes in these two groups are completely different from a metric perspective and the values for the secondary US group strongly resemble those of the tropic tree shapes. On the other hand, the connectivity characteristics do not change too much between the US main and secondary group. Hence, the metric $d_{\mathrm{CP},2}$ does not clearly divide the US trees into two groups, although the differences are certainly present. When carefully checking the phylogenetic trees, the reasons for the differences between trees in the US main group and US secondary group are not immediately apparent. Nevertheless, it is remarkable that trees from the secondary US cluster generally contain more samples from California and Florida (on average 1.92 and 0.88 more) and less from Maryland, Kentucky and Washington (on average 0.73, 0.83 and 0.72 less).

To conclude this section, we remark that using $\mathbf{SLB}_1$ instead of $\mathbf{SLB}_1^{\mathrm{ult}}$ for comparing the ultra-dissimilarity spaces $\mathcal{X}_i$, $1 \leq i \leq 400$, gives comparable results (cf. Figure 11, coloring and ordering as previously). Nevertheless, we observe (as we already have in Section 6) that

Tab. 1: Tree shape characteristics: The means of several metric and connectivity characteristics of the ultra-dissimilarity spaces $\mathcal{X}_i$ and the corresponding phylogenetic tree shapes $\mathcal{T}_i$, $1 \leq i \leq 400$, for the three groups $\mathcal{G}_i$, $1 \leq i \leq 3$.

	USA (main group)	USA (secondary group)	Tropics
Mean Avg. Dist.	36.16	61.88	53.45
Mean Max. Dist.	56.12	86.13	94.26
Mean Num. of 4-Struc.	15.61	14.08	7.81
Mean Num. of 5-Struc.	28.04	27.97	35.82

$\mathbf{SLB}_1^{\mathrm{ult}}$ is more discriminating than $\mathbf{SLB}_1$. Furthermore, we mention that so far we have only considered unweighted phylogenetic tree shapes. However, the branch lengths of the considered phylogenetic trees are relevant in many examples, because they can for instance reflect the (inferred) genetic distance between evolutionary events [25]. While the branch lengths cannot easily be included in the metric $d_{\mathrm{CP},2}$, the modeling of phylogenetic tree shapes as ultra-dissimilarity spaces is extremely flexible. It is straightforward to include branch lengths into the comparisons or to put emphasis on specific features (via weights on the corresponding tips). However, this is beyond the scope of this illustrative data analysis.

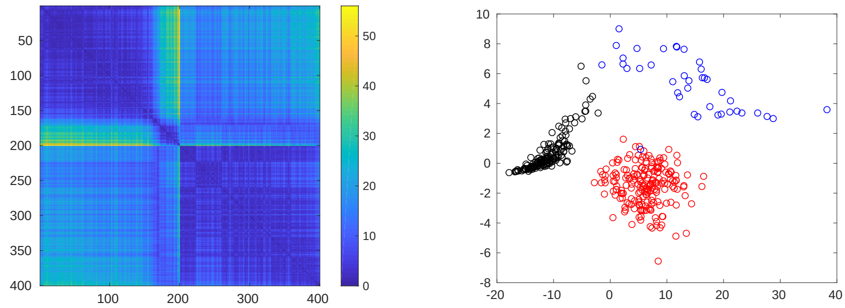

Fig. 11: Phylogenetic tree shape comparison based on $\mathbf{SLB}_1$: Representation of the dissimilarity matrices for the comparisons of the ultra-dissimilarity spaces $\mathcal{X}_i$, $1 \leq i \leq 400$, based on $\mathbf{SLB}_1$ as heat maps (left) and MDS-plots (right).

8 Concluding remarks

Since we suspect that computing $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ for finite p leads to NP-hard problems, it seems interesting to identify suitable collections of ultrametric measure spaces where these distances can be computed in polynomial time as done for the Gromov-Hausdorff distance in [70].

Acknowledgements

F.M. and Z.W. acknowledge funding from the National Science Foundation under grants CCF 1740761, DMS 1723003, and RI 1901360. A.M. and C.W. gratefully acknowledge support by the DFG Research Training Group 2088 and Cluster of Excellence MBExC 2067. F.M. and A.M. thank the Mathematisches Forschungsinstitut Oberwolfach. Conversations which eventually led to this project were initiated during the 2019 workshop "Statistical and Computational Aspects of Learning with Complex Structure".

Conflict of interest

The authors declare that they have no conflict of interest.

Data Availability statement

The code and datasets generated during and/or analysed during the current study are available in https://github.com/ndag/uGW and from http://dx.doi.org/10.5061/dryad.3r8v1.

References

[1] G. M. Adelson-Velskii and A. S. Kronrod. About level sets of continuous functions with partial derivatives. In *Dokl. Akad. Nauk SSSR*, volume 49, pages 239–241, 1945.

[2] P. K. Agarwal, K. Fox, A. Nath, A. Sidiropoulos, and Y. Wang. Computing the Gromov-Hausdorff distance for metric trees. *ACM Transactions on Algorithms (TALG)*, 14(2):1–20, 2018.

[3] A. V. Aho and J. E. Hopcroft. *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[4] D. Alvarez-Melis and T. S. Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018.

[5] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193. IEEE, 1996.

[6] L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.

[7] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

[8] M. G. Blum and O. François. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, 55(4):685–691, 2006.

[9] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[10] L. Bottou, M. Arjovsky, D. Lopez-Paz, and M. Oquab. Geometrical insights for implicit generative modeling. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, pages 229–268. Springer, 2018.
[11] D. Brinkman and P. J. Olver. Invariant histograms. *The American Mathematical Monthly*, 119(1): 4–24, 2012.
[12] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Efficient computation of isometry-invariant distances between surfaces. *SIAM Journal on Scientific Computing*, 28(5):1812–1836, 2006.
[13] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172, 2006.
[14] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel. Partial similarity of objects, or how to compare a centaur to a horse. *International Journal of Computer Vision*, 84(2):163, 2009.
[15] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Topology-invariant similarity of nonrigid shapes. *International journal of computer vision*, 81(3):281, 2009.
[16] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, and G. Sapiro. A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *International Journal of Computer Vision*, 89(2-3):266–286, 2010.
[17] P. Brown, W. Pullan, Y. Yang, and Y. Zhou. Fast and accurate non-sequential protein structure alignment using a new asymmetric linear sum assignment heuristic. *Bioinformatics*, 32(3):370–377, 2016.
[18] C. Bunne, D. Alvarez-Melis, A. Krause, and S. Jegelka. Learning generative models across incomparable spaces. *arXiv preprint arXiv:1905.05461*, 2019.
[19] G. Carlsson and F. Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of machine learning research*, 11(Apr):1425–1470, 2010.
[20] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. In *Proceedings of the Symposium on Geometry Processing*, pages 1393–1403, 2009.
[21] J. Chen and I. Safro. Algebraic distance on graphs. *SIAM Journal on Scientific Computing*, 33(6): 3468–3490, 2011.
[22] S. Chowdhury. *Metric and Topological Approaches to Network Data Analysis*. PhD thesis, The Ohio State University, 2019.
[23] S. Chowdhury and F. Mémoli. The Gromov-Wasserstein distance between networks and stable network invariants. *Information and Inference: A Journal of the IMA*, 8(4):757–787, 2019.
[24] S. Chowdhury and T. Needham. Generalized spectral clustering via Gromov-Wasserstein learning. *arXiv preprint arXiv:2006.04163*, 2020.
[25] C. Colijn and G. Plazzotta. A metric on phylogenetic tree shapes. *Systematic biology*, 67(1):113–126, 2018.
[26] G. David, S. W. Semmes, S. Semmes, and G. R. P. Pierre. *Fractured fractals and broken dreams: Self-similar geometry through metric and measure*, volume 7. Oxford University Press, 1997.
[27] A. Dayarian and B. I. Shraiman. How to infer relative fitness from a sample of genomic sequences. *Genetics*, 197(3):913–923, 2014.
[28] K. Do Ba, H. L. Nguyen, H. N. Nguyen, and R. Rubinfeld. Sublinear time algorithms for Earth Mover's distance. *Theory of Computing Systems*, 48(2):428–442, 2011.
[29] Y. Dong and W. Sawin. COPT: Coordinated optimal transport on graphs. *arXiv preprint arXiv:2003.03892*, 2020.
[30] D. Dordovskyi, O. Dovgoshey, and E. Petrov. Diameter and diametrical pairs of points in ultrametric spaces. *P-Adic Numbers, Ultrametric Analysis, and Applications*, 3(4):253–262, 2011.
[31] R. M. Dudley. *Real analysis and probability*. CRC Press, 2018.
[32] D. A. Edwards. The structure of superspace. In *Studies in topology*, pages 121–133. Elsevier, 1975.
[33] S. N. Evans. *Probability and Real Trees: École D'Été de Probabilités de Saint-Flour XXXV-2005*.

Springer, 2007.
[34] S. N. Evans and F. A. Matsen. The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3): 569–592, 2012.
[35] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485–497, 2004.
[36] G. B. Folland. *Real analysis: Modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.
[37] M. Gellert, M. F. Hossain, F. J. F. Berens, L. W. Bruhn, C. Urbainsky, V. Liebscher, and C. H. Lillig. Substrate specificity of thioredoxins and glutaredoxins–towards a functional classification. *Heliyon*, 5 (12):e02943, 2019.
[38] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, 31(2):231–240, 1984. doi: 10.1307/mmj/1029003026. URL https://doi.org/10.1307/mmj/1029003026.
[39] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). *Probability Theory and Related Fields*, 145(1-2):285–322, 2009.
[40] G. Grindstaff and M. Owen. Geometric comparison of phylogenetic trees with different leaf sets. *arXiv preprint arXiv:1807.04235*, 2018.
[41] M. Gromov. Groups of polynomial growth and expanding maps (with an appendix by Jacques Tits). *Publications Mathématiques de l'IHÉS*, 53:53–78, 1981.
[42] J. Hein. Reconstructing evolution of sequences subject to recombination using parsimony. *Mathematical biosciences*, 98(2):185–200, 1990.
[43] L. Holm and C. Sander. Protein structure comparison by alignment of distance matrices. *Journal of molecular biology*, 233(1):123–138, 1993.
[44] N. R. Howes. *Modern analysis and topology.* Springer Science & Business Media, 2012.
[45] A. K. Jain and C. Dorai. 3D object recognition: Representation and matching. *Statistics and Computing*, 10(2):167–182, 2000.
[46] N. Jardine and R. Sibson. *Mathematical taxonomy.* John Wiley & Sons, 1971.
[47] L. V. Kantorovich. On the translocation of masses, cr (dokl.) acad. *Sci. URSS (NS)*, 37:199, 1942.
[48] L. V. Kantorovich and G. Rubinstein. On a space of completely additive functions (russ.). *Vestnik Leningrad Univ*, 13:52–59, 1958.
[49] J. Kim, N. A. Rosenberg, and J. A. Palacios. A metric space of ranked tree shapes and ranked genealogies. *bioRxiv*, 2019.
[50] B. R. Kloeckner. A geometric study of Wasserstein spaces: Ultrametrics. *Mathematika*, 61(1):162–178, 2015.
[51] K. Koelle, P. Khatri, M. Kamradt, and T. B. Kepler. A two-tiered model for simulating the ecological and evolutionary dynamics of rapidly evolving viruses, with an application to influenza. *Journal of The Royal Society Interface*, 7(50):1257–1274, 2010.
[52] A. N. Kolmogorov and S. V. Fomin. *Elements of the theory of functions and functional analysis*, volume 1. Courier Corporation, 1957.
[53] S. Kolouri, K. Nadjahi, U. Simsekli, R. Badeau, and G. Rohde. Generalized sliced Wasserstein distances. In *Advances in Neural Information Processing Systems*, pages 261–272, 2019.
[54] I. Kufareva and R. Abagyan. Methods of protein structure comparison. In *Homology Modeling*, pages 231–257. Springer, 2011.
[55] H.-Y. Kuo, H.-R. Su, S.-H. Lai, and C.-C. Wu. 3D object detection and pose estimation from depth image for robotic bin picking. In *2014 IEEE international conference on automation science and engineering (CASE)*, pages 1264–1269. IEEE, 2014.
[56] M. Lafond, N. El-Mabrouk, K. T. Huber, and V. Moulton. The complexity of comparing multiply-labelled trees by extending phylogenetic-tree metrics. *Theoretical Computer Science*, 760:15–34, 2019.
[57] T. Le, N. Ho, and M. Yamada. Fast tree variants of Gromov-Wasserstein. *arXiv preprint*

arXiv:1910.04462, 2019.
[58] T. Le, M. Yamada, K. Fukumizu, and M. Cuturi. Tree-sliced variants of Wasserstein distances. In *Advances in neural information processing systems*, pages 12304–12315, 2019.
[59] V. Liebscher. New Gromov-inspired metrics on phylogenetic tree space. *Bulletin of mathematical biology*, 80(3):493–518, 2018.
[60] P. Liu, M. Gould, and C. Colijn. Polynomial phylogenetic analysis of tree shapes. *BioRxiv*, 2020.
[61] D. G. Lowe. Local feature view clustering for 3D object recognition. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.
[62] M. Łuksza and M. Lässig. A predictive fitness model for influenza. *Nature*, 507(7490):57–61, 2014.
[63] C. L. Mallows. A note on asymptotic joint normality. *The Annals of Mathematical Statistics*, pages 508–515, 1972.
[64] MATLAB. *MATLAB: Accelerating the pace of engineering and science*. The MathWorks, Inc., 2019. URL https://www.mathworks.com.
[65] A. McGregor and D. Stubbs. Sketching Earth-Mover distance on graph metrics. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 274–286. Springer, 2013.
[66] F. Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. ISBN 978-3-905673-51-7. doi: 10.2312/SPBG/SPBG07/081-090.
[67] F. Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
[68] F. Mémoli and T. Needham. Distance distributions and inverse problems for metric measure spaces. *arXiv preprint arXiv:1810.09646*, 2021.
[69] F. Mémoli and G. Sapiro. Comparing point clouds. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 32–40, 2004.
[70] F. Mémoli and Z. Wan. On p-metric spaces and the p-Gromov-Hausdorff distance. *arXiv preprint arXiv:1912.00564v4*, 2019.
[71] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
[72] A. O. Mooers and S. B. Heard. Inferring evolutionary process from phylogenetic tree shape. *The quarterly review of Biology*, 72(1):31–54, 1997.
[73] A. A. Morozov. Extension of Colijn-Plazotta tree shape distance metric to unrooted trees. *BioRxiv*, page 506022, 2018.
[74] D. Morozov, K. Beketayev, and G. Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49(22-45):52, 2013.
[75] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.
[76] M. Owen and J. S. Provan. A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):2–13, 2010.
[77] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka. Rigid 3D geometry matching for grasping of known objects in cluttered scenes. *The International Journal of Robotics Research*, 31 (4):538–553, 2012.
[78] P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global optimization*, 1(1):15–22, 1991.
[79] G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672, 2016.
[80] D. Qiu. Geometry of non-archimedean Gromov-Hausdorff distance. *P-Adic Numbers, Ultrametric Analysis, and Applications*, 1(4):317, 2009.
[81] R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Reviews of Modern*

Physics, 58:765–788, 1986.

[82] G. Reeb. Sur les points singuliers d'une forme de pfaff completement integrable ou d'une fonction numerique [on the singular points of a completely integrable pfaff form or of a numerical function]. *Comptes Rendus Acad. Sciences Paris*, 222:847–849, 1946.

[83] D. F. Robinson. Comparison of labeled trees with valency three. *Journal of combinatorial theory, Series B*, 11(2):105–119, 1971.

[84] D. F. Robinson and L. R. Foulds. Comparison of phylogenetic trees. *Mathematical biosciences*, 53 (1-2):131–147, 1981.

[85] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[86] C. A. Russell, T. C. Jones, I. G. Barr, N. J. Cox, R. J. Garten, V. Gregory, I. D. Gust, A. W. Hampson, A. J. Hay, A. C. Hurt, et al. The global circulation of seasonal influenza a (H3N2) viruses. *Science*, 320(5874):340–346, 2008.

[87] M. Scetbon, G. Peyré, and M. Cuturi. Linear-time Gromov-Wasserstein distances using low rank couplings and costs. *arXiv preprint arXiv:2106.01128*, 2021.

[88] F. Schmiedl. Computational aspects of the Gromov-Hausdorff distance and its application in non-rigid shape matching. *Discret. Comput. Geom.*, 57(4):854–880, 2017. doi: 10.1007/s00454-017-9889-4. URL https://doi.org/10.1007/s00454-017-9889-4.

[89] S. Semmes. An introduction to the geometry of ultrametric spaces. *arXiv preprint arXiv:0711.0709*, 2007.

[90] C. Semple, M. Steel, et al. *Phylogenetics*, volume 24. Oxford University Press on Demand, 2003.

[91] Z. Smith, S. Chowdhury, and F. Mémoli. Hierarchical representations of network data with optimal distortion bounds. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1834–1838. IEEE, 2016.

[92] K.-T. Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.

[93] K.-T. Sturm. The space of spaces: Curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012.

[94] D. Thorsley and E. Klavins. Model reduction of stochastic processes using Wasserstein pseudometrics. In *2008 American Control Conference*, pages 1374–1381. IEEE, 2008.

[95] V. Titouan, N. Courty, R. Tavenard, and R. Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pages 6275–6284, 2019.

[96] E. F. Touli and Y. Wang. FPT-algorithms for computing Gromov-Hausdorff and interleaving distances between trees. *arXiv preprint arXiv:1811.02425*, 2018.

[97] S. S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974.

[98] T. Vayer, R. Flamary, R. Tavenard, L. Chapel, and N. Courty. Sliced Gromov-Wasserstein. *arXiv preprint arXiv:1905.10124*, 2019.

[99] C. Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003.

[100] C. Villani. *Optimal transport: Old and new*, volume 338. Springer Science & Business Media, 2008.

[101] E. M. Volz, K. Koelle, and T. Bedford. Viral phylodynamics. *PLoS Comput Biol*, 9(3):e1002947, 2013.

[102] Z. Wan. A novel construction of Urysohn universal ultrametric space via the Gromov-Hausdorff ultrametric. *arXiv preprint arXiv:2007.08105*, 2020.

[103] K. B. Westgeest, M. de Graaf, M. Fourment, T. M. Bestebroer, R. van Beek, M. I. Spronken, J. C. de Jong, G. F. Rimmelzwaan, C. A. Russell, A. D. Osterhaus, et al. Genetic evolution of the neuraminidase of influenza a (H3N2) viruses from 1968 to 2009 and its correspondence to haemagglutinin evolution. *The Journal of general virology*, 93(Pt 9):1996, 2012.

[104] T. Wu and K. P. Choi. On joint subtree distributions under two evolutionary models. *Theoretical population biology*, 108:13–23, 2016.

[105] I. Zarichnyi. Gromov-Hausdorff ultrametric. *arXiv preprint math/0511437*, 2005.

A Missing details from Section 2

A.1 Proofs from Section 2

In this section we give the proofs of various results form Section 2.

A.1.1 Proof of Theorem 2.2

Recall that for a given $\theta \in \mathcal{D}(X)$, we define $u_\theta : X \times X \to \mathbb{R}_{\geq 0}$ as follows

$$u_\theta(x, x') := \inf\{t \geq 0 |\, x \text{ and } x' \text{ belong to the same block of } \theta(t)\}.$$

It is straightforward to verify that u_θ is an ultrametric. For any Cauchy sequence $\{x_n\}_{n\in\mathbb{N}}$ in (X, u_θ), let $D_i := \sup_{m,n\geq i} u_\theta(x_m, x_n)$ for each $i \in \mathbb{N}$. Then, each $D_i < \infty$ and $\lim_{i\to\infty} D_i = 0$. By definition of u_θ, we have that for each $i \in \mathbb{N}$ the set $\{x_n\}_{n=i}^{\infty}$ is contained in the block $[x_i]_{D_i} \in \theta(D_i)$. Let $X_i := [x_i]_{D_i}$ for each $i \in \mathbb{N}$. Then, obviously we have that $X_j \subseteq X_i$ for any $1 \leq i < j$. By condition (7) in Definition 2.1, we have that $\bigcap_{i\in\mathbb{N}} X_i \neq \emptyset$. Choose $x_* \in \bigcap_{i\in\mathbb{N}} X_i$, then it is easy to verify that $x_* = \lim_{n\to\infty} x_n$ and thus (X, u_θ) is a complete space. To prove that (X, u_θ) is a compact space, we need to verify that for each $t > 0$, X_t is a finite space (cf. Lemma A.7). Since $\theta(t)$ is finite by condition (6) in Definition 2.1, we have that $X_t = \{[x]_t |\, x \in X\} = \theta(t)$ is finite and thus X is compact. Therefore, we have proved that $u_\theta \in \mathcal{U}(X)$. Based on this, the map $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ defined by $\theta \mapsto u_\theta$ is well-defined.

Now given $u \in \mathcal{U}(X)$, we define a map $\theta_u : [0, \infty) \to \mathbf{Part}(X)$ as follows: for each $t \geq 0$, consider the equivalence relation $\sim_t$ with respect to u, i.e., $x \sim_t x'$ if and only if $u(x, x') \leq t$. This is actually the same equivalence relation defined in Section 2.2 for introducing quotient ultrametric spaces. We then let $\theta_u(t)$ to be the partition induced by $\sim_t$, i.e., $\theta_u(t) = X_t$. It is not hard to show that θ_u satisfies conditions (1)–(5) in Definition 2.1. Since X is compact, then $\theta_u(t) = X_t$ is finite for each $t > 0$ and thus θ_u satisfies condition (6) in Definition 2.1. Now, let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence such that $\lim_{n\to\infty} t_n = 0$ and let $X_n \in \theta_X(t_n)$ such that for any $1 \leq n < m$, $X_m \subseteq X_n$. Since each $X_n = [x_n]_{t_n}$ for some $x_n \in X$, X_n is a compact subset of X. Since X is also complete, we have that $\bigcap_{n\in\mathbb{N}} X_n \neq \emptyset$. Therefore, θ_u satisfies condition (7) in Definition 2.1 and thus $\theta_u \in \mathcal{D}(X)$. Then, we define the map $\Delta_X : \mathcal{U}(X) \to \mathcal{D}(X)$ by $u \mapsto \theta_u$.

It is easy to check that Δ_X is the inverse of Υ_X and thus we have established that $\Upsilon_X : \mathcal{D}(X) \to \mathcal{U}(X)$ is bijective.

A.1.2 Proof of Lemma 2.8

First of all, we show that the right hand side of Equation (16) is well defined. More precisely, we employ Lemma A.7 to prove that the supremum

$$\sup_{B\in V(X)\backslash\{X\} \text{ and } \alpha(B)\neq\beta(B)} \operatorname{diam}(B^*)$$

is attained. For arbitrary $B_0 \in V(X)\backslash\{X\}$ such that $\alpha(B_0) \neq \beta(B_0)$, we have that $\operatorname{diam}(B_0^*) > 0$. By Lemma A.7 the spaces X_t are finite for $t > 0$. Since $V(X) = \{[x]_t |\, x \in X, t > 0\} = \bigcup_{t>0} X_t$, there are only finitely many $B \in V(X)\backslash\{X\}$ such that $\operatorname{diam}(B) \geq \operatorname{diam}(B_0^*)$ and thus $\operatorname{diam}(B^*) \geq \operatorname{diam}(B_0^*)$. This implies that the supremum is attained and thus

$$\sup_{B\in V(X)\backslash\{X\} \text{ and } \alpha(B)\neq\beta(B)} \operatorname{diam}(B^*) = \max_{B\in V(X)\backslash\{X\} \text{ and } \alpha(B)\neq\beta(B)} \operatorname{diam}(B^*). \tag{27}$$

Let B_1 denote the maximizer in Equation (27) and let $\delta := \operatorname{diam}(B_1^*)$. It is easy to see that for any $x \in X$, $\alpha([x]_\delta) = \beta([x]_\delta)$.

By Strassen's theorem (see for example [31, Theorem 11.6.2]),

$$d_{\mathrm{W},\infty}(\alpha, \beta) = \inf\{r \geq 0 |\, \text{for any closed subset } A \subseteq X,\, \alpha(A) \leq \beta(A^r)\}, \tag{28}$$

where $A^r := \{x \in X |\, u_X(x, A) \leq r\}$.

Since $\alpha(B_1) \neq \beta(B_1)$, we assume without loss of generality that $\alpha(B_1) > \beta(B_1)$. By definition of B_1^*, it is obvious that $(B_1)^\delta = B_1^*$ (recall: $\delta := \operatorname{diam}(B_1^*)$) and $(B_1)^r = B_1$ for all $0 \leq r < \delta$. Therefore, $\alpha(B_1) \leq \beta((B_1)^r)$ only when $r \geq \delta$. By Equation (28), this implies that $d_{\mathrm{W},\infty}(\alpha, \beta) \geq \delta$. Conversely, for any closed set A, we have that $A^\delta = \bigcup_{x\in A}[x]_\delta$. For two closed balls in ultrametric spaces, either one includes the other or they have no intersection. Therefore, there exists a subset $S \subseteq A$ such that $[x]_\delta \cap [x']_\delta = \emptyset$ for all $x, x' \in S$ and $x \neq x'$, and that $A^\delta = \bigsqcup_{x\in S}[x]_\delta$. Then, $\alpha(A) \leq \alpha(A^\delta) = \sum_{x\in S} \alpha([x]_\delta) = \sum_{x\in S} \beta([x]_\delta) = \beta(A^\delta)$. Hence, $d_{\mathrm{W},\infty}(\alpha, \beta) \leq \delta$ and thus

$$d_{\mathrm{W},\infty}(\alpha, \beta) = \max_{B\in V(X)\backslash\{X\} \text{ and } \alpha(B)\neq\beta(B)} \operatorname{diam}(B^*).$$

A.2 Technical issues from Section 2

In the following, we address various technical issues from Section 2.

A.2.1 Synchronized rooted trees

A *synchronized rooted tree*, is a combinatorial tree $T = (V, E)$ with a root $o \in V$ and a height function $h : V \to [0, \infty)$ such that $h^{-1}(0)$ coincides with the leaf set and $h(v) < h(v^*)$ for each $v \in V \backslash \{o\}$, where v^* is the parent of v. Similar as in Theorem 2.2 that there exists a correspondence between ultrametric spaces and dendrograms, an ultrametric space X uniquely determines a synchronized rooted tree T_X [50].

Now given a compact ultrametric space (X, u_X), we construct the corresponding sychronized rooted tree T_X via the dendrogram θ_X associated with u_X. Recall from Section 2.3 that $V(X) := \bigcup_{t>0} \theta_X(t)$. For each $B \in V(X)\backslash\{X\}$, denote by B^* the smallest element in $V(X)$ such that $B \subsetneqq B^*$, whose existence is guaranteed by the following lemma:

Lemma A.1. *Let X be a compact ultrametric space and let $V(X) = \bigcup_{t>0} \theta_X(t)$, where θ_X is as defined in Remark 2.3. For each $B \in V(X)$ such that $B \neq X$, there exists $B^* \in V(X)$ such that $B^* \neq B$ and $B^* \subseteq B'$ for all $B' \in V(X)$ with $B \subsetneqq B'$.*

Proof. Let $\delta := \mathrm{diam}\,(B)$. Let $x \in B$, then $B = [x]_\delta$. By Lemma A.7, X_δ is a finite set. Consider $\delta^* := \min\{u_{X_\delta}([x]_\delta, [x']_\delta) |\, [x']_\delta \neq [x]_\delta\}$. Let $B^* := [x]_{\delta^*}$, then B^* is the smallest element in $V(X)$ containing B under inclusion. Indeed, $B^* \neq B$ and if $B \subseteq B'$ for some $B' \in V(X)$, then $B' = [x]_r$ for some $r > \delta$. It is easy to see that for all $\delta < r < \delta^*$, $[x]_r = [x]_\delta$. Therefore, if $B' \neq B$, we must have that $r \geq \delta^*$ and thus $B^* = [x]_{\delta^*} \subseteq [x]_r = B'$. $\square$

Now, we define a combinatorial tree $T_X = (V_X, E_X)$ as follows: we let $V_X := V(X)$; for any distinct $B, B' \in V_X$, we let $(B, B') \in E_X$ iff either $B = (B')^*$ or $B' = B^*$. We choose $X \in V_X$ to be the root of T_X, then any $B \neq X$ in V_X has a unique parent B^*. We define $h_X : V_X \to [0, \infty)$ such that $h_X(B) := \frac{\mathrm{diam}(B)}{2}$ for any $B \in V_X$. Now, T_X endowed with the root X and the height function h_X is a synchronized rooted tree. It is easy to see that X can be isometrically identified with $h_X^{-1}(0)$ of the so-called *metric completion* of T_X (see [50, Section 2.3] for details). With this construction Theorem 2.7 follows directly from [50, Lemma 3.1].

A.3 $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ between compactly supported measures

Next, we demonstrate that Theorem 2.9 extends naturally to the case of compactly supported probability measures in $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$. For this purpose, it is important to note that compact subsets of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$ have a very particular structure as shown by the subsequent lemma.

Lemma A.2. *Let $X \subseteq (\mathbb{R}_{\geq 0}, \Lambda_\infty)$. X is a compact subset if and only if X is either a finite set or a countable set with 0 being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1).*

Proof. If X is finite, then obviously X is compact. Assume that X is a countable set with 0 being the unique cluster point (w.r.t. the usual Euclidean distance Λ_1). If $\{x_n\}_{n\in\mathbb{N}} \subseteq X$ is a Cauchy sequence with respect to Λ_∞, then either x_n is a constant when n is large or $\lim_{n\to\infty} x_n = 0$. In either case, the limit of $\{x_n\}_{n\in\mathbb{N}}$ belongs to X and thus X is complete. Now for any $\varepsilon > 0$, by Lemma A.7, X_ε is a finite set. Denote $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$. Then, $\{x_1, \ldots, x_n\}$ is a finite ε-net of X. Therefore, X is totally bounded and thus X is compact.

Now, assume that X is compact. Then, for any $\varepsilon > 0$, X_ε is a finite set. Suppose $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$ where $0 \leq x_1 < x_2 < \cdots < x_n$. Further, we have that $\Lambda_\infty(x_i, x_j) = x_j$ whenever $1 \leq i < j \leq n$. This implies that

1. $x_i > \varepsilon$ for all $2 \leq i \leq n$;
2. $[x_i]_\varepsilon = \{x_i\}$ for all $2 \leq i \leq n$.

Therefore, $X \cap (\varepsilon, \infty) = \{x_2, \ldots, x_n\}$ is a finite set. Since $\varepsilon > 0$ is arbitrary, X is an at most countable set and has no cluster point (w.r.t. the usual Euclidean distance Λ_1) other than 0. If X is countable, then 0 must be a cluster point and by compactness of X, we have that $0 \in X$. ☐

Based on the special structure of compact subsets of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$, we derive the following extension of Theorem 2.9.

Theorem A.3 ($d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ between compactly supported measures)**.** *Suppose α, β are supported on a countable subset $X := \{0\} \cup \{x_i | \, i \in \mathbb{N}\}$ of $\mathbb{R}_{\geq 0}$ such that $0 < \ldots < x_n < x_{n-1} < \ldots < x_1$ and 0 is the only cluster point with respect to the usual Euclidean distance. Let $\alpha_i := \alpha(\{x_i\})$ for $i \in \mathbb{N}$ and $\alpha_0 := \alpha(\{0\})$. Similarly, let $\beta_i := \beta(\{x_i\})$ and $\beta_0 := \beta(\{0\})$. Then for $p \in [1, \infty)$,*

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha, \beta) = 2^{-\frac{1}{p}} \left(\sum_{i=2}^{\infty} \left| \sum_{j=i}^{\infty} (\alpha_j - \beta_j) \right| \cdot |x_{i-1}^p - x_i^p| + \sum_{i=1}^{\infty} |\alpha_i - \beta_i| \cdot x_i^p \right)^{\frac{1}{p}}. \tag{29}$$

Let F_α and F_β denote the cumulative distribution functions of α and β, respectively. Then, we obtain

$$d_{\mathrm{W},\infty}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha, \beta) = \max \left(\max_{2 \leq i < \infty, F_\alpha(x_i) \neq F_\beta(x_i)} x_{i-1}, \max_{1 \leq i < \infty, \alpha_i \neq \beta_i} x_i \right).$$

Proof. Note that $V(X) = \{\{0\} \cup \{x_j | \, j \geq i\} | \, i \in \mathbb{N}\} \cup \{\{x_i\} | \, i \in \mathbb{N}\}$ (recall that each set corresponds to a closed ball). Thus, we conclude the proof by applying Lemma 2.7 and Lemma 2.8. $\square$

A.3.1 Closed-form solution for $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}$

In the following, we will derive the subsequent theorem.

Theorem A.4. *Given $1 \leq p, q < \infty$ and two compactly supported probability measures α and β on $\mathbb{R}_{\geq 0}$, we have that*

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta) \leq \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p dt\right)^{\frac{1}{p}}.$$

When $q \leq p$, the equality holds whereas when $q > p$, the equality does not hold in general.

One important ingredient for the proof of Theorem A.4 is Lemma 3.2 of Chowdhury and Mémoli [23] which we restate here for convenience.

Lemma A.5 (Chowdhury and Mémoli [23, Lemma 3.2])**.** *Let X, Y be two Polish metric spaces and let $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be measurable maps. Denote by $f \times g : X \times Y \to \mathbb{R}^2$ the map $(x, y) \mapsto (f(x), g(y))$. Then, for any $\mu_Y \in \mathcal{P}(X)$ and $\mu_Y \in \mathcal{P}(Y)$*

$$(f \times g)_\# \mathcal{C}(\mu_X, \mu_Y) = \mathcal{C}(f_\# \mu_Y, g_\# \mu_Y).$$

Based on Lemma A.5, we can show the following auxiliary result.

Lemma A.6. *Let $1 \leq q \leq p < \infty$. Assume that α and β are compactly supported probability measures on $\mathbb{R}_{\geq 0}$. Then,*

$$\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p = \left(d_{\mathrm{W},\frac{p}{q}}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_\#\alpha,(S_q)_\#\beta)\right)^{\frac{p}{q}},$$

where $S_q : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ taking x to x^q is the q-snowflake transform defined in Section 3.3.

Proof.

$$\begin{aligned}\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)}\int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}}(\Lambda_q(x,y))^p\mu(dx\times dy)\\ &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)}\int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}}|S_q(x)-S_q(y)|^{\frac{p}{q}}\mu(dx\times dy)\\ &= \inf_{\mu\in\mathcal{C}(\alpha,\beta)}\int_{\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}}|s-t|^{\frac{p}{q}}(S_q\times S_q)_{\#}\mu(ds\times dt)\\ &= \left(d_{\mathrm{W},\frac{p}{q}}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_{\#}\alpha,(S_q)_{\#}\beta)\right)^{\frac{p}{q}},\end{aligned}$$

where we use $\frac{p}{q}\geq 1$ and Lemma A.5 in the last equality. □

With Lemma A.6 at our disposal, we can demonstrate Theorem A.4.

Proof of Theorem A.4. We first note that $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)=\inf_{(\xi,\eta)}\left(\mathbb{E}(\Lambda_q(\xi,\eta)^p)\right)^{\frac{1}{p}}$, where ξ and η are two random variables with marginal distributions α and β, respectively. Moreover, let ζ be the random variable uniformly distributed on $[0,1]$, then $F_\alpha^{-1}(\zeta)$ has distribution function F_α and $F_\beta^{-1}(\zeta)$ has distribution function F_β (see for example Vallender [97]). Let $\xi=F_\alpha^{-1}(\zeta)$ and $\eta=F_\beta^{-1}(\zeta)$, then we have

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\leq\left(\mathbb{E}(\Lambda_q(\xi,\eta)^p)\right)^{\frac{1}{p}}=\left(\int_0^1\Lambda_q(F_\alpha^{-1}(t),F_\beta^{-1}(t))^p dt\right)^{\frac{1}{p}}.$$

Next, we assume that $q\leq p$. By Lemma A.6, we have that

$$\left(d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)\right)^p=\left(d_{\mathrm{W},\frac{p}{q}}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_{\#}\alpha,(S_q)_{\#}\beta)\right)^{\frac{p}{q}}.$$

Then,

$$\left(d_{\mathrm{W},\frac{p}{q}}^{(\mathbb{R}_{\geq 0},\Lambda_1)}((S_q)_{\#}\alpha,(S_q)_{\#}\beta)\right)^{\frac{p}{q}}=\int_0^1|F_{\alpha,q}^{-1}(t)-F_{\beta,q}^{-1}(t)|^{\frac{p}{q}}dt,$$

where $F_{\alpha,q}$ and $F_{\beta,q}$ are distribution functions of $(S_q)_{\#}\alpha$ and $(S_q)_{\#}\beta$, respectively. It is easy to verify that $F_{\alpha,q}(t)=(F_\alpha^{-1}(t))^q$ and $F_{\beta,q}(t)=(F_\beta^{-1}(t))^q$. Therefore,

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha,\beta)=\left(\int_0^1\Lambda_q(F_\alpha^{-1}(t),F_\beta^{-1}(t))^p dt\right)^{\frac{1}{p}}$$

Finally, we demonstrate that for $q>p$ the equality does not hold in general. We first consider the extreme case $p=1$ and $q=\infty$ (though we require $q<\infty$ in the assumptions of the theorem, we relax this for now). Let $\alpha_0 = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$ and $\beta_0 = \frac{1}{2}\delta_2 + \frac{1}{2}\delta_3$ where δ_x means the Dirac measure at point $x \in \mathbb{R}_{\geq 0}$. Then, we have that

$$d_{\mathrm{W},1}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}(\alpha_0,\beta_0) = \frac{3}{2} < \frac{5}{2} = \int_0^1 \Lambda_\infty(F_\alpha^{-1}(t), F_\beta^{-1}(t))dt.$$

It is not hard to see that both $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha_0,\beta_0)$ and $\left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p dt\right)^{\frac{1}{p}}$ are continuous with respect to $p \in [1,\infty)$ and $q \in [1,\infty]$. Then, for p close to 1 and $q < \infty$ large enough, and in particular, $p < q$, we have that

$$d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_q)}(\alpha_0,\beta_0) < \left(\int_0^1 \Lambda_q(F_\alpha^{-1}(t), F_\beta^{-1}(t))^p dt\right)^{\frac{1}{p}}.$$

□

A.3.2 Miscellaneous

In the remainder of this section, we collect several technical results that find implicit or explicit usage throughout Section 2.

Lemma A.7. *Let X be a complete ultrametric space. Then, X is compact ultrametric space if and only if for any $t > 0$, X_t is a finite space.*

Proof. Wan [102, Lemma 2.3] proves that whenever X is compact, X_t is finite for any $t > 0$.

Conversely, we assume that X_t is finite for any $t > 0$. We only need to prove that X is totally bounded. For any $\varepsilon > 0$, X_ε is a finite set and thus there exists $x_1, \ldots, x_n \in X$ such that $X_\varepsilon = \{[x_1]_\varepsilon, \ldots, [x_n]_\varepsilon\}$. Now, for any $x \in X$, there exists x_i for some $i = 1, \ldots, n$ such that $x \in [x_i]_\varepsilon$. This implies that $u_X(x, x_i) \leq \varepsilon$. Therefore, the set $\{x_1, \ldots, x_n\} \subseteq X$ is an ε-net of X. Then, X is totally bounded and thus compact. □

Lemma A.8. *$V(X)$ is the collection of all closed balls in X except for singletons $\{x\}$ such that x is a cluster point in X. In particular, $X \in V(X)$ and for any $x \in X$, if x is not a cluster point, then $\{x\} \in V(X)$.*

Proof. Given any $t > 0$ and $x \in X$, $[x]_t = B_t(x) = \{x' \in X \,|\, u_X(x,x') \leq t\}$. Therefore, $V(X)$ is a collection of closed balls in X. On the contrary, any closed ball $B_t(x)$ with positive radius $t > 0$ coincides with $[x]_t \in \theta_X(t)$ and thus belongs to $V(X)$. Now, for any singleton $\{x\} = B_0(x)$. If x is not a cluster point, then there exists $t > 0$ such that $B_t(x) = \{x\}$ which implies that $\{x\} \in V(X)$. If x is a cluster point, then for any $t > 0$,

$\{x\} \subsetneqq B_t(x) = [x]_t$. In particular, this implies that $\{x\} \neq [x]_t$ for all $t > 0$ and thus $\{x\} \notin V(X)$. In conclusion, $V(X)$ is the collection of all closed balls in X except for singletons $\{x\}$ such that x is a cluster point in X.

If X is a one point space, then obviously $X \in V(X) = \{X\}$. Otherwise, let $\delta := \operatorname{diam}(X) > 0$, then for any $x \in X$ we have that $X = [x]_\delta \in V(X)$. As for singletons $\{x\}$ where $x \in X$ is not a cluster point, we have proved above that $\{x\} \in V(X)$. □

B Missing details from Section 3

B.1 Proofs from Section 3.1

Next, we give the missing proofs of the results stated in Section 3.1.

B.1.1 Proof of Proposition 3.3

1. This directly follows from the definitions of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ and $d_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Equation (10) and Equation (5)).
2. This simply follows from Jensen's inequality.
3. By (2), we know that $\{u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\}_{n\in\mathbb{N}}$ is an increasing sequence with a finite upper bound $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$. Therefore, $L := \lim_{n\to\infty} u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ exists and $L \leq u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$.

   Next, we come to the opposite inequality. By Proposition B.1, there exist $u_n \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\mu_n \in \mathcal{C}(\mu_X,\mu_Y)$ such that
   $$\left(\int_{X\times Y} (u_n(x,y))^n \mu_n(dx\times dy)\right)^{\frac{1}{n}} = u_{\mathrm{GW},n}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}).$$
   By Lemma B.19 and Lemma B.21, the sequence $\{u_n\}_{n\in\mathbb{N}}$ uniformly converges to some $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges to some $\mu \in \mathcal{C}(\mu_X,\mu_Y)$ (after taking appropriate subsequences of both sequences). Let $M := \sup_{(x,y)\in\mathrm{supp}(\mu)} u(x,y)$. Let $\varepsilon > 0$ and let $U = \{(x,y) \in X\times Y \,|\, u(x,y) > M-\varepsilon\}$. Then, $\mu(U) > 0$. Since U is open, it follows that there exists a small $\varepsilon_1 > 0$ such that $\mu_n(U) > \mu(U)-\varepsilon_1 > 0$ for all n large enough (see e.g. Billingsley [7, Thm. 2.1]). Moreover, by uniform convergence of the sequence $\{u_n\}_{n\in\mathbb{N}}$, we have $|u(x,y)-u_n(x,y)| \leq \varepsilon$ for any $(x,y) \in X\times Y$ when n is large enough. Therefore, we obtain for n large enough
   $$\left(\int_{X\times Y} (u_n(x,y))^n \mu_n(dx\times dy)\right)^{\frac{1}{n}} \geq (\mu_n(U))^{\frac{1}{n}}(M-2\varepsilon) \geq (\mu(U)-\varepsilon_1)^{\frac{1}{n}}(M-2\varepsilon).$$

Letting $n \to \infty$, we obtain $L \geq M - 2\varepsilon$. Since $\varepsilon > 0$ is arbitrary, we obtain $L \geq M \geq u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y})$.

B.1.2 Proof of Theorem 3.4

In this section, we devote to prove Theorem 3.4. To this end, we will first verify the existence of optimal metrics and optimal couplings in Equation (18).

Proposition B.1 (Existence of optimal couplings). *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, there always exist an ultrametric $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that for $1 \leq p < \infty$*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \left(\int_{X \times Y} (u(x,y))^p \mu(dx \times dy) \right)^{\frac{1}{p}}$$

and such that

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \sup_{(x,y) \in \mathrm{supp}(\mu)} u(x,y).$$

Proof. The following proof is a suitable adaptation from proof of Lemma 3.3 in [92]. We will only prove the claim for the case $p < \infty$ since the case $p = \infty$ can be shown in a similar manner. Let $u_n \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ be such that

$$\left(\int_{X \times Y} (u_n(x,y))^p \mu_n(dx \times dy) \right)^{\frac{1}{p}} \leq u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) + \frac{1}{n}.$$

By Lemma B.19, $\{\mu_n\}_{n \in \mathbb{N}}$ weakly converges (after taking an appropriate subsequence) to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. By Lemma B.21, $\{u_n\}_{n \in \mathbb{N}}$ uniformly converges (after taking an appropriate subsequence) to some $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. Then, it is easy to verify that

$$\left(\int_{X \times Y} (u(x,y))^p \mu(dx \times dy) \right)^{\frac{1}{p}} \leq u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}).$$

□

As a direct consequence of the proposition, we get the subsequent result.

Corollary B.2. *Fix $1 \leq p \leq \infty$. Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, there exist a compact ultrametric space Z and isometric embeddings $\phi : X \hookrightarrow Z$ and $\psi : Y \hookrightarrow Z$ such that*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},p}^{Z}(\phi_\# \mu_X, \psi_\# \mu_Y).$$

Before we come to the proof of Theorem 3.4, it remains to establish another auxiliary result. We ensure that the Wasserstein pseudometric of order p on a compact pseudo-ultrametric space (X, u_X) is for $p \in [1, \infty)$ a p-pseudometric and for $p = \infty$ a pseudo-ultrametric, i.e., we prove for $1 \leq p < \infty$ that for all $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(X)$

$$d_{\mathrm{W},p}^{(X,u_X)}(\mu_1, \mu_3) \leq \left(\left(d_{\mathrm{W},p}^{(X,u_X)}(\mu_1, \mu_2)\right)^p + \left(d_{\mathrm{W},p}^{(X,u_X)}(\mu_2, \mu_3)\right)^p\right)^{1/p}$$

and for $p = \infty$ that for all $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(X)$

$$d_{\mathrm{W},p}^{(X,u_X)}(\mu_1, \mu_3) \leq \max\left(d_{\mathrm{W},p}^{(X,u_X)}(\mu_1, \mu_2), d_{\mathrm{W},p}^{(X,u_X)}(\mu_2, \mu_3)\right).$$

Lemma B.3. *Let (X, u_X) be a compact pseudo-ultrametric space. Then, the p-Wasserstein metric $d_{\mathrm{W},p}^{(X,u_X)}$ is a p-pseudometric on $\mathcal{P}(X)$ for $1 \leq p \leq \infty$. In particular, when $p = \infty$, it is an pseudo-ultrametric on $\mathcal{P}(X)$.*

Proof. We prove the statement by adapting the proof of the triangle inequality for the p-Wasserstein distance (see e.g., [99, Theorem 7.3]). We only prove the case when $p < \infty$ whereas the case $p = \infty$ follows by analogous arguments.

Let $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{P}(X)$, denote by μ_{12} an optimal transport plan between α_1 and α_2 and by μ_{23} an optimal transport plan between α_2 and α_3 (see [100, Theorem 4.1] for the existence of μ_{12} and μ_{23}). Furthermore, let X_i be the support of α_i, $1 \leq i \leq 3$. Then, by the Gluing Lemma [99, Lemma 7.6] there exists a measure $\mu \in \mathcal{P}(X_1 \times X_2 \times X_3)$ with marginals μ_{12} on $X_1 \times X_2$ and μ_{23} on $X_2 \times X_3$. Clearly, we obtain

$$\begin{aligned}\left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1, \alpha_3)\right)^p &\leq \int_{X_1 \times X_2 \times X_3} u_X^p(x, z)\ \mu(dx \times dy \times dz)\\ &\leq \int_{X_1 \times X_2 \times X_3} \left(u_X^p(x, y) + u_X^p(y, z)\right)\ \mu(dx \times dy \times dz).\end{aligned}$$

Here, we used that u_X is an ultrametric, i.e., in particular a p-metric [70, Proposition 1.16]. With this we obtain that

$$\begin{aligned}\left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1, \alpha_2)\right)^p \leq \int_{X_1 \times X_2} u_X^p(x, y)\ \mu_{12}(dx \times dy) + \int_{X_2 \times X_3} u_X^p(y, z)\ \mu_{23}(dy \times dz)\\ = \left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_1, \alpha_2)\right)^p + \left(d_{\mathrm{W},p}^{(X,u_X)}(\alpha_2, \alpha_3)\right)^p.\end{aligned}$$

□

With Proposition B.1 and Lemma B.3 at our disposal we are now ready to prove Theorem 3.4 which states that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is indeed a p-metric on $\mathcal{U}^w$.

Proof of Theorem 3.4. It is clear that $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ is symmetric and that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = 0$ if $\mathcal{X} \cong_w \mathcal{Y}$. Furthermore, we remark that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \geq d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ by Proposition 3.3. Since $d_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_w \mathcal{Y}$ ([93]), we have that $u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_w \mathcal{Y}$. It remains to verify the p-triangle inequality. To this end, we only prove the case when $p < \infty$ whereas the case $p = \infty$ follows by analogous arguments.

Let $\mathcal{X},\mathcal{Y},\mathcal{Z} \in \mathcal{U}^w$. Suppose $u_{XY} \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $u_{YZ} \in \mathcal{D}^{\mathrm{ult}}(u_Y,u_Z)$ are optimal metric couplings such that

$$\left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\right)^p = \left(d_{\mathrm{W},p}^{(X\sqcup Y,u_{XY})}(\mu_X,\mu_Y)\right)^p \text{ and } \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{Y},\mathcal{Z})\right)^p = \left(d_{\mathrm{W},p}^{(Y\sqcup Z,u_{YZ})}(\mu_Y,\mu_Z)\right)^p.$$

Further, define u_{XYZ} on $X \sqcup Y \sqcup Z$ as

$$u_{XYZ}(x_1,x_2) = \begin{cases} u_{XY}(x_1,x_2) & x_1,x_2 \in X \sqcup Y \\ u_{YZ}(x_1,x_2) & x_1,x_2 \in Y \sqcup Z \\ \inf\{\max(u_{XY}(x_1,y),u_{YZ}(y,x_2)) \,|\, y \in Y\} & x_1 \in X, x_2 \in Z \\ \inf\{\max(u_{XY}(x_2,y),u_{YZ}(y,x_1)) \,|\, y \in Y\} & x_1 \in Z, x_2 \in X. \end{cases}$$

Then, by Lemma 1.1 of Zarichnyi [105] u_{XYZ} is a pseudo-ultrametric on $X \sqcup Y \sqcup Z$ that coincides with u_{XY} on $X \sqcup Y$ and with u_{YZ} on $Y \sqcup Z$. With this we obtain by Lemma B.3 that

$$\begin{aligned} \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Z})\right)^p &\leq \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z,u_{XYZ})}(\mu_X,\mu_Z)\right)^p \\ &\leq \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z,u_{XYZ})}(\mu_X,\mu_Y)\right)^p + \left(d_{\mathrm{W},p}^{(X\sqcup Y\sqcup Z,u_{XYZ})}(\mu_Y,\mu_Z)\right)^p \\ &= \left(d_{\mathrm{W},p}^{(X\sqcup Y,u_{XY})}(\mu_X,\mu_Y)\right)^p + \left(d_{\mathrm{W},p}^{(Y\sqcup Z,u_{YZ})}(\mu_Y,\mu_Z)\right)^p \\ &= \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\right)^p + \left(u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{Y},\mathcal{Z})\right)^p \end{aligned}$$

This gives the claim for $p < \infty$. □

B.1.3 Proof of Theorem 3.7

In order to proof Theorem 3.7, we will first establish the statement for *finite* ultrametric measure spaces. For this purpose, we need to introduce some notation. Given $\mathcal{X},\mathcal{Y} \in \mathcal{U}^w$, let $\mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X,u_Y)$ denote the collection of all admissible pseudo-ultrametrics on $X \sqcup Y$, where $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ is called *admissible*, if there exists no $u^* \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ such that $u^* \neq u$ and $u^*(x,y) \leq u(x,y)$ for all $x,y \in X \sqcup Y$.

Lemma B.4. *For any $\mathcal{X},\mathcal{Y} \in \mathcal{U}^w$, $\mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X,u_Y) \neq \emptyset$. Moreover,*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = \inf_{u \in \mathcal{D}_{\mathrm{adm}}^{\mathrm{ult}}(u_X,u_Y)} d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X,\mu_Y).$$

Proof. If $\{u_n\}_{n\in\mathbb{N}} \subseteq \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ is a decreasing sequence (with respect to pointwise inequality), it is easy to verify that $u := \inf_{n\in\mathbb{N}} u_n \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and thus u is a lower bound of $\{u_n\}_{n\in\mathbb{N}}$. Then, by Zorn's lemma $\mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y) \neq \emptyset$. Therefore, we obtain that

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \inf_{u\in\mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X,u_Y)} d^{(X\sqcup Y,u)}_{\mathrm{W},p}(\mu_X, \mu_Y).$$

□

Combined with Example 3.6, the following result implies that each $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$ gives rise to an element in $\mathcal{A}$.

Lemma B.5. *Given* finite *spaces* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$*, for each* $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$*,* $u^{-1}(0) \neq \emptyset$*.*

Proof. Assume otherwise that $u^{-1}(0) = \emptyset$. Then, u is a metric (instead of pseudo-metric). Let $(x_0, y_0) \in X \times Y$ such that $u(x_0, y_0) = \min_{x\in X,y\in Y} u(x, y)$. The existence of (x_0, y_0) is guaranteed by the finiteness of X and Y. We define $u_{(x_0,y_0)} : X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\geq 0}$ as follows:

1. $u_{(x_0,y_0)}|_{X\times X} := u_X$ and $u_{(x_0,y_0)}|_{Y\times Y} := u_Y$;
2. For $(x, y) \in X \times Y$,
$$u_{(x_0,y_0)}(x, y) := \min\left(u(x, y), \max(u_X(x, x_0), u_Y(y, y_0))\right);$$
3. For any $(y, x) \in Y \times X$, $u_{(x_0,y_0)}(y, x) := u_{(x_0,y_0)}(x, y)$.

It is easy to verify that $u_{(x_0,y_0)} \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. Further, it is obvious that $u_{(x_0,y_0)}(x_0, y_0) = 0 < u(x_0, y_0)$ and that $u_{(x_0,y_0)}(x, y) \leq u(x, y)$ for all $x, y \in X \sqcup Y$ which contradicts with $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$. Therefore, $u^{-1}(0) \neq \emptyset$. □

Theorem B.6. *Let* $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$ *be* finite *spaces. Then, we have for each* $p \in [1, \infty)$ *that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}} d^{Z_A}_{\mathrm{W},p}\left(\left(\phi^X_{(A,\varphi)}\right)_\#\mu_X, \left(\psi^Y_{(A,\varphi)}\right)_\#\mu_Y\right). \tag{30}$$

Proof. By Lemma B.4 it is sufficent to prove that each $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$ induces $(A, \varphi) \in \mathcal{A}$ such that

$$d^{(X\sqcup Y,u)}_{\mathrm{W},p}(\mu_X, \mu_Y) \geq d^{Z_A}_{\mathrm{W},p}\left(\left(\phi^X_{(A,\varphi)}\right)_\#\mu_X, \left(\psi^Y_{(A,\varphi)}\right)_\#\mu_Y\right).$$

Let $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X, u_Y)$. We define $A_0 := \{x \in X | \exists y \in Y \text{ such that } u(x, y) = 0\}$ ($A_0 \neq \emptyset$ by Lemma B.5). By Example 3.6, the map $\varphi_0 : A_0 \to Y$ defined by taking x to y such that $u(x, y) = 0$ is a well-defined isometric embedding. This means in particular that $(A_0, \varphi_0) \in \mathcal{A}$.

If $u(x,y) \geq u_{Z_{A_0}}\left(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\right)$ holds for all $(x,y) \in X \times Y$, then we set $A := A_0$ and $\varphi := \varphi_0$. This gives

$$d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X,\mu_Y) \geq d_{\mathrm{W},p}^{Z_A}\left(\left(\phi^X_{(A,\varphi)}\right)_\# \mu_X, \left(\psi^Y_{(A,\varphi)}\right)_\# \mu_Y\right).$$

Otherwise, there exists $(x,y) \in X \backslash A_0 \times Y \backslash \varphi_0(A_0)$ such that

$$u(x,y) < u_{Z_{A_0}}\left(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\right)$$

(if $x \in A_0$ or $y \in \varphi_0(A_0)$, then $u(x,y) \geq u_{Z_{A_0}}\left(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\right)$ must hold). Let $(x_1,y_1) \in X \backslash A_0 \times Y \backslash \varphi_0(A_0)$ be such that

$$\begin{aligned} u(x_1,y_1) = \min\Big\{ & u(x,y) \,|\, (x,y) \in X \backslash A_0 \times Y \backslash \varphi_0(A_0) \\ & \text{and } u(x,y) < u_{Z_{A_0}}\left(\phi^X_{(A_0,\varphi_0)}(x), \psi^Y_{(A_0,\varphi_0)}(y)\right) \Big\} > 0. \end{aligned}$$

The existence of (x_1,y_1) follows from finiteness of X and Y. It is easy to check that φ_0 extends to an isometry from $A_0 \cup \{x_1\}$ to $\varphi_0(A_0) \cup \{y_1\}$ by taking x_1 to y_1. We denote the new isometry φ_1 and set $A_1 := A_0 \cup \{x_1\}$. If for any $(x,y) \in X \times Y$, we have that $u(x,y) \geq u_{Z_{A_1}}\left(\phi^X_{(A_1,\varphi_1)}(x), \psi^Y_{(A_1,\varphi_1)}(y)\right)$, then we define $A := A_1$ and $\varphi := \varphi_1$. Otherwise, we continue the process to obtain $A_2, A_3, \ldots$. This process will eventually stop since we are considering finite spaces. Suppose the process stops at A_n, then $A := A_n$ and $\varphi := \varphi_n$ satisfy that $u(x,y) \geq u_{Z_A}\left(\phi^X_{(A,\varphi)}(x), \psi^Y_{(A,\varphi)}(y)\right)$ for any $(x,y) \in X \times Y$. Therefore,

$$d_{\mathrm{W},p}^{(X\sqcup Y,u)}(\mu_X,\mu_Y) \geq d_{\mathrm{W},p}^{Z_A}\left(\left(\phi^X_{(A,\varphi)}\right)_\# \mu_X, \left(\psi^Y_{(A,\varphi)}\right)_\# \mu_Y\right).$$

Since $u \in \mathcal{D}^{\mathrm{ult}}_{\mathrm{adm}}(u_X,u_Y)$ is arbitrary, this gives the claim. □

As a direct consequence of Theorem B.6, we obtain that it is sufficient, as claimed in Remark 3.8, for finite spaces to infimize in Equation (30) over the collection of all maximal pairs $\mathcal{A}^* \subseteq \mathcal{A}$. Recall that a pair $(A,\varphi_1) \in \mathcal{A}$ is denoted as *maximal*, if for all pairs $(B,\varphi_2) \in \mathcal{A}$ with $A \subseteq B$ and $\varphi_2|_A = \varphi_1$ it holds $A = B$.

Corollary B.7. *Let $\mathcal{X},\mathcal{Y} \in \mathcal{U}^w$ be* finite *spaces. Then, we have for each $p \in [1,\infty]$ that*

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = \inf_{(A,\varphi)\in\mathcal{A}^*} d_{\mathrm{W},p}^{Z_A}\left(\left(\phi^X_{(A,\varphi)}\right)_\# \mu_X, \left(\psi^Y_{(A,\varphi)}\right)_\# \mu_Y\right). \tag{31}$$

By proving Theorem B.6, we have verified Theorem 3.7 for finite ultrametric measure spaces. In the following, we will use Theorem B.6 and weighted quotients to demonstrate Theorem 3.7. However, before we come to this, we need to establish the following two auxiliary results.

Lemma B.8. *Let $X \in \mathcal{U}$ be a compact ultrametric space. Let $t > 0$ and let $p \in [1, \infty)$. Then, for any $\alpha, \beta \in \mathcal{P}(X)$, we have that*

$$\left(d_{\mathrm{W},p}^{X_t}(\alpha_t, \beta_t)\right)^p \geq \left(d_{\mathrm{W},p}^{X}(\alpha, \beta)\right)^p - t^p,$$

where α_t is the push forward of α under the canonical quotient map $Q_t : X \to X_t$ taking $x \in X$ to $[x]_t \in X_t$.

Proof. For any $\mu_t \in \mathcal{C}(\alpha_t, \beta_t)$, it is easy to see that there exists $\mu \in \mathcal{C}(\alpha, \beta)$ such that $\mu_t = (Q_t \times Q_t)_\# \mu$ where $Q_t \times Q_t : X \times X \to X_t \times X_t$ maps $(x, x') \in X \times X$ to $([x]_t, [x']_t)$. For example, suppose $X_t = \{[x_1]_t, \ldots, [x_n]_t\}$, then one can let

$$\mu := \sum_{i,j=1}^{n} \mu_t(([x_i]_t, [x_j]_t)) \frac{\alpha|_{[x_i]_t}}{\alpha([x_i]_t)} \otimes \frac{\beta|_{[x_j]_t}}{\beta([x_j]_t)},$$

where $\alpha|_{[x_i]_t}$ is the restriction of α on $[x_i]_t$.

For any $x, x' \in X$, we have that $(u_X(x, x'))^p \leq (u_{X_t}([x]_t, [x']_t))^p + t^p$. Then,

$$\begin{aligned}
\left(d_{\mathrm{W},p}^{X}(\alpha, \beta)\right)^p &\leq \int_{X \times X} \left(u_X(x, x')\right)^p \mu(dx \times dx') \\
&\leq \int_{X \times X} \left(\left(u_{X_t}([x]_t, [x']_t)\right)^p + t^p\right) \mu(dx \times dx') \\
&= \int_{X \times X} \left(u_X(Q_t(x), Q_t(x'))\right)^p \mu(dx \times dx') + t^p \\
&= \int_{X_t \times X_t} \left(u_{X_t}([x]_t, [x']_t)\right)^p \mu_t(d[x]_t \times d[x']_t) + t^p
\end{aligned}$$

Infimizing over all $\mu_t \in \mathcal{C}(\alpha_t, \beta_t)$, we obtain that

$$\left(d_{\mathrm{W},p}^{X_t}(\alpha_t, \beta_t)\right)^p \geq \left(d_{\mathrm{W},p}^{X}(\alpha, \beta)\right)^p - t^p.$$

□

Lemma B.9. *Let $\mathcal{X} \in \mathcal{U}^w$ and let $p \in [1, \infty]$. Then, for any $t > 0$, we have that*

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \leq t.$$

In particular, $\lim_{t \to 0} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) = 0$.

Proof. It is obvious that $(\mathcal{X}_t)_t \cong_w \mathcal{X}_t$. In consequence, it holds by Theorem 3.13 that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \leq t$. By Proposition 3.3 we have that for any $p \in [1, \infty]$

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \leq u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{X}) \leq t.$$

□

With Lemma B.8 and Lemma B.9 available, we can come to the proof of Theorem 3.7.

Proof of Theorem 3.7. Clearly, it follows from the definition of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Equation (10)) that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) \leq \inf_{(A,\varphi)\in\mathcal{A}} d_{\mathrm{W},p}^{Z_A}\left(\left(\phi_{(A,\varphi)}^X\right)_\# \mu_X, \left(\psi_{(A,\varphi)}^Y\right)_\# \mu_Y\right)$$

Hence, we focus on proving the opposite inequality.

Given any $t > 0$, by Lemma A.7, both $\mathcal{X}_t$ and $\mathcal{Y}_t$ are finite spaces. By Theorem B.6 we have that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) = \inf_{(A_t,\varphi_t)\in\mathcal{A}_t} d_{\mathrm{W},p}^{Z_{A_t}}\left(\left(\phi_{(A_t,\varphi_t)}^{X_t}\right)_\# (\mu_X)_t, \left(\psi_{(A_t,\varphi_t)}^{Y_t}\right)_\# (\mu_Y)_t\right),$$

where

$$\mathcal{A}_t := \{(A_t, \varphi_t) \,|\, \emptyset \neq A_t \subseteq X_t \text{ is closed and } \varphi_t : A_t \hookrightarrow Y_t \text{ is an isometric embedding }\}.$$

For any $(A_t, \varphi_t) \in \mathcal{A}_t$, assume that $A_t = \{[x_1]_t^X, \ldots, [x_n]_t^X\}$ and that $\varphi_t([x_i]_t) = [y_i]_t \in Y_t$ for all $i = 1, \ldots, n$. Let $A := \{x_1, \ldots, x_n\}$. Then, the map $\varphi : A \to Y$ defined by $x_i \mapsto y_i$ for $i = 1, \ldots, n$ is an isometric embedding. Therefore, $(A, \varphi) \in \mathcal{A}$.

Claim 1: $\left((Z_A)_t, u_{(Z_A)_t}\right) \cong \left(Z_{A_t}, u_{Z_{A_t}}\right)$.

Proof of the Claim. We define a map $\Psi : (Z_A)_t \to Z_{A_t}$ by $[x]_t^{Z_A} \mapsto [x]_t^X$ for $x \in X$ and $[y]_t^{Z_A} \mapsto [y]_t^Y$ for $y \in Y\backslash\varphi(A)$. We first show that Ψ is well-defined. For any $x' \in X$, if $u_{Z_A}(x, x') \leq t$, then obviously we have that $u_X(x, x') = u_{Z_A}(x, x') \leq t$ and thus $[x]_t^X = [x']_t^X$. Now, assume that there exists $y \in Y\backslash\varphi(A)$ such that $u_{Z_A}(x, y) \leq t$, i.e., $[x]_t^{Z_A} = [y]_t^{Z_A}$. Then, by finiteness of A and definition of Z_A, there exists $x_i \in A$ such that $u_{Z_A}(x, y) = \max\left(u_X(x, x_i), u_Y(\varphi(x_i), y)\right) \leq t$. This gives that

$$u_{Z_{A_t}}([x]_t^X, [y]_t^Y) \leq \max\left(u_{X_t}\left([x]_t^X, [x_i]_t^X\right), u_{Y_t}\left([\varphi(x_i)]_t^Y, [y]_t^Y\right)\right) \leq t.$$

However, this happens only if $u_{Z_{A_t}}([x]_t^X, [y]_t^Y) = 0$, that is, $[x]_t^X$ is identified with $[y]_t^Y$ under the map φ_t. Therefore, Ψ is well-defined.

It is easy to see from the definition that Ψ is surjective. Thus, it suffices to show that Ψ is an isometric embedding to finish the proof. For any $x, x' \in X$ such that $u_X(x, x') > t$, we have that

$$u_{(Z_A)_t}\left([x]_t^{Z_A}, [x']_t^{Z_A}\right) = u_{Z_A}(x, x') = u_X(x, x') = u_{X_t}\left([x]_t^X, [x']_t^X\right) = u_{Z_{A_t}}\left([x]_t^X, [x']_t^X\right).$$

Similarly, for any $y, y' \in Y \backslash \varphi(A)$ such that $u_Y(y, y') > t$, we have that

$$u_{(Z_A)_t}\left([y]_t^{Z_A}, [y']_t^{Z_A}\right) = u_{Z_{A_t}}\left([y]_t^Y, [y']_t^Y\right).$$

Now, consider $x \in X$ and $y \in Y \backslash \varphi(A)$. Assume that $u_{Z_A}(x, y) > t$ (otherwise $[x]_t^{Z_A} = [y]_t^{Z_A}$). Then, we have that

$$u_{Z_A}(x, y) = \min_{i=1,\ldots,n} \max\left(u_X(x, x_i), u_Y(\varphi(x_i), y)\right) > t.$$

This implies that

$$\begin{aligned}
u_{Z_{A_t}}\left([x]_t^X, [y]_t^Y\right) &= \min_{i=1,\ldots,n} \max\left(u_{X_t}\left([x]_t^X, [x_i]_t^X\right), u_{Y_t}\left(\varphi_t([x_i]_t^X), [y]_t^Y\right)\right) \\
&= \min_{i=1,\ldots,n} \max\left(u_X(x, x_i), u_Y(\varphi(x_i), y)\right) \\
&= u_{Z_A}(x, y) = u_{(Z_A)_t}\left([x]_t^{Z_A}, [y]_t^{Z_A}\right).
\end{aligned}$$

Therefore, Ψ is an isometric embedding and thus we conclude the proof. □

By Lemma B.8 we have that

$$\begin{aligned}
&\left(d_{\mathrm{W},p}^{Z_{A_t}}\left(\left(\phi_{(A_t,\varphi_t)}^{X_t}\right)_\# (\mu_X)_t, \left(\psi_{(A_t,\varphi_t)}^{Y_t}\right)_\# (\mu_Y)_t\right)\right)^p \\
&\geq \left(d_{\mathrm{W},p}^{Z_A}\left(\left(\phi_{(A,\varphi)}^{X}\right)_\# \mu_X, \left(\psi_{(A,\varphi)}^{Y}\right)_\# \mu_Y\right)\right)^p - t^p
\end{aligned}$$

Therefore,

$$\begin{aligned}
u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) &= \inf_{(A_t,\varphi_t)\in\mathcal{A}_t} d_{\mathrm{W},p}^{Z_{A_t}}\left(\left(\phi_{(A_t,\varphi_t)}^{X_t}\right)_\# (\mu_X)_t, \left(\psi_{(A_t,\varphi_t)}^{Y_t}\right)_\# (\mu_Y)_t\right) \\
&\geq \inf_{(A,\varphi)\in\mathcal{A}} \left(\left(d_{\mathrm{W},p}^{Z_A}\left(\left(\phi_{(A,\varphi)}^{X}\right)_\# \mu_X, \left(\psi_{(A,\varphi)}^{Y}\right)_\# \mu_Y\right)\right)^p - t^p\right)^{\frac{1}{p}}.
\end{aligned}$$

Notice that the last inequality already holds when we only consider (A, φ) corresponding to $(A_t, \varphi_t) \in \mathcal{A}_t$.

By Lemma B.9, we have that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \lim_{t\to 0} u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X}_t, \mathcal{Y}_t) \geq \inf_{(A,\varphi)\in\mathcal{A}} d_{\mathrm{W},p}^{Z_A}\left(\left(\phi_{(A,\varphi)}^{X}\right)_\# \mu_X, \left(\psi_{(A,\varphi)}^{Y}\right)_\# \mu_Y\right),$$

which concludes the proof. □

B.2 Proofs from Section 3.2

In the following, we give the complete proofs of the results stated in Section 3.2.

B.2.1 Proof of Proposition 3.9

1. This follows directly from the definitions of $u_{\mathrm{GW},p}$ and $d_{\mathrm{GW},p}$ (see Equation (13) and Equation (7)).
2. By Jensen's inequality we have that $\mathrm{dis}_p^{\mathrm{ult}}(\mu) \leq \mathrm{dis}_q^{\mathrm{ult}}(\mu)$ for any $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Therefore, $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \leq u_{\mathrm{GW},q}(\mathcal{X},\mathcal{Y})$.
3. By (2), we know that $\{u_{\mathrm{GW},n}(\mathcal{X},\mathcal{Y})\}_{n\in\mathbb{N}}$ is an increasing sequence with a finite upper bound $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$. Therefore, $L \coloneqq \lim_{n\to\infty} u_{\mathrm{GW},n}(\mathcal{X},\mathcal{Y})$ exists and it holds $L \leq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

   To prove the opposite inequality, by Proposition B.10, there exists for each $n \in \mathbb{N}$ $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ such that

   $$\left(\iint_{X\times Y\times X\times Y} \Lambda_\infty(u_X(x,x'), u_Y(y,y'))^n \mu_n(dx\times dy)\mu_n(dx'\times dy')\right)^{\frac{1}{n}} = u_{\mathrm{GW},n}(\mathcal{X},\mathcal{Y}).$$

   By Lemma B.19, $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges (after taking an appropriate subsequence) to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Let

   $$M = \sup_{(x,y),(x',y')\in \mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'), u_Y(y,y'))$$

   and for any given $\varepsilon > 0$ let

   $$U = \{((x,y),(x',y')) \in X\times Y\times X\times Y \,|\, \Lambda_\infty(u_X(x,x'), u_Y(y,y')) > M - \varepsilon\}.$$

   Then, we have $\mu\otimes\mu(U) > 0$. As μ_n weakly converges to μ, we have that $\mu_n \otimes \mu_n$ weakly converges to $\mu\otimes\mu$. Since U is open, there exists a small $\varepsilon_1 > 0$ such that $\mu_n\otimes\mu_n(U) > \mu\otimes\mu(U) - \varepsilon_1 > 0$ for n large enough (see e.g. Billingsley [7, Thm. 2.1]). Therefore,

   $$\begin{aligned}&\left(\iint_{X\times Y\times X\times Y} \Lambda_\infty(u_X(x,x'), u_Y(y,y'))^n \mu_n(dx\times dy)\mu_n(dx'\times dy')\right)^{\frac{1}{n}}\\ \geq&(\mu_n\otimes\mu_n(U))^{\frac{1}{n}}(M-\varepsilon) \geq (\mu\otimes\mu(U)-\varepsilon_1)^{\frac{1}{n}}(M-\varepsilon).\end{aligned}$$

   Letting $n\to\infty$, we obtain $L \geq M - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we obtain $L \geq M \geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

B.2.2 Proof of Theorem 3.10

One main step to verify Theorem 3.10 is to demonstrate the existence of optimal couplings.

Proposition B.10. *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, for any $p \in [1, \infty]$, there always exists an optimal coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu)$.*

Proof. We will only prove the claim for the case $p < \infty$ since the case $p = \infty$ can be proven in a similar manner. Let $\mu_n \in \mathcal{C}(\mu_X, \mu_Y)$ be such that

$$\left(\iint_{X\times Y\times X\times Y} \Lambda_\infty(u_X(x,x'), u_Y(y,y'))^p \, \mu_n(dx\times dy)\mu_n(dx'\times dy')\right)^{\frac{1}{p}} \leq u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) + \frac{1}{n}.$$

By Lemma B.19, $\{\mu_n\}_{n\in\mathbb{N}}$ weakly converges to some $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ (after taking an appropriate subsequence). Then, by the boundedness and continuity of $\Lambda_\infty(u_X, u_Y)$ on $X \times Y \times X \times Y$ (cf. Lemma B.22) as well as the weak convergence of $\mu_n \otimes \mu_n$, we have that that

$$\mathrm{dis}_p^{\mathrm{ult}}(\mu) = \lim_{n\to\infty} \mathrm{dis}_p^{\mathrm{ult}}(\mu_n) \leq u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}).$$

Hence, $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu)$. □

Based on Proposition B.10, it is straightforward to prove Theorem 3.10.

Proof of Theorem 3.10. It is clear that $u_{\mathrm{GW},p}$ is symmetric and that $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = 0$ if $\mathcal{X} \cong_w \mathcal{Y}$. Furthermore, we remark that $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})$ by Proposition 3.9. Since $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_w \mathcal{Y}$ (see [67]), we have that $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) = 0$ implies that $\mathcal{X} \cong_w \mathcal{Y}$. It remains to verify the p-triangle inequality. To this end, we only prove the case when $p < \infty$ whereas the case $p = \infty$ follows by analogous arguments.

Now let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be three ultrametric measure spaces. Let $\mu_{XY} \in \mathcal{C}(\mu_X, \mu_Y)$ and $\mu_{YZ} \in \mathcal{C}(\mu_Y, \mu_Z)$ be optimal (cf. Proposition B.10). By the Gluing Lemma [99, Lemma 7.6], there exists a measure $\mu_{XYZ} \in \mathcal{P}(X \times Y \times Z)$ with marginals μ_{XY} on $X \times Y$ and μ_{YZ} on $Y \times Z$. Further, we define $\mu_{XZ} = (\pi_{XZ})_\# \mu \in \mathcal{P}(X \times Z)$, where π_{XZ} denotes the canonical projection $X \times Y \times Z \to X \times Z$. Then,

$$
\begin{aligned}
&(u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Z}))^p \leq \iint_{X\times Z\times X\times Z} \left(\Lambda_\infty(u_X(x,x'),u_Z(z,z'))\right)^p \mu_{XZ}(dx\times dz)\mu_{XZ}(dx'\times dz')\\
=&\iint_{X\times Y\times Z\times X\times Y\times Z} \left(\Lambda_\infty(u_X(x,x'),u_Z(z,z'))\right)^p \mu_{XYZ}(dx\times dy\times dz)\mu_{XYZ}(dx'\times dy'\times dz')\\
\leq&\iint_{X\times Y\times Z\times X\times Y\times Z} \left(\Lambda_\infty(u_X(x,x'),u_Y(y,y'))\right)^p \mu_{XYZ}(dx\times dy\times dz)\mu_{XYZ}(dx'\times dy'\times dz')\\
+&\iint_{X\times Y\times Z\times X\times Y\times Z} \left(\Lambda_\infty(u_Y(y,y'),u_Z(z,z'))\right)^p \mu_{XYZ}(dx\times dy\times dz)\mu_{XYZ}(dx'\times dy'\times dz')\\
=&\iint_{X\times Y\times X\times Y} \left(\Lambda_\infty(u_X(x,x'),u_Y(y,y'))\right)^p \mu_{XY}(dx\times dy)\mu_{XY}(dx'\times dy')\\
+&\iint_{Y\times Z\times Y\times Z} \left(\Lambda_\infty(u_Y(y,y'),u_Z(z,z'))\right)^p \mu_{YZ}(dy\times dz)\mu_{YZ}(dy'\times dz')\\
=&(u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}))^p + (u_{\mathrm{GW},p}(\mathcal{Y},\mathcal{Z}))^p,
\end{aligned}
$$

where the second inequality follows from the fact that Λ_∞ in an ultrametric on $\mathbb{R}_{\geq 0}$ (cf. [70, Remark 1.14]) and the observation that an ultrametric is automatically a p-metric for any $p \in [1,\infty]$ [70, Proposition 1.16]. □

B.2.3 Proof of Theorem 3.13

We first prove that

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\} \tag{32}$$

and then show that the infimum is attainable.

Since $\mathcal{X}_0 \cong_w \mathcal{X}$ and $\mathcal{Y}_0 \cong_w \mathcal{Y}$, if $\mathcal{X}_0 \cong_w \mathcal{Y}_0$, then $\mathcal{X} \cong_w \mathcal{Y}$ and thus by Theorem 3.10

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = 0 = \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$$

Now, assume that for some $t > 0$, $\mathcal{X}_t \cong_w \mathcal{Y}_t$. By Lemma A.7, for some $n \in \mathbb{N}$ we can write $X_t = \{[x_1]_t, \ldots, [x_n]_t\}$ and $Y_t = \{[y_1]_t, \ldots, [y_n]_t\}$ such that $u_{X_t}([x_i]_t, [x_j]_t) = u_{Y_t}([y_i]_t, [y_j]_t)$ and $\mu_X([x_i]_t) = \mu_Y([y_i]_t)$. Let $\mu_X^i := \mu_X|_{[x_i]_t}$ and $\mu_Y^i := \mu_Y|_{[y_i]_t}$ for all $i = 1, \ldots, n$. Let $\mu := \sum_{i=1}^n \mu_X^i \otimes \mu_Y^i$. It is easy to check that $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ and $\mathrm{supp}(\mu) = \bigcup_{i=1}^n [x_i]_t \times [y_i]_t$. Assume $(x,y) \in [x_i]_t \times [y_i]_t$ and $(x',y') \in [x_j]_t \times [y_j]_t$. If $i \neq j$, then $u_{X_t}([x_i]_t, [x_j]_t) = u_{Y_t}([y_i]_t, [y_j]_t)$ and thus

$$\Lambda_\infty(u_X(x,x'), u_Y(y,y')) = \Lambda_\infty(u_{X_t}([x_i]_t,[x_j]_t), u_{Y_t}([y_i]_t,[y_j]_t)) = 0.$$

If $i = j$, then $u_X(x,x'), u_Y(y,y') \leq t$ and thus $\Lambda_\infty(u_X(x,x'), u_Y(y,y')) \leq t$. In either case, we have that

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \leq \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'), u_Y(y,y')) \leq t.$$

Therefore, $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \leq \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$.

Conversely, suppose $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ and let

$$t := \sup_{(x,y),(x',y')\in \mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'), u_Y(y,y')).$$

By Mémoli [67, Lemma 2.2], we know that $\mathrm{supp}(\mu)$ is a correspondence between X and Y. We define a map $f_t : X_t \to Y_t$ by taking $[x]_t^X \in X_t$ to $[y]_t^Y \in Y_t$ such that $(x,y) \in \mathrm{supp}(\mu)$. It is easy to check that f_t is well-defined and moreover f_t is an isometry (see for example the proof of Mémoli and Wan [70, Theorem 5.7]). Next, we prove that f_t is actually an isomorphism between $\mathcal{X}_t$ and $\mathcal{Y}_t$. For any $[x]_t^X \in X_t$, let $y \in Y$ be such that $(x,y) \in \mathrm{supp}\,(\mu)$ (in this case, $[y]_t^Y = f_t([x]_t^X)$). If there exists $(x',y') \in \mathrm{supp}(\mu)$ such that $x' \in [x]_t^X$ and $y' \notin [y]_t^Y$, then $\Lambda_\infty(u_X(x,x'), u_Y(y,y')) = u_Y(y,y') > t$, which is impossible. Consequently, $\mu([x]_t^X \times (Y \setminus [y]_t^Y)) = 0$ and similarly, $\mu((X \setminus [x]_t^X) \times [y]_t^Y) = 0$. This yields that

$$\mu_X([x]_t^X) = \mu([x]_t^Y \times Y) = \mu([x]_t^X \times [y]_t^Y) = \mu(X \times [y]_t^Y) = \mu_Y([y]_t^Y).$$

Therefore, f_t is an isomorphism between $\mathcal{X}_t$ and $\mathcal{Y}_t$. Hence, we have that $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \geq \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$ and hence $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$.

Finally, we show that the infimum of $\inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$ is attainable. Let $\delta := \inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$. If $\delta > 0$, let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence converging to δ such that $\mathcal{X}_{t_n} \cong_w \mathcal{Y}_{t_n}$ for all t_n. Since $\mathcal{X}_\delta$ and $\mathcal{Y}_\delta$ are finite spaces, we actually have that $\mathcal{X}_{t_n} = \mathcal{X}_\delta$ and $\mathcal{Y}_{t_n} = \mathcal{Y}_\delta$ when n is large enough. This immediately implies that $\mathcal{X}_\delta \cong_w \mathcal{Y}_\delta$. Now, if $\delta = 0$, then by Equation (32) we have that $u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) = \delta = 0$. By Theorem 3.10, $\mathcal{X} \cong_w \mathcal{Y}$. This is equivalent to $\mathcal{X}_\delta \cong_w \mathcal{Y}_\delta$. Therefore, the infimum of $\inf\{t \geq 0 \,|\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$ is always attainable.

B.2.4 Proof of Theorem 3.17

An important observation for the proof of Theorem 3.17 is that the snowflake transform relates the p-Wasserstein pseudometric on a pseudo-ultrametric space X with the 1-Wasserstein pseudometric on the space $S_p(X)$, $1 \leq p < \infty$.

Lemma B.11. *Given a pseudo-ultrametric space (X, u_X) and $p \geq 1$, we have for any $\alpha, \beta \in \mathcal{P}(X)$ that*

$$d_{\mathrm{W},p}^{(X,u_X)}(\alpha,\beta) = \left(d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta)\right)^{\frac{1}{p}}.$$

Remark B.12. Since $S_p \circ u_X$ and u_X induce the same topology and thus the same Borel sets on X, we have that $\mathcal{P}(X) = \mathcal{P}(S_p(X))$ and thus the expression $d_{\mathrm{W},1}^{S_p(X)}(\alpha,\beta)$ in the lemma is well defined.

Proof of Lemma B.11. Suppose $\mu_1, \mu_2 \in \mathcal{C}(\alpha,\beta)$ are the optimal couplings for $d^X_{\mathrm{W},p}(\alpha,\beta)$ and $d^{S_p(X)}_{\mathrm{W},1}(\alpha,\beta)$, respectively (see Appendix B.5.1 for the existence of μ_1 and μ_2). Then,

$$\left(d^{(X,u_X)}_{\mathrm{W},p}(\alpha,\beta)\right)^p = \int_{X\times X}(u_X(x,y))^p\mu_1(dx\times dy)$$
$$= \int_{X\times X} S_p(u_X)(x,y)\mu_1(dx\times dy) \geq d^{S_p(X)}_{\mathrm{W},1}(\alpha,\beta),$$

and

$$d^{S_p(X)}_{\mathrm{W},1}(\alpha,\beta) = \int_{X\times X} S_p(u_X)(x,y)\mu_2(dx\times dy)$$
$$= \int_{X\times X}(u_X(x,y))^p\mu_2(dx\times dy) \geq \left(d^{(X,u_X)}_{\mathrm{W},p}(\alpha,\beta)\right)^p.$$

Therefore, $d^{(X,u_X)}_{\mathrm{W},p}(\alpha,\beta) = \left(d^{S_p(X)}_{\mathrm{W},1}(\alpha,\beta)\right)^{\frac{1}{p}}$. □

With Lemma B.11 at our disposal we can prove Theorem 3.17.

Proof of Theorem 3.17. Let $\mu \in \mathcal{C}(\mu_X,\mu_Y)$. Then,

$$\iint_{X\times Y\times X\times Y}\left(\Lambda_\infty(u_X(x,x'),u_Y(y,y'))\right)^p\mu(dx\times dy)\,\mu(dx'\times dy')$$
$$= \iint_{X\times Y\times X\times Y}\Lambda_\infty\left(u_X(x,x')^p,u_Y(y,y')^p\right)\mu(dx\times dy)\,\mu(dx'\times dy').$$

By infimizing over $\mu\in\mathcal{C}(\mu_X,\mu_Y)$ on both sides, we obtain that

$$(u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}))^p = u_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y})).$$

To prove the second part of the claim, let $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$. By Lemma B.11 we have that

$$\left(d^{(X\sqcup Y,u)}_{\mathrm{W},p}(\mu_X,\mu_Y)\right)^p = d^{(S_p(X)\sqcup S_p(Y),S_p(u))}_{\mathrm{W},1}(\mu_X,\mu_Y).$$

Finally, infimizing over $u\in\mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ yields

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})^p = u^{\mathrm{sturm}}_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y})).$$

□

As a direct consequence of Theorem 3.17, we obtain the following relation between the spaces $(\mathcal{U}^w, u^{\mathrm{sturm}}_{\mathrm{GW},1})$ and $\left(\mathcal{U}^w, u^{\mathrm{sturm}}_{\mathrm{GW},p}\right)$ for $p\in[1,\infty)$.

Corollary B.13. *For each* $p \in [1,\infty)$*, the metric space* $(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ *is isometric to the snowflake transform of* $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$*, i.e.,*

$$S_p\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right) \cong \left(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}}\right)$$

Proof. Consider the snowflake transform map $S_p : \mathcal{U}^w \to \mathcal{U}^w$ sending $X \in \mathcal{U}^w$ to $S_p(X) \in \mathcal{U}^w$. It is obvious that S_p is bijective. By Theorem 3.17, we know that S_p is an isometry from $S_p\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ to $\left(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}}\right)$. Therefore, $S_p\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right) \cong \left(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}}\right)$. □

B.3 Proofs from Section 3.3

Throughout the following, we demonstrate the open claims from Section 3.3.

B.3.1 Proof of Theorem 3.18

First, we focus on the statement for $p = 1$, i.e., on showing

$$u_{\mathrm{GW},1}(\mathcal{X}, \mathcal{Y}) \leq 2\, u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}). \tag{33}$$

Let $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be such that $u_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X}, \mathcal{Y}) = \int u(x,y)\mu(dx \times dy)$. The existence of u and μ follows from Proposition B.1

Claim 1: For any $(x,y),(x',y') \in X \times Y$, we have

$$\Lambda_\infty(u_X(x,x'), u_Y(y,y')) \leq \max(u(x,y), u(x',y')) \leq u(x,y) + u(x',y').$$

Proof. We only need to show that

$$\Lambda_\infty(u_X(x,x'), u_Y(y,y')) \leq \max(u(x,y), u(x',y')).$$

If $u_X(x,x') = u_Y(y,y')$, then there is nothing to prove. Otherwise, we assume without loss of generality that $u_X(x,x') < u_Y(y,y')$. If $\max(u(x,y), u(x',y')) < u_Y(y,y')$, then by the strong triangle inequality we must have $u(x,y') = u_Y(y,y') = u(x',y)$. However, $u(x',y) \leq \max(u_X(x,x'), u(x,y)) < u_Y(y,y')$, which leads to a contradiction. Therefore, $\Lambda_\infty(u_X(x,x'), u_Y(y,y')) \leq \max(u(x,y), u(x',y'))$. □

By Claim 1, we have

$$\begin{aligned}
&\iint_{X\times Y\times X\times Y} \Lambda_\infty(u_X(x,x'),u_Y(y,y'))\,\mu(dx\times dy)\,\mu(dx'\times dy')\\
\le&\iint_{X\times Y\times X\times Y} u(x,y)\,\mu(dx\times dy)\,\mu(dx'\times dy')\\
+&\iint_{X\times Y\times X\times Y} u(x',y')\,\mu(dx\times dy)\,\mu(dx'\times dy')\\
=&\int_{X\times Y} u(x,y)\,\mu(dx\times dy)+\int_{X\times Y} u(x',y')\,\mu(dx'\times dy')\le 2u^{\mathrm{sturm}}_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y}).
\end{aligned}$$

Therefore, $u_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y})\le 2\,u^{\mathrm{sturm}}_{\mathrm{GW},1}(\mathcal{X},\mathcal{Y})$.

Applying Theorem 3.17 and Equation (33), yields that for any $p\in[1,\infty)$

$$u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y})=(u_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y})))^{\frac{1}{p}}\le\left(2\,u^{\mathrm{sturm}}_{\mathrm{GW},1}(S_p(\mathcal{X}),S_p(\mathcal{Y}))\right)^{\frac{1}{p}}=2^{\frac{1}{p}}\,u^{\mathrm{sturm}}_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}).$$

B.3.2 Proof of result in Example 3.20

It follows from [67, Remark 5.17] that

$$d^{\mathrm{sturm}}_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)\ge\frac{1}{4}\text{ and }d_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)\le\frac{1}{2}\left(\frac{3}{2n}\right)^{\frac{1}{p}}.$$

Then, by Proposition 3.3, we have that

$$u^{\mathrm{sturm}}_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)\ge d^{\mathrm{sturm}}_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)\ge\frac{1}{4}.$$

Let μ_n denote the uniform probability measure of $\hat\Delta_n(1)$. Since $\hat\Delta_n(1)$ has the constant interpoint distance 1, it is obvious that for any coupling $\mu\in\mathcal{C}(\mu_n,\mu_{2n})$,

$$\mathrm{dis}_p(\mu)=\mathrm{dis}^{\mathrm{ult}}_p(\mu)$$

This implies that

$$u_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)=2\,d_{\mathrm{GW},p}\left(\hat\Delta_n(1),\hat\Delta_{2n}(1)\right)\le\left(\frac{3}{2n}\right)^{\frac{1}{p}}.$$

B.3.3 Proof of Theorem 3.21

First, we prove that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$. Indeed, for any $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and $\mu \in \mathcal{C}(\mu_X,\mu_Y)$, we have that

$$\begin{aligned}\sup_{(x,y)\in\mathrm{supp}(\mu)} u(x,y) &= \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \max(u(x,y),u(x',y'))\\ &\geq \sup_{(x,y),(x',y')\in\mathrm{supp}(\mu)} \Lambda_\infty(u_X(x,x'),u_Y(y,y'))\\ &\geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}),\end{aligned}$$

where the first inequality follows from Claim 1 in the proof of Theorem 3.18. Then, by a standard limit argument, we conclude that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \geq u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

Next, we prove that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq \min\{t \geq 0 |\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$. Let $t > 0$ be such that $\mathcal{X}_t \cong_w \mathcal{Y}_t$ and let $\varphi : \mathcal{X}_t \to \mathcal{Y}_t$ denote such an isomorphism. Then, we define a function $u : X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\geq 0}$ as follows:

1. $u|_{X\times X} \coloneqq u_X$ and $u|_{Y\times Y} \coloneqq u_Y$;
2. for any $(x,y) \in X \times Y$, $u(x,y) \coloneqq \begin{cases} u_{Y_t}(\varphi([x]_t^X),[y]_t^Y), & \text{if } \varphi([x]_t^X) \neq [y]_t^Y \\ t, & \text{if } \varphi([x]_t^X) = [y]_t^Y. \end{cases}$
3. for any $(y,x) \in Y \times X$, $u(y,x) \coloneqq u(x,y)$.

Then, it is easy to verify that $u \in \mathcal{D}^{\mathrm{ult}}(u_X,u_Y)$ and that u is actually an ultrametric. Let $Z \coloneqq (X \sqcup Y, u)$. By Lemma 2.8, we have

$$u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) = \max_{B\in V(Z)\setminus\{Z\} \text{ and } \mu_X(B)\neq\mu_Y(B)} \mathrm{diam}\,(B^*)\,.$$

We verify that $d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) \leq t$ in the following. It is obvious that $Z_t \cong X_t \cong Y_t$. Write $X_t = \{[x_i]_t^X\}_{i=1}^n$ and $Y_t = \{[y_i]_t^Y\}_{i=1}^n$ such that $[y_i]_t^Y = \varphi([x_i]_t^X)$ for each $i = 1,\ldots,n$. Then, $[x_i]_t^Z = [y_i]_t^Z$ and $Z_t = \{[x_i]_t^Z |\, i = 1,\ldots,n\}$. Since φ is an isomorphism, for any $i = 1,\ldots,n$ we have that $\mu_X([x_i]_t^X) = \mu_Y([y_i]_t^Y)$ and thus $\mu_X([x_i]_t^Z) = \mu_Y([y_i]_s^Z) = \mu_Y([x_i]_t^Z)$ when μ_X and μ_Y are regarded as pushforward measures under the inclusion map $X \hookrightarrow Z$ and $Y \hookrightarrow Z$, respectively. Now for any $B \in V(Z)$ (cf. Section 2.3), if $\mathrm{diam}\,(B) \geq t$, then B is the union of certain $[x_i]_t^Z$'s in Z_t and thus $\mu_X(B) = \mu_Y(B)$. If $\mathrm{diam}\,(B) < t$ and $\mathrm{diam}\,(B^*) > t$, then there exists some x_i such that $B = [x_i]_s^Z$ and $[x_i]_s^Z = [x_i]_t^Z$ where $s \coloneqq \mathrm{diam}\,(B)$. This implies that $\mu_X(B) = \mu_Y(B)$. In consequence, we have that $d_{\mathrm{W},\infty}^{Z}(\mu_X,\mu_Y) \leq t$ and thus $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq d_{\mathrm{W},\infty}^{(X\sqcup Y,u)}(\mu_X,\mu_Y) \leq t$. Therefore, $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) \leq \inf\{t \geq 0 |\, \mathcal{X}_t \cong_w \mathcal{Y}_t\}$.

Finally, by invoking Theorem 3.13, we conclude that $u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y})$.

B.3.4 Proof of Theorem 3.22

We prove the result via an explicit construction. By Theorem 3.21, we have

$$s = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}).$$

By Theorem 3.13, there exists an isomorpism $\varphi : \mathcal{X}_s \to \mathcal{Y}_s$. Since $s > 0$, by Lemma A.7, both $\mathcal{X}_s$ and $\mathcal{Y}_s$ are finite spaces. Let $X_s = \{[x_1]_s^X, \ldots, [x_n]_s^X\}$, $Y_s = \{[y_1]_s^Y, \ldots, [y_n]_s^Y\}$ and assume $[y_i]_s^Y = \varphi([x_i]_s^X)$ for each $i = 1, \ldots, n$. Let $A \coloneqq \{x_1, \ldots, x_n\}$ and define $\phi : A \to Y$ by sending x_i to y_i for each $i = 1, \ldots, n$. We prove that (A, ϕ) satisfies the conditions in the statement.

Since φ is an isomorphism, for any $1 \le i < j \le n$,

$$u_Y(y_i, y_j) = u_{Y_s}([y_i]_s^Y, [y_j]_s^Y) = u_{Y_s}(\varphi([x_i]_s^X), \varphi([x_j]_s^X)) = u_{X_s}([x_i]_s^X, [x_j]_s^X) = u_X(x_i, x_j).$$

This implies that $\phi : A \to Y$ is an isometric embedding and thus $(A, \phi) \in \mathcal{A}$.

It is obvious that $(Z_A)_s$ is isometric to both X_s and Y_s. In fact, $[x_i]_s^{Z_A} = [y_i]_s^{Z_A}$ in Z_A for each $i = 1, \ldots, n$ and $(Z_A)_s = \{[x_i]_s^{Z_A} | \, i = 1, \ldots, n\}$. Since φ is an isomorphism, for any $i = 1, \ldots, n$ we have that $\mu_X([x_i]_s^X) = \mu_Y([y_i]_s^Y)$ and thus $\mu_X([x_i]_s^{Z_A}) = \mu_Y([y_i]_s^{Z_A}) = \mu_Y([x_i]_s^{Z_A})$ when μ_X and μ_Y are regarded as pushforward measures under the inclusion maps $X \to Z_A$ and $Y \to Z_A$, respectively. Now for any $B \in V(Z_A)$ (cf. Section 2.3), if $\mathrm{diam}\,(B) \ge s$, then B is the union of certain $[x_i]_s^{Z_A}$'s and thus $\mu_X(B) = \mu_Y(B)$. If otherwise $\mathrm{diam}\,(B) < s$ and $\mathrm{diam}\,(B^*) > s$, then there exists x_i such that $B = [x_i]_t^{Z_A}$ and $[x_i]_t^{Z_A} = [x_i]_s^{Z_A}$ where $t \coloneqq \mathrm{diam}\,(B)$. This implies that $\mu_X(B) = \mu_Y(B)$. By Lemma 2.8, we have $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y) \le s$ and thus $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y) = s$ since $d_{\mathrm{W},\infty}^{Z_A}(\mu_X, \mu_Y)$ is an upper bound for $s = u_{\mathrm{GW},\infty}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})$ due to Equation (10).

B.3.5 Proof of Theorem 3.24

In this section, we prove Theorem 3.24 by slightly modifying the proof of Proposition 5.3 in [67].

Lemma B.14. *Let (X, u_X) and (Y, u_Y) be compact ultrametric spaces and let $S \subseteq X \times Y$ be non-empty. Assume that $\sup_{(x,y),(x',y')\in S} \Lambda_\infty(u_X(x,x'), u_Y(y,y')) \le \eta$. Define $u_S : X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\ge 0}$ as follows:*

1. *$u_S|_{X\times X} \coloneqq u_X$ and $u_S|_{Y\times Y} \coloneqq u_Y$;*
2. *for any $(x,y) \in X \times Y$, $u_S(x,y) \coloneqq \inf_{(x',y')\in S} \max\left(u_X(x,x'), u_Y(y,y'), \eta\right).$*
3. *for any $(x,y) \in X \times Y$, $u_S(y,x) \coloneqq u_S(x,y)$.*

Then, $u_S \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and $u_S(x,y) \le \eta$ for all $(x,y) \in S$.

Proof. That $u_S \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ essentially follows by Zarichnyi [105, Lemma 1.1]. It remains to prove the second half of the statement. For $(x, y) \in S$, we set $(x', y') := (x, y)$. This yields

$$u_S(x, y) \leq \max(u_X(x, x'), u_Y(y, y'), \eta) = \max(0, 0, \eta) = \eta.$$

□

Proof of Theorem 3.24. Let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be a coupling such that $\left\| \Gamma^{\infty}_{X,Y} \right\|_{L^p(\mu \otimes \mu)} < \delta^5$. Set $\varepsilon := 4v_\delta(X) \leq 4$.

By Mémoli [67, Claim 10.1], there exist a positive integer $N \leq [1/\delta]$ and points $x_1, \ldots, x_N$ in X such that $\min_{i \neq j} u_X(x_i, x_j) \geq \frac{\varepsilon}{2}$, $\min_i \mu_X \left(B^X_\varepsilon(x_i) \right) > \delta$ and $\mu_X \left(\bigcup_{i=1}^N B^X_\varepsilon(x_i) \right) \geq 1 - \varepsilon$.

Claim 1: For every $i = 1, \ldots, N$ there exists $y_i \in Y$ such that

$$\mu \left(B^X_\varepsilon(x_i) \times B^Y_{2(\varepsilon+\delta)}(y_i) \right) \geq (1 - \delta^2) \mu_X \left(B^X_\varepsilon(x_i) \right).$$

Proof. Assume the claim is false for some i and let $Q_i(y) = B^X_\varepsilon(x_i) \times \left(Y \setminus B^Y_{2(\varepsilon+\delta)}(y) \right)$. Then, as $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ it holds

$$\begin{aligned} \mu_X \left(B^X_\varepsilon(x_i) \right) =& \mu \left(B^X_\varepsilon(x_i) \times Y \right) \\ =& \mu \left(B^X_\varepsilon(x_i) \times B^Y_{2(\varepsilon+\delta)}(y) \right) + \mu \left(B^X_\varepsilon(x_i) \times \left(Y \setminus B^Y_{2(\varepsilon+\delta)}(y) \right) \right). \end{aligned}$$

Consequently, we have that $\mu(Q_i(y)) \geq \delta^2 \mu_X \left(B^X_\varepsilon(x_i) \right)$. Further, let

$$\mathcal{Q}_i := \left\{ (x, y, x', y') \in X \times Y \times X \times Y \,|\, x, x' \in B^X_\varepsilon(x_i) \text{ and } u_Y(y, y') \geq 2(\varepsilon + \delta) \right\}.$$

Clearly, it holds for $(x, y, x', y') \in \mathcal{Q}_i$ that

$$\Gamma^{\infty}_{X,Y}(x, y, x', y') = \Lambda_\infty \left(u_X(x, x'), u_Y(y, y') \right) = u_Y(y, y') \geq 2\delta.$$

Further, we have that $\mu \otimes \mu(\mathcal{Q}_i) \geq \delta^4$. Indeed, it holds

$$\begin{aligned} \mu \otimes \mu(\mathcal{Q}_i) =& \int_{B^X_\varepsilon(x_i) \times Y} \int_{Q_i(y)} 1 \, \mu(dx' \times dy') \, \mu(dx \times dy) \\ =& \int_{B^X_\varepsilon(x_i) \times Y} \mu(Q_i(y)) \mu(dx \times dy) \\ =& \mu_X \left(B^X_\varepsilon(x_i) \right) \int_Y \mu(Q_i(y)) \mu_Y(dy) \\ \geq& \left(\mu_X \left(B^X_\varepsilon(x_i) \right) \right)^2 \delta^2 \\ \geq& \delta^4. \end{aligned}$$

However, this yields that

$$\left\|\Gamma^{\infty}_{X,Y}\right\|_{L^p(\mu\otimes\mu)} \geq \left\|\Gamma^{\infty}_{X,Y}\right\|_{L^1(\mu\otimes\mu)} \geq \left\|\Gamma^{\infty}_{X,Y}\mathbb{1}_{\mathcal{Q}_i}\right\|_{L^1(\mu\otimes\mu)} \geq 2\delta\cdot\mu\otimes\mu(\mathcal{Q}_i) \geq 2\delta^5,$$

which contradicts $\left\|\Gamma^{\infty}_{X,Y}\right\|_{L^p(\mu\otimes\mu)} < \delta^5$. □

Define for each $i = 1,\ldots,N$

$$S_i := B^X_\varepsilon(x_i) \times B^Y_{2(\varepsilon+\delta)}(y_i).$$

Then, by Claim 1, $\mu(S_i) \geq \delta(1-\delta^2)$, for all $i = 1,\ldots,N$.

Claim 2: $\Gamma^{\infty}_{X,Y}(x_i,y_i,x_j,y_j) \leq 6(\varepsilon+\delta)$ for all $i,j = 1,\ldots,N$.

Proof. Assume the claim fails for some (i_0,j_0), i.e.,

$$\Lambda_\infty(u_X(x_{i_0},x_{j_0}),u_Y(y_{i_0},y_{j_0})) > 6(\varepsilon+\delta) > 0.$$

Then, we have $\Lambda_\infty(u_X(x_{i_0},x_{j_0}),u_Y(y_{i_0},y_{j_0})) = \max(u_X(x_{i_0},x_{j_0}),u_Y(y_{i_0},y_{j_0}))$. We assume without loss of generality that

$$u_X(x_{i_0},x_{j_0}) = \Lambda_\infty(u_X(x_{i_0},x_{j_0}),u_Y(y_{i_0},y_{j_0})) > u_Y(y_{i_0},y_{j_0}).$$

Consider any $(x,y) \in S_{i_0}$ and $(x',y') \in S_{j_0}$. By the strong triangle inequality and the fact that $u_X(x_{i_0},x_{j_0}) > 6(\varepsilon+\delta) > \varepsilon$, it is easy to verify that $u_X(x,x') = u_X(x_{i_0},x_{j_0})$. Moreover,

$$\begin{aligned}u_Y(y,y') &\leq \max(u_Y(y,y_{i_0}),u_Y(y_{i_0},y_{j_0}),u_Y(y_{j_0},y'))\\ &< \max(2(\varepsilon+\delta),u_X(x_{i_0},x_{j_0}),2(\varepsilon+\delta)) = u_X(x_{i_0},x_{j_0}) = u_X(x,x').\end{aligned}$$

Therefore,

$$\Gamma^{\infty}_{X,Y}(x,y,x',y') = u_X(x,x') = u_X(x_{i_0},x_{j_0}) = \Gamma^{\infty}_{X,Y}(x_{i_0},y_{i_0},x_{j_0},y_{j_0}) > 6(\varepsilon+\delta) > 2\delta.$$

Consequently, we have that

$$\begin{aligned}&\left\|\Gamma^{\infty}_{X,Y}\right\|_{L^p(\mu\otimes\mu)} \geq \left\|\Gamma^{\infty}_{X,Y}\right\|_{L^1(\mu\otimes\mu)} \geq \left\|\Gamma^{\infty}_{X,Y}\mathbb{1}_{S_{i_0}}\mathbb{1}_{S_{j_0}}\right\|_{L^1(\mu\otimes\mu)} \geq 2\delta\mu(S_{i_0})\mu(S_{j_0})\\ &> 2\delta\left(\delta(1-\delta^2)\right)^2.\end{aligned}$$

However, for $\delta \leq 1/2$, $2\delta\left(\delta(1-\delta^2)\right)^2 \geq 2\delta^5$. This leads to a contradiction. □

Consider $S \subseteq X \times Y$ given by $S := \{(x_i, y_i) | \, i = 1, \ldots, N\}$. Let u_S be the ultrametric on $X \sqcup Y$ given by Lemma B.14. By Claim 2, $\sup_{(x,y),(x',y') \in S} \Gamma^{\infty}_{X,Y}(x, y, x', y') \leq 6(\varepsilon + \delta)$. Then, for all $i = 1, \ldots, N$ we have that $u_S(x_i, y_i) \leq 6(\varepsilon + \delta)$ and for any $(x, y) \in X \times Y$ we have that

$$u_S(x, y) \leq \max(\operatorname{diam}(X), \operatorname{diam}(Y), 6(\varepsilon + \delta)) \leq \max(\operatorname{diam}(X), \operatorname{diam}(Y), 27) =: M'.$$

Here in the second inequality we use the assumption that $\delta < \frac{1}{2}$ and the fact that $\varepsilon = 4v_\delta(X) \leq 4$.

Claim 3: Fix $i \in \{1, \ldots, N\}$. Then, for all $(x, y) \in S_i$, it holds $u_S(x, y) \leq 6(\varepsilon + \delta)$.

Proof. Let $(x, y) \in S_i$. Then, $u_X(x, x_i) \leq \varepsilon$ and $u_Y(y, y_i) \leq 2(\varepsilon + \delta)$. Then, by the strong triangle inequality for u_S we obtain

$$\begin{aligned} u_S(x, y) &\leq \max\{u_X(x, x_i), u_Y(y, y_i), u_S(x_i, y_i)\} \\ &\leq \max\{\varepsilon, 2(\varepsilon + \delta), 6(\varepsilon + \delta)\} \leq 6(\varepsilon + \delta). \end{aligned}$$

□

Let $L := \bigcup_{i=1}^N S_i$. The next step is to estimate the mass of μ in the complement of L.

Claim 4: $\mu(X \times Y \backslash L) \leq \varepsilon + \delta$.

Proof. For each $i = 1, \ldots, N$, let $A_i := B^X_\varepsilon(x_i) \times \left(Y \setminus B^Y_{2(\varepsilon+\delta)}(y_i)\right)$. Then,

$$A_i = \left(B^X_\varepsilon(x_i) \times Y\right) \setminus \left(B^X_\varepsilon(x_i) \times B^Y_{2(\varepsilon+\delta)}(y_i)\right) = \left(B^X_\varepsilon(x_i) \times Y\right) \setminus S_i.$$

Hence,

$$\mu(A_i) = \mu\left(B^X_\varepsilon(x_i) \times Y\right) - \mu(S_i) = \mu_X\left(B^X_\varepsilon(x_i)\right) - \mu(S_i),$$

where the last equality follows from the fact that $\mu \in \mathcal{M}(\mu_X, \mu_Y)$. By Claim 1, we have that $\mu(S_i) \geq \mu_X\left(B^X_\varepsilon(x_i)\right)(1 - \delta^2)$. Consequently, we obtain

$$\mu(A_i) \leq \mu_X\left(B^X_\varepsilon(x_i)\right)\delta^2.$$

Notice that

$$X \times Y \setminus L \subseteq \left(X \setminus \bigcup_{i=1}^N B^X_\varepsilon(x_i)\right) \times Y \cup \left(\bigcup_{i=1}^N A_i\right).$$

Hence,

$$\begin{aligned}\mu(X\times Y\setminus L) &\leq \mu_X\left(X\Big\backslash\bigcup_{i=1}^N B_\varepsilon^X(x_i)\right)+\sum_{i=1}^N\mu(A_i)\\ &\leq 1-\mu_X\left(\bigcup_{i=1}^N B_\varepsilon^X(x_i)\right)+\sum_{i=1}^N\delta^2\mu_X\left(B_\varepsilon^X(x_i)\right)\\ &\leq \varepsilon+N\cdot\delta^2\leq\varepsilon+\delta.\end{aligned}$$

Here, the third inequality follows from the construction of x_is in the beginning of this section and from the fact that $N \leq [1/\delta]$. □

Now,

$$\begin{aligned}\int_{X\times Y}u_S^p(x,y)\,\mu(dx\times dy) &= \left(\int_L+\int_{X\times Y\setminus L}\right)u_S^p(x,y)\,\mu(dx\times dy)\\ &\leq (6(\varepsilon+\delta))^p+M'^p\cdot(\varepsilon+\delta).\end{aligned}$$

Since we have for any $a, b \geq 0$ and $p \geq 1$ that $a^{1/p}+b^{1/p}\geq(a+b)^{1/p}$, we obtain

$$\begin{aligned}u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) &\leq (\varepsilon+\delta)^{\frac{1}{p}}\left(6(\varepsilon+\delta)^{1-\frac{1}{p}}+M'\right)\leq(\varepsilon+\delta)^{\frac{1}{p}}\left(27+M'\right)\\ &\leq (4v_\delta(\mathcal{X})+\delta)^{\frac{1}{p}}\cdot M,\end{aligned}$$

where we used $\varepsilon = 4v_\delta(\mathcal{X})$ and $M := 2\max(\mathrm{diam}\,(X)\,,\mathrm{diam}\,(Y))+54 \geq M'+27$. Since the roles of $\mathcal{X}$ and $\mathcal{Y}$ are symmetric, we have that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y})\leq(4\min(v_\delta(\mathcal{X}),v_\delta(Y))+\delta)^{\frac{1}{p}}\cdot M.$$

This concludes the proof. □

B.4 Proofs from Section 3.4

The subsequent section contains the full proofs of the statements in Section 3.4.

B.4.1 Proof of Theorem 3.26

1. We first prove that $(\mathcal{U}^w, u_{\mathrm{GW},p})$ is non-separable for each $p\in[1,\infty]$. Recall notations in Example 3.5 and consider the family $\{\hat{\Delta}_2(a)\}_{a\in[1,2]}$.

   Claim 1: $\forall a\neq b\in[1,2]$, $u_{\mathrm{GW},p}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)=2^{-\frac{1}{p}}\Lambda_\infty(a,b)\geq 2^{-\frac{1}{p}}$, where we let $2^{-\frac{1}{\infty}}=1$.

Proof of Claim 1 . First note by Theorem 4.1 that

$$u_{\mathrm{GW},p}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)\geq \mathbf{SLB}_p^{\mathrm{ult}}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right).$$

It is easy to verify that $\mathbf{SLB}_p^{\mathrm{ult}}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)=2^{-\frac{1}{p}}\Lambda_\infty(a,b)$. On the other hand, consider the diagonal coupling between μ_a and μ_b, then for $p\in[1,\infty)$

$$u_{\mathrm{GW},p}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)\leq\left(2\cdot\Lambda_\infty(a,b)^p\cdot\frac{1}{2}\cdot\frac{1}{2}\right)^{\frac{1}{p}}=2^{-\frac{1}{p}}\Lambda_\infty(a,b),$$

and for $p=\infty$

$$u_{\mathrm{GW},\infty}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)\leq\Lambda_\infty(a,b).$$

Therefore,

$$u_{\mathrm{GW},p}\left(\hat{\Delta}_2(a),\hat{\Delta}_2(b)\right)=2^{-\frac{1}{p}}\Lambda_\infty(a,b).$$

□

By Claim 1, we have that $\left\{\hat{\Delta}_2(a)\right\}_{a\in[1,2]}$ is an uncountable subset of $\mathcal{U}^w$ with pairwise distance greater than $2^{-\frac{1}{p}}$, which implies that $(\mathcal{U}^w,u_{\mathrm{GW},p})$ is non-separable.

Now for $p\in[1,\infty)$, we show that $u_{\mathrm{GW},p}$ is not complete. Consider the family $\{\Delta_{2^n}(1)\}_{n\in\mathbb{N}}$ of 2^n-point spaces with unitary interpoint distances. Endow each space $\Delta_{2^n}(1)$ with the uniform measure μ_n and denote the corresponding ultrametric measure space by $\hat{\Delta}_{2^n}(1)$. It is proven in [93, Example 2.2] that $\{\hat{\Delta}_{2^n}(1)\}_{n\in\mathbb{N}}$ is a Cauchy sequence with respect to $d_{\mathrm{GW},p}$ without a compact metric measure space as limit. It is not hard to check that

$$u_{\mathrm{GW},p}\left(\hat{\Delta}_{2^m}(1),\hat{\Delta}_{2^n}(1)\right)=2d_{\mathrm{GW},p}\left(\hat{\Delta}_{2^m}(1),\hat{\Delta}_{2^n}(1)\right),\quad\forall n,m\in\mathbb{N}.$$

Therefore, $\{\hat{\Delta}_{2^n}(1)\}_{n\in\mathbb{N}}$ is a Cauchy sequence with respect to $u_{\mathrm{GW},p}$ without limit in $\mathcal{U}^w$. This implies that $(\mathcal{U}^w,u_{\mathrm{GW},p})$ is not complete.

2. By Theorem 3.18 and (1), we have that $\left(\mathcal{U}^w,u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is not separable. As for completeness, consider the subset $X:=\{1-\frac{1}{n}\}_{n\in\mathbb{N}}\subseteq(\mathbb{R}_{\geq 0},\Lambda_\infty)$. By Lemma A.2, X is not a compact ultrametric space. Let $\mu_0\in\mathcal{P}(X)$ be a probability defined as follows:

$$\mu_0\left(\left\{1-\frac{1}{n}\right\}\right):=2^{-n},\quad\forall n\in\mathbb{N}.$$

For each $N\in\mathbb{N}$, let $X_N:=\{1-\frac{1}{n}|\,n=1,\ldots,N\}$. Since each X_N is finite, (X_N,Λ_∞) is a compact ultrametric space. Let $\mu_N\in\mathcal{P}(X_N)$ be a probability defined as follows:

$$\mu_N\left(\left\{1-\frac{1}{n}\right\}\right):=\begin{cases}2^{-n}, & 1\leq n<N\\ 2^{-N+1} & n=N\end{cases}.$$

Then, it is easy to verify (e.g. via Theorem 3.7) that $\{(X_N, \Lambda_\infty, \mu_N)\}_{N\in\mathbb{N}}$ is a $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ Cauchy sequence with $(X, \Lambda_\infty, \mu_0)$ being the limit. Since the set X is not compact, $(X, \Lambda_\infty, \mu_0) \notin \mathcal{U}^w$ and thus $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is not complete.

3. That $(\mathcal{U}^w, u_{\mathrm{GW},\infty})$ is non-separable is already proved in (1). Given a Cauchy sequence $\{\mathcal{X}_n = (X_n, u_n, \mu_n)\}_{n\in\mathbb{N}}$ with respect to $u_{\mathrm{GW},\infty}$, we have that the underlying ultrametric spaces $\{X_n\}_{n\in\mathbb{N}}$ form a Cauchy sequence with respect to u_{GH} due to Corollary 3.15. Since $(\mathcal{U}, u_{\mathrm{GH}})$ is complete (see [105, Proposition 2.1]), there exists a compact ultrametric space (X, u_X) such that

$$\lim_{n\to\infty} u_{\mathrm{GH}}(X_n, X) = 0.$$

For each $n \in \mathbb{N}$, let $\delta_n \coloneqq u_{\mathrm{GH}}(X_n, X)$. By Theorem 2.5, we have that $(X_n)_{\delta_n} \cong X_{\delta_n}$. Denote by $\hat{\mu}_n \in \mathcal{P}(X_{\delta_n})$ the pushforward of $(\mu_n)_{\delta_n}$ under the isometry. Furthermore, we have by Lemma A.7 that X_{δ_n} is finite and we let $X_{\delta_n} = \{[x_1]_{\delta_n}, \ldots, [x_k]_{\delta_n}\}$ for $x_1, \ldots, x_k \in X$. Based on this, we define

$$\nu_n \coloneqq \sum_{i=1}^{k} \hat{\mu}_n([x_i]_{\delta_n}) \cdot \delta_{x_i} \in \mathcal{P}(X),$$

where δ_{x_i} is the Dirac measure at x_i. Since X is compact, $\mathcal{P}(X)$ is weakly compact. Therefore, the sequence $\{\nu_n\}_{n\in\mathbb{N}}$ has a cluster point $\nu \in \mathcal{P}(X)$.

Now we show that $\mathcal{X} \coloneqq (X, u_X, \nu)$ is a $u_{\mathrm{GW},\infty}$ cluster point of $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ and thus the limit of $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ since $\{\mathcal{X}_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. Without loss of generality, we assume that $\{\nu_n\}_{n\in\mathbb{N}}$ weakly converges to ν. Fix any $\varepsilon > 0$, we need to show that $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{X}_n) \leq \varepsilon$ when n is large enough. For any fixed $x_* \in X$, $[x_*]_\varepsilon$ is both an open and closed ball in X. Therefore, $\nu([x_*]_\varepsilon) = \lim_{n\to\infty} \nu_n([x_*]_\varepsilon)$ (see e.g. Billingsley [7, Thm. 2.1]). Since $\delta_n \to 0$ as $n \to \infty$, there exists $N_1 > 0$ such that for any $n > N_1$, $\delta_n < \varepsilon$. We specify an isometry $\varphi_n : (X_n)_{\delta_n} \to X_{\delta_n}$ that gives rise to the construction of ν_n. Then, we let $\psi_n : (X_n)_\varepsilon \to X_\varepsilon$ be the isometry such that the following diagram commutes:

$$\begin{array}{ccc} (X_n)_{\delta_n} & \xrightarrow{\varphi_n} & X_{\delta_n} \\ {\scriptstyle \varepsilon\text{-quotient}}\downarrow & & \downarrow{\scriptstyle \varepsilon\text{-quotient}} \\ (X_n)_\varepsilon & \xrightarrow{\psi_n} & X_\varepsilon \end{array}$$

Assume that $[x_*]_\varepsilon^X = \bigcup_{i=1}^{l} [x_i]_{\delta_n}^X$. Let $x_*^n \in X_n$ be such that $\psi_n([x_*^n]_\varepsilon^{X_n}) = [x_*]_\varepsilon^X$ and let $x_1^n, \ldots, x_l^n \in X_n$ be such that $\varphi_n([x_i^n]_{\delta_n}^{X_n}) = [x_i]_{\delta_n}^X$ for each $i = 1, \ldots, l$. Then,

$[x_*^n]_\varepsilon^{X_n} = \bigcup_{i=1}^l [x_i^n]_{\delta_n}^{X_n}$. Therefore,

$$\nu_n([x_*]_\varepsilon^X) = \sum_{i=1}^l \nu_n([x_i]_{\delta_n}^X) = \sum_{i=1}^l \hat{\mu}_n([x_i]_{\delta_n}^X) = \sum_{i=1}^l \mu_n([x_i^n]_{\delta_n}^{X_n}) = \mu_n([x_*^n]_\varepsilon^{X_n}).$$

Since $\mathcal{X}_n$ is a Cauchy sequence, there exists $N_2 > 0$ such that $u_{\mathrm{GW},\infty}(\mathcal{X}_n, \mathcal{X}_m) < \varepsilon$ when $n, m > N_2$. Then, by Theorem 3.13, $(\mathcal{X}_n)_\varepsilon \cong_w (\mathcal{X}_m)_\varepsilon$ for all $n, m > N_2$. By Lemma A.7, $(X_n)_\varepsilon$ is finite, then $(X_n)_\varepsilon$ has cardinality independent of n when $n > N_2$. For all $n > N_2$, we define the finite set $A_n \coloneqq \left\{\mu_n([x^n]_\varepsilon^{X_n}) | \, x^n \in X_n\right\}$. A_n is independent of n since $(\mathcal{X}_n)_\varepsilon \cong_w (\mathcal{X}_m)_\varepsilon$ for all $n, m > N_2$. This implies that $\mu_n([x_*^n]_\varepsilon^{X_n})$ only takes value in a finite set A_n. Combining with the fact that $\lim_{n\to\infty} \mu_n([x_*^n]_\varepsilon^{X_n}) = \lim_{n\to\infty} \nu_n([x]_\varepsilon^X) = \nu([x_*]_\varepsilon^X)$ exists, there exists $N_3 > 0$ such that when $n > N_3$, $\mu_n([x_*^n]_\varepsilon) \equiv C$ for some constant C. This implies that

$$\nu([x_*]_\varepsilon^X) = \mu_n([x_*^n]_\varepsilon^{X_n}), \qquad \text{when } n > \max(N_1, N_2, N_3).$$

Since X_ε is finite, there exists a common $N > 0$ such that for all $n > N$ and $\forall [x_*]_\varepsilon \in X_\varepsilon$ we have

$$\nu([x_*]_\varepsilon^X) = \mu_n([x_*^n]_\varepsilon^{X_n}),$$

where $[x_*^n]_\varepsilon^{X_n} = \psi_n^{-1}([x_*]_\varepsilon^X) \in (X_n)_\varepsilon$. This indicates that $\nu_\varepsilon = (\psi_n)_\#(\mu_n)_\varepsilon$ when $n > N$. Therefore, $\mathcal{X}_\varepsilon \cong_w (\mathcal{X}_n)_\varepsilon$ and thus $u_{\mathrm{GW},\infty}(\mathcal{X}, \mathcal{X}_n) \le \varepsilon$.

B.4.2 Proof of Proposition 3.27

Next, we will demonstrate Theorem 3.27. However, before we come to this we recall some facts about p-metric and p-geodesic spaces.

Lemma B.15 (Mémoli and Wan [70, Proposition 7.10])**.** *Given $p \in [1, \infty)$, if X is a p-metric space, then X is not q-geodesic for all $1 \le q < p$.*

Lemma B.16 (Mémoli and Wan [70, Theorem 7.7])**.** *Let X be a geodesic metric space. Then, for any $p \ge 1$, $S_{\frac{1}{p}}(X)$ is p-geodesic, where S_α denotes the snowflake transform for $\alpha > 0$ (cf. Section 3.3).*

For $p = 1$, the proof is based on the following property of the 1-Wasserstein space.

Lemma B.17 (Bottou et al. [10, Theorem 5.1])**.** *Let X be a compact metric space. Then, the space $W_1(X) \coloneqq (\mathcal{P}(X), d_{\mathrm{W},1}^X)$ is a geodesic space.*

Based on the above results and Corollary B.2, the proof of Theorem 3.27 is straightforward.

Proof of Theorem 3.27. Let $\mathcal{X}$ and $\mathcal{Y}$ be two compact ultrametric measure spaces. First, we consider the case $p = 1$. By Corollary B.2, there exist a compact ultrametric space Z and isometric embeddings $\phi : X \hookrightarrow Z$ and $\psi : Y \hookrightarrow Z$ such that

$$u_{\mathrm{GW},p}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}) = d_{\mathrm{W},p}^{Z}(\phi_\# \mu_X, \psi_\# \mu_Y).$$

The space $W_1(Z)$ is geodesic (cf. Lemma B.17). Therefore, there exists a Wasserstein geodesic $\tilde{\gamma} : [0,1] \to W_1(Z)$ connecting $\phi_\# \mu_X$ and $\psi_\# \mu_Y$. This induces a curve $\gamma : [0,1] \to \mathcal{U}^w$ where for each $t \in [0,1]$, $\gamma(t) := (\mathrm{supp}(\tilde{\gamma}(t)), u|_{\mathrm{supp}(\tilde{\gamma}(t))\times \mathrm{supp}(\tilde{\gamma}(t))}, \tilde{\gamma}(t))$. Note that $\gamma(0) \cong_w \mathcal{X}$ and $\gamma(1) \cong_w \mathcal{Y}$ and hence we simply replace $\gamma(0)$ and $\gamma(1)$ with $\mathcal{X}$ and $\mathcal{Y}$, respectively. Now, for each $s, t \in [0,1]$, we have that

$$d_{\mathrm{GW},1}^{\mathrm{sturm}}(\gamma(s),\gamma(t)) \le d_{\mathrm{W},1}^{Z}(\tilde{\gamma}(s),\tilde{\gamma}(t)) = |s-t| d_{\mathrm{W},1}^{Z}(\tilde{\gamma}(0),\tilde{\gamma}(1)) = |s-t| d_{\mathrm{GW},1}^{\mathrm{sturm}}(\mathcal{X},\mathcal{Y}).$$

Therefore, γ is a geodesic connecting $\mathcal{X}$ and $\mathcal{Y}$ and thus $(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}})$ is geodesic.

Next, we come to the case $p > 1$. By Corollary B.13, $S_p\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right) \cong \left(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}}\right)$. This implies that $S_{\frac{1}{p}}(\mathcal{U}^w, u_{\mathrm{GW},1}^{\mathrm{sturm}}) \cong \left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$. Hence, by Lemma B.16, we have that $\left(\mathcal{U}^w, u_{\mathrm{GW},p}^{\mathrm{sturm}}\right)$ is p-geodesic. □

B.5 Technical issues from Section 3

In the following, we address various technical issues from Section 3.

B.5.1 The Wasserstein pseudometric

Given a set X, a pseudometric is a symmetric function $d_X : X \times X \to \mathbb{R}_{\ge 0}$ satisfying the triangle inequality and $d_X(x,x) = 0$ for all $x \in X$. Note that if moreover $d_X(x,y) = 0$ implies $x = y$, then d_X is a metric. There is a canonical identification on pseudometric spaces (X, d_X): $x \sim x'$ if $d_X(x,x') = 0$. Then, $\sim$ is in fact an equivalence relation and we define the quotient space $\tilde{X} = X/\sim$. Define a function $\tilde{d}_X : \tilde{X} \times \tilde{X} \to \mathbb{R}_{\ge 0}$ as follows:

$$\tilde{d}_X([x],[x']) := \begin{cases} d_X(x,x') & \text{if } d_X(x,x') \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

$\tilde{d}_X$ turns out to be a metric on $\tilde{X}$. In the following, the metric space $(\tilde{X}, \tilde{d}_X)$ is referred to as the *metric space induced by the pseudometric space* (X, d_X). Note that $\tilde{d}_X$ preserves the induced topology (see e.g. [44]) and thus the quotient map $\Psi : X \to \tilde{X}$ is continuous.

Analogously to the Wasserstein distance, which is defined for probability measures on metric spaces, we define the *Wasserstein pseudometric* for measures on compact pseudometric spaces as done in [94]. Let $\alpha, \beta \in \mathcal{P}(X)$. Then, we define for $p \in [1, \infty)$ the Wasserstein pseudometric of order p as

$$d_{\mathrm{W},p}^{(X,d_X)}(\alpha, \beta) := \left(\inf_{\mu \in \mathcal{C}(\alpha,\beta)} \int_{X \times X} d_X^p(x, y)\, \mu(dx \times dy) \right)^{\frac{1}{p}} \tag{34}$$

and for $p = \infty$ as

$$d_{\mathrm{W},\infty}^{(X,d_X)}(\alpha, \beta) := \inf_{\mu \in \mathcal{C}(\alpha,\beta)} \sup_{(x,y) \in \mathrm{supp}(\mu)} u(x, y). \tag{35}$$

It is easy to see that the Wasserstein pseudometric is closely related to the Wasserstein distance on the induced metric space. More precisely, one can show the following.

Lemma B.18. *Let (X, d_X) denote a compact pseudometric space, let $\alpha, \beta \in \mathcal{P}(X)$. Then, it follows for $p \in [1, \infty]$ that*

$$d_{\mathrm{W},p}^{(X,d_X)}(\alpha, \beta) = d_{\mathrm{W},p}^{(\tilde{X},\tilde{d}_X)}(\Psi_{\#}\alpha, \Psi_{\#}\beta) \tag{36}$$

and in particular that the infimum in Equation (34) (resp. in Equation (35) if $p = \infty$) is attained for some $\mu \in \mathcal{C}(\alpha, \beta)$.

Proof. In the course of this proof we focus on the case $p < \infty$ and remark that the case $p = \infty$ follows by similar arguments. The quotient map allows us to define the map $\theta : \mathcal{C}(\alpha, \beta) \to \mathcal{C}(\Psi_{\#}\alpha, \Psi_{\#}\beta)$ via $\mu \mapsto (\Psi \times \Psi)_{\#}\mu$. It is easy to see that θ is well defined and surjective. Furthermore, it holds by construction that

$$\int_{X \times X} d_X^p(x, y)\, \mu(dx \times dy) = \int_{\tilde{X} \times \tilde{X}} \tilde{d}_X^p(x, y)\, \theta(\mu)(dx \times dy)$$

for all $\mu \in \mathcal{C}(\alpha, \beta)$. Hence, Equation (36) follows.

We come to the second part of the claim. By [100, Sec.4] there exists an optimal coupling $\tilde{\mu}^* \in \mathcal{C}(\Psi_{\#}\alpha, \Psi_{\#}\beta)$ such that

$$d_{\mathrm{W},p}^{(\tilde{X},\tilde{d}_X)}(\Psi_{\#}\alpha, \Psi_{\#}\beta) = \left(\int_{\tilde{X} \times \tilde{X}} \tilde{d}_X^p(x, y)\, \tilde{\mu}^*(dx \times dy) \right)^{\frac{1}{p}}.$$

In consequence, we find using our previous results that for any $\mu^* \in \theta^{-1}(\tilde{\mu}^*)$ it holds

$$\begin{aligned} d_{\mathrm{W},p}^{(\tilde{X},\tilde{d}_X)}(\Psi_{\#}\alpha, \Psi_{\#}\beta) &= \left(\int_{\tilde{X} \times \tilde{X}} \tilde{d}_X^p(x, y)\, \tilde{\mu}^*(dx \times dy) \right)^{\frac{1}{p}} \\ &= \left(\int_{X \times X} d_X^p(x, y)\, \mu^*(dx \times dy) \right)^{\frac{1}{p}} = d_{\mathrm{W},p}^{(X,d_X)}(\alpha, \beta). \end{aligned}$$

This yields the claim. ∎

B.5.2 Regularity of the cost functionals of $u_{\mathrm{GW},p}$ and $u_{\mathrm{GW},p}^{\mathrm{sturm}}$

In the remainder of this section, we collect various technical results required to demonstrate the existence of optimizers in the definitions of $u_{\mathrm{GW},p}^{\mathrm{sturm}}$ (see Equation (10)) and $u_{\mathrm{GW},p}$ (see Equation (13)).

Lemma B.19. *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Then, $\mu \in \mathcal{C}(\mu_X, \mu_Y) \subseteq \mathcal{P}(X \times Y, \max(u_X, u_Y))$ is compact with respect to weak convergence.*

Proof. The proof follows directly from Chowdhury and Mémoli [23, Lemma 2.2]. □

Lemma B.20. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}^w$. Let $D_1 \subseteq \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ be a non-empty subset satisfying the following: there exist $(x_0, y_0) \in X \times Y$ and $C > 0$ such that $u(x_0, y_0) \le C$ for all $u \in D_1$. Then, D_1 is pre-compact with respect to uniform convergence.*

Proof. Let $\{u_n\}_{n\in\mathbb{N}} \subseteq D_1$ be a sequence. Note that $X \times Y \subseteq X \sqcup Y \times X \sqcup Y$. Let $v_n := u_n|_{X\times Y}$. For any $n \in \mathbb{N}$ and any $(x, y), (x', y') \in X \times Y$, we have that

$$|u_n(x, y) - u_n(x', y')| \le u_X(x, x') + u_Y(y, y') \le 2 \max(u_X, u_Y) \big((x, y), (x', y')\big).$$

This means that $\{v_n\}_{n\in\mathbb{N}}$ is equicontinuous with respect to the ultrametric $\max\{u_X, u_Y\}$ on $X \times Y$. Now, since $u_n(x_0, y_0) \le C$, we have that for any $(x, y) \in X \times Y$,

$$u_n(x, y) \le 2 \max(u_X, u_Y)((x, y), (x_0, y_0)) + u_n(x_0, y_0) \le 2 \max(\mathrm{diam}(X), \mathrm{diam}(Y)) + C.$$

Consequently, $\{v_n\}_{n\in\mathbb{N}}$ is uniformly bounded. By the Arzéla-Ascoli theorem ([52, Theorem 7 on page 61]), we have that each subsequence of $\{v_n\}_{n\in\mathbb{N}}$ has a uniformly convergent subsequence. Hence, we can assume without loss of generality that the sequence $\{v_n\}_{n\in\mathbb{N}}$ converges to $v : X \times Y \to \mathbb{R}_{\ge 0}$.

Now, we define $u : X \sqcup Y \times X \sqcup Y \to \mathbb{R}_{\ge 0}$ as follows:

1. $u|_{X\times X} := u_X$ and $u|_{Y\times Y} := u_Y$;
2. $u|_{X\times Y} := v$;
3. for $(y, x) \in Y \times X$, we let $u(y, x) := u(x, y)$.

It is easy to verify that $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ and that u is a cluster point of the sequence $\{u_n\}_{n\in\mathbb{N}}$. Therefore, D_1 is pre-compact. □

Lemma B.21. *Let $\mathcal{X} = (X, u_X, \mu_X)$ and $\mathcal{Y} = (Y, u_Y, \mu_Y)$ be compact ultrametric measure spaces. Let $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathcal{C}(\mu_X, \mu_Y)$ be a sequence weakly converging to $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Let*

$\{u_n\}_{n\in\mathbb{N}} \subseteq \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$. *Suppose that there exist a non-decreasing sequence* $\{p_n\}_{n\in\mathbb{N}} \subseteq [1,\infty)$ *and* $C > 0$ *such that*

$$\left(\int_{X\times Y} (u_n(x,y))^{p_n} \mu_n(dx \times dy)\right)^{\frac{1}{p_n}} \leq C$$

for all $n \in \mathbb{N}$. *Then,* $\{u_n\}_{n\in\mathbb{N}}$ *uniformly converges to some* $u \in \mathcal{D}^{\mathrm{ult}}(u_X, u_Y)$ *(up to taking a subsequence).*

Proof. The following argument adapts the proof of Lemma 3.3 in [92] to the current setting. For any $(x_0, y_0) \in \mathrm{supp}\,(\mu)$, there exist $\varepsilon, \delta > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$

$$\begin{aligned} C &\geq \left(\int_{X\times Y} (u_n(x,y))^{p_n} \mu_n(dx \times dy)\right)^{\frac{1}{p_n}} \geq \int_{X\times Y} u_n(x,y)\mu_n(dx \times dy) \\ &\geq \int_{B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0)} u_n(x,y)\mu_n(dx\times dy) \geq \int_{B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0)} (u_n(x_0,y_0) - 2\varepsilon)\mu_n(dx\times dy) \\ &\geq (u_n(x_0,y_0) - 2\varepsilon)\left(\mu\left(B_\varepsilon^X(x_0)\times B_\varepsilon^Y(y_0)\right) - \delta\right). \end{aligned}$$

Therefore, $\{u_n(x_0,y_0)\}_{n\geq N}$ is uniformly bounded. By Lemma B.20, we have that $\{u_n\}_{n\in\mathbb{N}}$ has a uniformly convergent subsequence. □

Lemma B.22. *Let* X, Y *be ultrametric spaces, then* $\Lambda_\infty(u_X, u_Y) : X\times Y\times X\times Y \to \mathbb{R}_{\geq 0}$ *is continuous with respect to the product topology (induced by* $\max(u_X, u_Y, u_X, u_Y)$*).*

Proof. Fix $(x, y, x', y') \in X \times Y \times X \times Y$ and $\varepsilon > 0$. Choose $0 < \delta < \varepsilon$ such that $\delta < u_X(x, x')$ if $x \neq x'$ and $\delta < u_Y(y, y')$ if $y \neq y'$. Then, consider any point $(x_1, y_1, x_1', y_1') \in X\times Y\times X\times Y$ such that $u_X(x,x_1), u_Y(y,y_1), u_X(x',x_1'), u_Y(y',y_1') \leq \delta$. For $u_X(x_1, x_1')$, we have the following two situations:

1. $x = x'$: $u_X(x_1, x_1') \leq \max(u_X(x_1, x), u_X(x, x_1')) \leq \delta < \varepsilon$;
2. $x \neq x'$: $u_X(x_1, x_1') \leq \max(u_X(x_1, x), u_X(x, x'), u_X(x', x_1')) = u_X(x, x')$. Similarly, $u_X(x, x') \leq u_X(x_1, x_1')$ and thus $u_X(x, x') = u_X(x_1, x_1')$.

Similar result holds for $u_Y(y_1, y_1')$. This leads to four cases for $\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1'))$:

1. $x = x', y = y'$: In this case we have $u_X(x_1, x_1'), u_Y(y_1, y_1') < \varepsilon$. Then,

$$|\Lambda_\infty(u_X(x_1,x_1'), u_Y(y_1,y_1')) - \Lambda_\infty(u_X(x,x'), u_Y(y,y'))| = \Lambda_\infty(u_X(x_1,x_1'), u_Y(y_1,y_1')) \leq \varepsilon;$$

2. $x = x', y \neq y'$: Now $u_X(x_1, x_1') < \varepsilon$ and $u_Y(y_1, y_1') = u_Y(y, y')$. If $u_Y(y, y') \geq \varepsilon > u_X(x_1, x_1')$, then

$$|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))| = |u_Y(y, y') - u_Y(y, y')| = 0.$$

Otherwise $u_Y(y, y') < \varepsilon$, which implies that $\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) \leq \varepsilon$ and $\Lambda_\infty(u_X(x, x'), u_Y(y, y')) = u_Y(y, y') \leq \varepsilon$. Therefore,

$$|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))| \leq \varepsilon;$$

3. $x \neq x', y = y'$: Similar with (2) we have

$$|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))| \leq \varepsilon;$$

4. $x \neq x', y \neq y'$: Now $u_X(x_1, x_1') = u_X(x, x')$ and $u_Y(y_1, y_1') = u_Y(y, y')$. Therefore,

$$|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))| = 0.$$

In conclusion, whenever $u_X(x, x_1), u_Y(y, y_1), u_X(x', x_1'), u_Y(y', y_1') \leq \delta$ we have that

$$|\Lambda_\infty(u_X(x_1, x_1'), u_Y(y_1, y_1')) - \Lambda_\infty(u_X(x, x'), u_Y(y, y'))| \leq \varepsilon.$$

Therefore, $\Lambda_\infty(u_X, u_Y)$ is continuous with respect to the metric $\max(u_X, u_Y, u_X, u_Y)$. □

B.5.3 $u_{\mathrm{GW},p}$ and the one point space

It is possible to explicitly write down $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, in some simple settings. In the following, we derive an explicit formulation of $u_{\mathrm{GW},p}$, $1 \leq p \leq \infty$, between an arbitrary ultrametric measure space $\mathcal{X}$ and the one point ultrametric measure space $*$. For this purpose, we need to introduce some notation. Let $\mathcal{X} = (\mathcal{X}, d_X, \mu_X)$ be a ultrametric measure space. Let its p-diameter (see e.g., [67]) for $1 \leq p < \infty$ be defined as

$$\mathrm{diam}_p(\mathcal{X}) := \left(\iint_{X \times X} \left(d_X(x, x')\right)^p \mu_X(dx)\, \mu_X(dx') \right)^{1/p}$$

and for $p = \infty$ as

$$\mathrm{diam}_\infty(\mathcal{X}) := \sup_{(x,x') \in \mathrm{supp}(\mu_X)} d_X(x, x').$$

Then, one can show the subsequent proposition.

Proposition B.23. *Let $* \in \mathcal{U}^w$ be the one-point space. Then, it holds for any $1 \leq p \leq \infty$ that*

$$u_{\mathrm{GW},p}(\mathcal{X}, *) = \mathrm{diam}_p(\mathcal{X}).$$

Proof. Denote by μ the unique coupling $\mu_X \otimes \delta_*$ between μ_X and δ_*. Then, for any $p < \infty$ we have

$$\begin{aligned} u_{\mathrm{GW},p}(\mathcal{X}, *) &= \left(\iint_{X\times * \times X \times *} \left(\Lambda_\infty(u_X(x,x'), u_*(y,y'))\right)^p \mu(dx \times dy)\, \mu(dx' \times dy') \right)^{1/p} \\ &= \left(\iint_{X\times X} \left(u_X(x,x')\right)^p \mu_X(dx)\, \mu_X(dx') \right)^{1/p} = \mathrm{diam}_p(\mathcal{X}). \end{aligned}$$

The case $p = \infty$ follows by analogous arguments. □

C Missing details from Section 4

C.1 Proofs from Section 4

In the following, we state the full proofs of the results from Section 4.

C.1.1 Proof of Theorem 4.1

We start by proving the first statement. To this end, we observe that for any point x in an ultrametric space X, there always exists a point $x' \in X$ such that $u_X(x,x') = \mathrm{diam}\,(X)$ (see [30]). By assumption μ_X is fully supported on X. Hence, $s_{X,\infty} \equiv \mathrm{diam}\,(X)$ is a constant function. Therefore,

$$\Lambda_\infty(s_{X,\infty}(x), s_{Y,\infty}(y)) \equiv \Lambda_\infty(\mathrm{diam}\,(X), \mathrm{diam}\,(Y)), \quad \forall x \in X, y \in Y.$$

This implies that $\mathbf{FLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = \Lambda_\infty(\mathrm{diam}\,(X), \mathrm{diam}\,(Y))$. By Corollary 5.8 of Mémoli and Wan [70] and Corollary 3.15, we have that

$$u_{\mathrm{GW},\infty}(\mathcal{X},\mathcal{Y}) \geq u_{\mathrm{GH}}(X,Y) \geq \Lambda_\infty(\mathrm{diam}\,(X), \mathrm{diam}\,(Y)) = \mathbf{FLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).$$

It remains to prove the second statement. The proof for $d_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq \mathbf{TLB}_p(\mathcal{X},\mathcal{Y})$ in [67, Sec. 6] can be used essentially without any change for showing $u_{\mathrm{GW},p}(\mathcal{X},\mathcal{Y}) \geq \mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y})$. Hence, it only remains to show that $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y})$, i.e., the claim follows once we have established Proposition C.1.

Proposition C.1. *Let $\mathcal{X},\mathcal{Y} \in \mathcal{U}^w$ and let $p \in [1,\infty]$. Then,*

$$\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) \geq \mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).$$

In order to prove Proposition C.1, we need the following technical lemma.

Lemma C.2. *Let* $\mathcal{X} = (X, d_X, \mu_X) \in \mathcal{U}^w$. *Then,* $\operatorname{spec}(X) := \{u_X(x, x') \mid x, x' \in \mathcal{X}\}$ *is a compact subset of* $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$.

Proof. By Lemma A.7, we have that for each $t > 0$, X_t is a finite set. Let $\{t_n\}_{n=1}^\infty$ be a positive sequence decreasing to 0. Then, it is easy to see that

$$\operatorname{spec}(X) = \bigcup_{n=1}^\infty \operatorname{spec}(X_{t_n}).$$

Since each $\operatorname{spec}(X_{t_n})$ is a finite set, $\operatorname{spec}(X)$ is a countable set.

Now, pick any $0 \neq t \in \operatorname{spec}(X)$. Suppose t is a cluster point in $\operatorname{spec}(X)$. Then, there exists infinitely many $s \in \operatorname{spec}(X)$ greater than $\frac{t}{2}$. However, this will result in $X_{\frac{t}{2}}$ being an infinite set, which contradicts the fact that $X_{\frac{t}{2}}$ is finite. Therefore, 0 is the only possible cluster point of $\operatorname{spec}(X)$. By Lemma A.2, we have that $\operatorname{spec}(X)$ is compact. □

With the above auxiliary result available, we can demonstrate Proposition C.1 and hence finish the proof of Theorem 4.1.

Proof of Proposition C.1. We first prove the case when $p < \infty$. Let $dh_\mathcal{X}(x) := u_X(x, \cdot)_\# \mu_X$ and let $dh_\mathcal{Y}(y) := u_Y(y, \cdot)_\# \mu_Y$. Futher, define $dH_\mathcal{X} := (u_X)_\#(\mu_X \otimes \mu_X)$ and $dH_\mathcal{Y} := (u_Y)_\#(\mu_Y \otimes \mu_Y)$. Lemma C.2 implies that the set $S := \operatorname{spec}(X) \cup \operatorname{spec}(Y)$ is a compact subset of $(\mathbb{R}_{\geq 0}, \Lambda_\infty)$. It is easy to see that $\operatorname{supp}(dh_\mathcal{X}), \operatorname{supp}(dh_\mathcal{Y}), \operatorname{supp}(dH_\mathcal{X}), \operatorname{supp}(dH_\mathcal{Y}) \subseteq S \subseteq \mathbb{R}_{\geq 0}$. Now, recall that by Proposition 4.4

$$\mathbf{SLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_\mathcal{X}, dH_\mathcal{Y})$$

and

$$\mathbf{TLB}_p^{\text{ult}}(\mathcal{X}, \mathcal{Y}) = \left(\inf_{\pi \in \mathcal{C}(\mu_X, \mu_Y)} \int_{X \times Y} \left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_\mathcal{X}(x), dh_\mathcal{Y}(y)) \right)^p \mu(dx \times dy) \right)^{1/p}.$$

Further, we observe for any $x \in X$ and $y \in Y$ that

$$d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_\mathcal{X}(x), dh_\mathcal{Y}(y)) = \inf_{\pi_{xy} \in \mathcal{C}(dh_\mathcal{X}(x), dh_\mathcal{Y}(y))} \left(\int_{S \times S} \Lambda_\infty^p(s, t)\, \pi_{xy}(ds \times dt) \right)^{\frac{1}{p}}.$$

For the remainder of this proof, the metric on metric on $S \subseteq \mathbb{R}_{\geq 0}$ is always given by Λ_∞. Additionally, $\mathcal{P}(S)$ denotes the set of probability measures on S and we equip $\mathcal{P}(S)$ with the Borel σ-field with respect to the topology induced by weak convergence.

Claim 1: There is a measurable choice $(x, y) \mapsto \pi_{xy}^*$ such that for each $(x, y) \in X \times Y$, $\pi_{x,y}^*$ is an optimal transport plan between $dh_\mathcal{X}(x)$ and $dh_\mathcal{Y}(y)$.

Proof of Claim 1. It is easy to see that both Λ_1 and Λ_∞ induce the same topology and thus Borel sets on S. This therefore implies that $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_1)}$ and $d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0},\Lambda_\infty)}$ metrize the same weak topology on $\mathcal{P}(S)$. By Mémoli and Needham [68, Remark 2.5], the following two maps are continuous with respect to the weak topology and thus measurable:

$$\Phi_1 : X \to \mathcal{P}(S),\ x \mapsto dh_{\mathcal{X}}(x)$$

and

$$\Phi_2 : Y \to \mathcal{P}(S),\ y \mapsto dh_{\mathcal{Y}}(y).$$

Since S is a compact space, the space $\left(\mathcal{P}(S), d_{\mathrm{W},p}^{(S,\Lambda_\infty)}\right)$ is separable [100, Theorem 6.18]. This yields that $\mathscr{B}(\mathcal{P}(S) \times \mathcal{P}(S)) = \mathscr{B}(\mathcal{P}(S)) \otimes \mathscr{B}(\mathcal{P}(S))$ [36, Proposition 1.5]. Hence, the product Φ of Φ_1 and Φ_2, defined by

$$\Phi : X \times Y \to \mathcal{P}(S) \times \mathcal{P}(S),\ (x, y) \mapsto (dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))$$

is measurable [36, Proposition 2.4]. Since Φ is measurable, a direct application of Villani [100, Corollary 5.22] gives the claim. $\square$

Now, we have that for every $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ that

$$\begin{aligned}&\int_{X\times Y} \left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x), dh_{\mathcal{Y}}(y))\right)^p \mu(dx \times dy)\\ &= \int_{X\times Y}\int_{S\times S} \Lambda_\infty^p(s,t)\, \pi_{xy}^*(ds \times dt)\, \mu(dx \times dy)\\ &= \int_{S\times S} \Lambda_\infty^p(s,t)\, \bar{\mu}(ds \times dt),\end{aligned}$$

by Fubini's Theorem, where $\bar{\mu} \in \mathcal{P}(S \times S)$ is defined as

$$\bar{\mu}(A) := \int_{X\times Y} \pi_{xy}^*(A)\, \mu(dx \times dy) \tag{37}$$

for measurable $A \subseteq S \times S$. We remark that by Claim 1 the measure $\bar{\mu}$ in Equation (37) is well defined. Next, we verify that $\bar{\mu} \in \mathcal{C}(dH_{\mathcal{X}}, dH_{\mathcal{Y}})$. For any measurable $A \subseteq (S, \Lambda_\infty)$ we have

$$\begin{aligned}\bar{\mu}(A\times S) &= \int_{X\times Y}\pi^*_{x,y}(A\times S)\,\mu(dx\times dy)\\ &= \int_{X\times Y} dh_{\mathcal{X}}(x)(A)\,\mu(dx\times dy)\\ &= \int_X dh_{\mathcal{X}}(x)(A)\,\mu_X(dx)\\ &\overset{(i)}{=}\int_X\int_X \mathbb{1}_{\{d_X(x,x')\in A\}}\,\mu_X(dx')\,\mu_X(dx)\\ &= dH_{\mathcal{X}}(A),\end{aligned}$$

where we have applied the marginal constraints for π_{xy} and μ. Further, (i) follows by the change-of-variables formula. The analogous arguments give that

$$\bar{\mu}(S\times B) = dH_{\mathcal{Y}}(B),$$

for any measurable $B\subseteq S$. Thus, we conclude that for every $\mu\in\mathcal{C}(\mu_X,\mu_Y)$

$$\begin{aligned}\int_{X\times Y}\left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(x),dh_{\mathcal{Y}}(y))\right)^p\,\mu(dx\times dy) &= \int_{S\times S}\Lambda_\infty^p(s,t)\,\bar{\mu}(ds\times dt)\\ &\geq \inf_{\pi\in\mathcal{C}(dH_{\mathcal{X}},dH_{\mathcal{Y}})}\int_{S\times S}\Lambda_\infty(s,t)\,\pi(ds\times dt)\\ &= \left(d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}},dH_{\mathcal{Y}})\right)^p.\end{aligned}$$

This gives the claim for $p<\infty$.

Next, we prove the assertion for the case $p=\infty$. Note that for any $p<\infty$

$$\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)}\left\|d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(\cdot),dh_{\mathcal{Y}}(\cdot))\right\|_{L^p(\mu)} \tag{38}$$

$$\leq \inf_{\mu\in\mathcal{C}(\mu_X,\mu_Y)}\left\|d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}(dh_{\mathcal{X}}(\cdot),dh_{\mathcal{Y}}(\cdot))\right\|_{L^\infty(\mu)} \tag{39}$$

$$= \mathbf{TLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}), \tag{40}$$

where the inequality holds since $d_{\mathrm{W},p}^{(S,\Lambda_\infty)}\leq d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}$ and $\|\cdot\|_{L^p(\mu)}\leq\|\cdot\|_{L^\infty(\mu)}$.

By Givens and Shortt [38, Proposition 3] we have that

$$\mathbf{SLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = d_{\mathrm{W},\infty}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}},dH_{\mathcal{Y}}) = \lim_{p\to\infty} d_{\mathrm{W},p}^{(S,\Lambda_\infty)}(dH_{\mathcal{X}},dH_{\mathcal{Y}}) = \lim_{p\to\infty}\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).$$

Therefore,

$$\mathbf{SLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) = \lim_{p\to\infty}\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) \leq \limsup_{p\to\infty}\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}) \leq \mathbf{TLB}_\infty^{\mathrm{ult}}(\mathcal{X},\mathcal{Y}).$$

□

C.1.2 Proof of Proposition 4.4

We only prove the first statement for $p \in [1, \infty)$. The case $p = \infty$ as well as the second statement can be proven in a similar manner.

By directly using the change-of-variables formula, we have the following:

$$
\begin{aligned}
&\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) \\
=& \inf_{\gamma \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \int_{X \times X \times Y \times Y} \left(\Lambda_\infty\left(u_X(x, x'), u_Y(y, y')\right)\right)^p \gamma(d(x, x') \times d(y, y')) \\
=& \inf_{\gamma \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \int_{\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}} \left(\Lambda_\infty(s, t)\right)^p (u_X \times u_Y)_\# \gamma(ds \times dt),
\end{aligned}
$$

where $u_X \times u_Y : X \times X \times Y \times Y \to \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ maps (x, x', y, y') to $(u_X(x, x'), u_Y(y, y'))$. By Lemma A.5, we have that

$$
(u_X \times u_Y)_\# \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y) = \mathcal{C}\left((u_X)_\#(\mu_X \otimes \mu_X), (u_Y)_\#(\mu_Y \otimes \mu_Y)\right).
$$

Therefore,

$$
\begin{aligned}
\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}, \mathcal{Y}) =& \inf_{\gamma \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \int_{\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}} \left(\Lambda_\infty(s, t)\right)^p (u_X \times u_Y)_\# \gamma(ds \times dt) \\
=& \inf_{\tilde{\gamma} \in \mathcal{C}\left((u_X)_\#(\mu_X \otimes \mu_X), (u_Y)_\#(\mu_Y \otimes \mu_Y)\right)} \int_{\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}} \left(\Lambda_\infty(s, t)\right)^p \tilde{\gamma}(ds \times dt) \\
=& d_{\mathrm{W},p}^{(\mathbb{R}_{\geq 0}, \Lambda_\infty)}((u_X)_\#(\mu_X \otimes \mu_X), (u_Y)_\#(\mu_Y \otimes \mu_Y)).
\end{aligned}
$$

C.1.3 The relation between $\mathbf{SLB}^{\mathrm{ult}}$ and $\mathbf{TLB}^{\mathrm{ult}}$

Next, we will demonstrate that there are ultrametric measure spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ such that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}_1, \mathcal{X}_2) = 0$, while it holds $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}_1, \mathcal{X}_2) > 0$. To this end, consider the three point space $\Delta_3(1) = (\{x_1, x_2, x_3\}, u)$ where $u(x_i, x_j) = 1$ whenever $i \neq j$. Let $\mu_1 := \frac{2}{3}\delta_{x_1} + \frac{1}{6}\delta_{x_2} + \frac{1}{6}\delta_{x_3}$ and let $\mu_2 := \frac{1}{3}\delta_{x_1} + \left(\frac{1}{3} - \frac{1}{2\sqrt{3}}\right)\delta_{x_2} + \left(\frac{1}{3} + \frac{1}{2\sqrt{3}}\right)\delta_{x_3}$. Both μ_1 and μ_2 are probability measures on $\Delta_3(1)$. We then let $\mathcal{X}_1 := (\Delta_3(1), \mu_1)$ and $\mathcal{X}_2 := (\Delta_3(1), \mu_2)$. It is easy to check that

$$
u_\#(\mu_1 \otimes \mu_1) = u_\#(\mu_2 \otimes \mu_2) = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.
$$

Then, by Proposition 4.4 we immediately have that $\mathbf{SLB}_p^{\mathrm{ult}}(\mathcal{X}_1, \mathcal{X}_2) = 0$ for any $p \in [1, \infty]$. Now, note that

$$
u(x_1, \cdot)_\# \mu_1 = \frac{2}{3}\delta_0 + \frac{1}{3}\delta_1,
$$

which is obviously different from all $u(x_i, \cdot)_{\#}\mu_2$ for $i = 1, 2, 3$. This implies (by Proposition 4.4) that we have $\mathbf{TLB}_p^{\mathrm{ult}}(\mathcal{X}_1, \mathcal{X}_2) > 0$ for any $p \in [1, \infty]$.

In fact, this example works as well for showing that $\mathbf{TLB}_p(\mathcal{X}_1, \mathcal{X}_2) > \mathbf{SLB}_p(\mathcal{X}_1, \mathcal{X}_2) = 0$.

D Missing details from Section 5

D.1 Proofs from Section 5

Next, we give the complete proofs of the results stated in Section 5.

D.1.1 Proof of Theorem 5.6

The first step to prove this is to verify the existence of an optimal coupling. To this end, we make the following obvious observation.

Lemma D.1. *Let X, Y be finite ultra-dissimilarity spaces, then $\Lambda_\infty(u_X, u_Y) : X \times Y \times X \times Y \to \mathbb{R}_{\geq 0}$ is continuous with respect to the discrete topology.*

This allows us to verify the subsequent analogue to Proposition B.10.

Proposition D.2. *Let $\mathcal{X}, \mathcal{Y} \in \mathcal{U}_{\mathrm{dis}}^w$. Then, for any $p \in [1, \infty]$, there always exists an optimal coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu)$.*

Proof. The proof is essentially the same as the one for Proposition B.10. We only replace Lemma B.22 with Lemma D.1. The details are left to the reader. □

With Proposition D.2 available and Theorem 3.10 already proven, it is immediately clear how to verify the symmetry and the p-triangle inequality for $u_{\mathrm{GW},p}$ on $\mathcal{U}_{\mathrm{dis}}^w$. Hence it only remains to demonstrate identity of indiscernibles.

Proof of Theorem 5.6. Due to the similarity between Theorem 5.6 and Theorem 3.10, we only verify that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$ if and only if $\mathcal{X} \cong_w \mathcal{Y}$. If $\mathcal{X} \cong_w \mathcal{Y}$, then obviously $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$.

Next, we assume that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = 0$. By Proposition D.2 there exists $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ such that $u_{\mathrm{GW},p}(\mathcal{X}, \mathcal{Y}) = \mathrm{dis}_p^{\mathrm{ult}}(\mu) = 0$. Now, we define a map $\varphi : X \to Y$ as follows: For any $x \in X$ we have $\mu_X(\{x\}) > 0$, since μ_X has full support and X is finite. As a result, $\mu(\{(x, y)\}) > 0$ for some $y \in Y$, then we let $\varphi(x) \mapsto y$. This map is well-defined. Indeed, if there are $x \in X$ and $y, y' \in Y$ such that $\mu(\{(x,y)\}), \mu(\{(x,y')\}) > 0$, then by $\mathrm{dis}_p^{\mathrm{ult}}(\mu) = 0$ we must have that

$$\Delta_\infty\left(u_X(x,x), u_Y(y,y')\right) = \Delta_\infty\left(u_X(x,x), u_Y(y,y)\right) = \Delta_\infty\left(u_X(x,x), u_Y(y',y')\right) = 0.$$

This implies that $u_Y(y,y') = u_Y(y,y) = u_Y(y',y') = u_X(x,x)$. Since u_Y is an ultra-dissimilarity, we have that $y = y'$ (cf. condition (3) in Definition 5.1). Essentially the same argument gives that $\varphi : X \to Y$ is an injective map. As $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ and φ is injective, it follows $\mu_X(\{x\}) = \mu(\{(x,\varphi(x))\}) \leq \mu_Y(\{\varphi(x)\})$ for any $x \in X$. Since

$$1 = \sum_{x \in X} \mu_X(\{x\}) \leq \sum_{x \in X} \mu_Y(\{\varphi(x)\}) \leq 1,$$

we have that $\mu_X(\{x\}) = \mu_Y(\{\varphi(x)\})$ for all $x \in X$. Since μ_Y is fully supported, this implies that φ is a bijective measure preserving map. Now, for any $x, x' \in X$, $\mathrm{dis}_p^{\mathrm{ult}}(\mu) = 0$ implies that $\Delta_\infty(u_X(x,x'), u_Y(\varphi(x),\varphi(x'))) = 0$ and thus $u_X(x,x') = u_Y(\varphi(x),\varphi(x'))$. Therefore, φ is also an isometry and thus an isomorphism. In consequence, $\mathcal{X} \cong_w \mathcal{Y}$. □

E Missing details from Section 6

E.1 Missing details from Section 6.2

Here, we list the precise results for the comparisons of the spaces $\mathcal{X}_i$, $1 \leq i \leq 4$, illustrated in Figure 6. They are gathered in Table 2 and Table 3.

Tab. 2: Comparison of different ultrametric measure spaces I: The values of $u_{\mathrm{GW},1}(\mathcal{X}_i, \mathcal{X}_j)$ (approximated by Algorithm 1) and $u_{\mathrm{GW},\infty}(\mathcal{X}_i, \mathcal{X}_j)$, $1 \leq i \leq j \leq 4$, where $\mathcal{X}_i$, $1 \leq i \leq 4$, denote the ultrametric measure spaces displayed in Figure 6.

	$u_{\mathrm{GW},1}$				$u_{\mathrm{GW},\infty}$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.9333	0.2444	0.7071	0.0000	2.1000	1.1000	2.000
$\mathcal{X}_2$	0.9333	0.0000	1.1778	1.5107	2.1000	0.0000	2.1000	2.1000
$\mathcal{X}_3$	0.2444	1.1778	0.0000	0.4493	1.1000	2.1000	0.0000	2.0000
$\mathcal{X}_4$	0.7071	1.5107	0.4493	0.0000	2.0000	2.1000	2.0000	0.0000

Tab. 3: Comparison of different ultrametric measure spaces II: The values of $\mathbf{SLB}_1^{\mathrm{ult}}(\mathcal{X}_i,\mathcal{X}_j)$, $1 \le i \le j \le 4$, where $\mathcal{X}_i$, $1 \le i \le 4$, denote the ultrametric measure spaces displayed in Figure 6.

	$\mathbf{SLB}_1^{\mathrm{ult}}$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.9333	0.2444	0.0778
$\mathcal{X}_2$	0.9333	0.0000	1.1778	1.4522
$\mathcal{X}_3$	0.2444	1.1778	0.0000	0.2764
$\mathcal{X}_4$	0.0778	1.5107	0.2764	0.0000

E.2 Missing details from Section 6.3

Here, we state more results for the comparison of the ultrametric measure spaces illustrated in Figure 6 and give the precise construction of the ultrametric spaces $Z_{k,t}^i$, $2 \le k \le 5$, $t = 0, 0.2, 0.4, 0.4$, $1 \le i \le 15$.

The ultrametric measure spaces from Figure 6 First, we give the precise results for comparing the ultametric dissimilarity spaces in Figure 6 based on $d_{\mathrm{GW},1}$ and $\mathbf{SLB}_1$. They are gathered in Table 4.

Tab. 4: Comparison of different ultrametric measure spaces III: The values of $d_{\mathrm{GW},1}(\mathcal{X}_i,\mathcal{X}_j)$ (approximated by Algorithm 1) and $\mathbf{SLB}_1(\mathcal{X}_i,\mathcal{X}_j)$, $1 \le i \le j \le 4$, where $(X_i, d_{X_i}, \mu_{X_i})$, $1 \le i \le 4$, denote the ultrametric measure spaces displayed in Figure 6.

	$d_{\mathrm{GW},1}$				$\mathbf{SLB}_1$			
	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$	$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	$\mathcal{X}_4$
$\mathcal{X}_1$	0.0000	0.0444	0.0222	0.2111	0.0000	0.0444	0.0222	0.0422
$\mathcal{X}_2$	0.0444	0.0000	0.0667	0.2556	0.0444	0.0000	0.0667	0.0867
$\mathcal{X}_3$	0.0222	0.0667	0.0000	0.2253	0.0222	0.0667	0.0000	0.0573
$\mathcal{X}_4$	0.2111	0.2556	0.2253	0.0000	0.0422	0.0867	0.0573	0.0000

Perturbations at level t Next, we give the precise construction of the ultrametric measure spaces $Z_{k,t}^i$, $2 \le k \le 5$, $t = 0, 0.2, 0.4, 0.4$, $1 \le i \le 15$. For each $k = 2, 3, 4, 5$ we first draw a sample with $100 \times k$ points from the mixture distribution

$$\sum_{i=0}^{k} \frac{1}{k} U[1.5(k-1), 1.5(k-1)+1],$$

where $U[a,b]$ denotes the uniform distribution on $[a,b]$. For each sample, we employ the single linkage algorithm to create a dendrogram, which then induces an ultrametric on the given sample. We further draw a 30-point subspace from each ultrametric space and denote it by Z_k. These four spaces have similar diameter values between 0.5 and 0.6. Each space Z_k is equipped with the uniform probability measure and the resulting ultrametric measure spaces are denoted by $\mathcal{Z}_k = (Z_k, u_{Z_k}, \mu_{Z_k})$, $k = 2, 3, 4, 5$. We remark that k can be regarded as the number of blocks in the dendrogram representation of the obtained ultrametric measure spaces (see the top row of Figure 7 for a visualization of three 3-block spaces).

CHAPTER C

From Small Scales to Large Scales: Distance-to-Measure Density based Geometric Analysis of Complex Data

From Small Scales to Large Scales: Distance-to-Measure Density based Geometric Analysis of Complex Data

Katharina Proksch * Christoph Alexander Weitkamp† Thomas Staudt †
Benoît Lelandais ‡ Christophe Zimmer ‡

May 24, 2022

Abstract

How can we tell complex point clouds with different small scale characteristics apart, while disregarding global features? Can we find a suitable transformation of such data in a way that allows to discriminate between differences in this sense with statistical guarantees?

In this paper, we consider the analysis and classification of complex point clouds as they are obtained, e.g., via single molecule localization microscopy. We focus on the task of identifying differences between noisy point clouds based on small scale characteristics, while disregarding large scale information such as overall size. We propose an approach based on a transformation of the data via the so-called Distance-to-Measure (DTM) function, a transformation which is based on the average of nearest neighbor distances. For each data set, we estimate the probability density of average local distances of all data points and use the estimated densities for classification. While the applicability is immediate and the practical performance of the proposed methodology is very good, the theoretical study of the density estimators is quite challenging, as they are based on *non-i.i.d.* observations that have been obtained via a complicated transformation. In fact, the transformed data are stochastically dependent in a non-local way that is not captured by commonly considered dependence measures. Nonetheless, we show that the asymptotic behaviour of the density estimator is driven by a kernel density estimator of certain i.i.d. random variables by using theoretical properties of U-statistics, which allows to handle the dependencies via a Hoeffding decomposition. We show via a numerical study and in an application to simulated single molecule localization microscopy data of chromatin fibers that unsupervised classification tasks based on estimated DTM-densities achieve excellent separation results.

*Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Hallenweg 19, 7522NH Enschede
†Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen
‡Institut Pasteur, Imaging and Modeling Unit, UMR 3691, CNRS, Paris

Keywords Geometric data analysis, Distance-to-Measure signature, kernel density estimators, nearest neighbor distributions

1 Introduction

The analysis and extraction of information from complex point clouds has become a main task in many applications. Prominent examples can be found in geomorphology, where structure in point-clouds obtained from laser scanners is investigated to infer on the shape of the Earth [27, 54], or in cosmology, where the Cosmic Web is analysed based on a discrete set of points from N-body simulations or galaxy studies [32]. Related questions also arise in biology, when data from single molecule localization microscopy (SMLM), which is based on the localization of fluorescent molecules that appear at different times, are analyzed [31, 42]. Data obtained in SMLM are 2D or 3D point clouds, where the points correspond to particular molecular localization events. In this paper, we consider a specific example which is related to the analysis of super-resolution visualization of human chromosomal regions as it has recently been investigated in Hao et al. [26]. In this application, the goal is to better understand the 3D organization of the chromatin fiber in cell nuclei, which plays a key role in the regulation of gene expression.
In all aforementioned examples, it is important to identify significant differences between noisy point clouds, where a focus is on general structure and small scale information rather than on global features such as the overall shape of a point cloud.

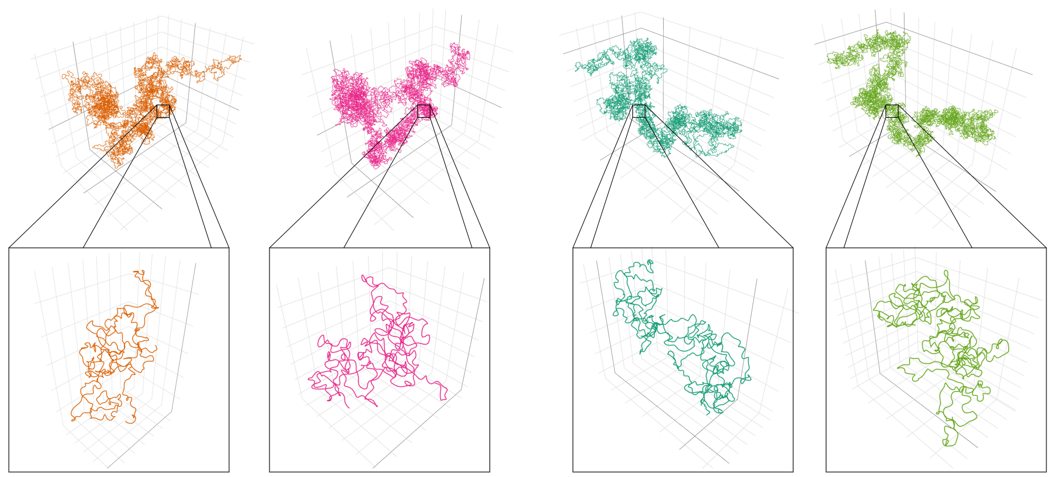

Fig. 1: Example Data: Four different simulated chromatin fibers in two different conditions: Condition A (orange (far left) and blue-green (middle right)) and Condition B (pink (middle left) and green (far right)) for the purpose of comparison.

For illustration, Figure 1 shows four simulated chromatin fibers in two different conditions. The displayed structures form loops of different sizes and frequencies, based on the condition under which they were simulated, where the differences are very subtle. In the application considered in this paper, we analyse noisy samples of such simulated structures. The noise accounts for localization errors as they are present in real SMLM data. The loops are of sizes comparable to the resolution of the images (see Section 4 and Hao et al. [26] for more details), which makes the problem tractable but difficult. The aim is to classify the point clouds based on their loop distribution (i.e. based on their small scale characteristics), while disregarding their total size or large scale shapes. It is natural to transform such complicated data prior to the analysis, in particular when one has a clear objective in mind. In the above reference, the statistical analysis of the simulated and real data was based on a transformation of each data cloud onto a set of two parameters, capturing smoothness and local curvature of the point clouds. While this transformation provided a clear discrimination between different groups, the amount of information preserved in a two-dimensional parameter is not sufficient as a basis for point-by-point classification. In this paper, we propose an approach which is similar in spirit, but which provides a transformation into a curve, with different characteristics for the different conditions. In our analysis, the whole curves are then used as features. To this end, we perform the following two steps.

(i) A transformation of the data points in a point cloud based on the *Distance-to-Measure (DTM) signature* [13, 14] to a one-dimensional data set,

(ii) The analysis of the distribution of the DTM-transformed data via their estimated probability density.

The DTM signature is closely related to certain nearest neighbor distributions, which makes this approach very intuitive. In particular, this framework allows for a comprehensive exploratory analysis of complex data, for which we might seek a simple graphical representation that captures and summarizes the local structural information well.

1.1 The DTM-Density as a Representation for Local Features

We now introduce the statistical framework of the paper and carefully define the previously mentioned DTM-signature. Throughout the following, we consider random point clouds as samples from a *Euclidean metric measure space* $\mathcal{X} = (\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$, i.e., a triple, where $\mathcal{X} \subset \mathbb{R}^d$ denotes a compact set, $||\cdot||$ stands for the Euclidean distance and $\mu_{\mathcal{X}}$ denotes a probability measure that is fully supported on the compact set $\mathcal{X}$. If, additionally, $\mu_{\mathcal{X}}$ has a Lipschitz continuous density with respect to the d-dimensional Lebesgue measure, then we call $\mathcal{X}$ a *regular Euclidean metric measure space*. For each metric measure space $\mathcal{X}$, we can define the corresponding *Distance-to-Measure (DTM) function* with *mass parameter*

$m \in (0,1]$ for $x \in \mathbb{R}^d$ as

$$\mathrm{d}^2_{\mathcal{X},m}(x) = \frac{1}{m}\int_0^m F_x^{-1}(u)\,du, \tag{1}$$

where $F_x(t) = P(\|X-x\|^2 \leq t)$, $X \sim \mu_{\mathcal{X}}$, and F_x^{-1} denotes the corresponding quantile function. The DTM function, which is essential for the definition of the DTM-signature, is a population quantity that is generally unknown in practice and thus has to be estimated from the data. In order to do so, we replace the quantile function in the definition (1) by its empirical version as follows. Let $X_1 \ldots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and denote the corresponding empirical measure by $\widehat{\mu}_{\mathcal{X}}$. We define for $t \geq 0$

$$\widehat{F}_{x,n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{||x-X_i||^2 \leq t\}} \tag{2}$$

and denote by $\widehat{F}_{x,n}^{-1}$ the corresponding quantile function, giving rise to a plug-in estimator for the Distance-to-Measure function $\mathrm{d}^2_{\mathcal{X},m}(x)$:

$$\delta^2_{\mathcal{X},m}(x) = \frac{1}{m}\int_0^m \widehat{F}_{x,n}^{-1}(u)\,du. \tag{3}$$

In the special case that $m = \frac{k}{n}$, it is possible to rewrite (3) as a nearest neighbor statistic as follows

$$\delta^2_{\mathcal{X},m}(x) = \frac{1}{k}\sum_{X_i \in N_k(x)} ||X_i - x||^2, \tag{4}$$

where $N_k(x)$ is the set containing the k nearest neighbors of x among the data points $X_1, \ldots, X_n$.

As discussed previously, we require a good descriptor for the small scale behavior of our data. Hence, in a similar spirit as Brécheteau [7], we reduce the potentially complex Euclidean metric measure space to a one-dimensional probability distribution by considering the *Distance-to-Measure (DTM) signature* $\mathrm{d}^2_{\mathcal{X},m}(X)$, where $X \sim \mu_{\mathcal{X}}$. That is, the deterministic point $x \in \mathcal{X}$ is replaced by the random variable X. The distribution of $\mathrm{d}^2_{\mathcal{X},m}(X)$ captures the relative frequency of the mean of the distances of a random point in $\mathcal{X}$ to its "$m \cdot 100\%$ nearest neighbors". We will empirically illustrate that the distribution of $\mathrm{d}^2_{\mathcal{X},m}(X)$ is a good descriptor for the small scale behavior of the considered data for small values of m and verify that it is well-suited for chromatin loop analysis. Furthermore, it is easy to see that for $m = 1$ the random quantity $\mathrm{d}^2_{\mathcal{X},1}(X)$ is closely related to the lower bound FLB_p of the Gromov-Wasserstein distance defined in Mémoli [37] and is well suited for object discrimination with a focus on large scale characteristics. Although this case is not of interest in our specific data example, we include it in our analysis, since variants of

$\mathrm{d}^2_{\mathcal{X},1}(X)$ have been proven very useful for pose invariant object discrimination [22, 25].

Since we propose to reduce (possibly complex) multi-dimensional metric measure spaces to a one-dimensional probability distribution, the next step is to visualize and investigate these distributions. It is well known that probability densities (if they exist) usually provide a useful visual insight into the probability distributions considered. In this regard, they are usually better suited than cumulative distribution functions (see, e.g., Chen and Pokojovy [15]). Therefore, we focus on the estimation of the density of $\mathrm{d}^2_{\mathcal{X},m}(X)$ in this paper. A natural estimator for the density of $\mathrm{d}^2_{\mathcal{X},m}(X)$, in the following denoted as *DTM-density*, in case of a known DTM-function, is given by

$$\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right). \tag{5}$$

However, since $\mu_{\mathcal{X}}$ is usually unknown, we cannot calculate $\mathrm{d}^2_{\mathcal{X},m}$ and consequently it is generally not feasible to estimate $f_{\mathrm{d}^2_{\mathcal{X},m}}$ via $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$. Instead, we propose to replace $\mathrm{d}^2_{\mathcal{X},m}$ by its empirical version $\delta^2_{\mathcal{X},m}$ and estimate $f_{\mathrm{d}^2_{\mathcal{X},m}}$ based on the plug-in estimator

$$\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\delta^2_{\mathcal{X},m}(X_i) - y}{h}\right). \tag{6}$$

It is important to note that, in contrast to $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$, the plug-in estimator $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ is based on the *non-i.i.d.* observations $\delta^2_{\mathcal{X},m}(X_1), \dots, \delta^2_{\mathcal{X},m}(X_n)$. In fact, for each $i \neq j$, $\delta^2_{\mathcal{X},m}(X_i)$ and $\delta^2_{\mathcal{X},m}(X_j)$ are stochastically dependent. The asymptotic behaviour of kernel density estimators under dependence has been studied extensively in the literature for various mixing and linear processes connected to weakly dependent time series [10, 34, 35, 45, 56]. In all these settings, results on asymptotic normality similar to the i.i.d. case can be derived. Related results for spatial processes can be found, e.g., in Hallin et al. [24]. For *long-range dependent* data, the asymptotic behaviour of kernel density estimators changes drastically. Here, the empirical density process (based on kernel estimators of the marginal densities) converges weakly to a tight limit (see Csorgo and Mielniczuk [16]). For the sequence $\delta^2_{\mathcal{X},m}(X_1), \dots, \delta^2_{\mathcal{X},m}(X_n)$, however, a structure as in the above examples (in space or time) is not given. For each $i \neq j$, $\delta^2_{\mathcal{X},m}(X_i)$ and $\delta^2_{\mathcal{X},m}(X_j)$ are stochastically dependent in a way that is not captured by the dependency models considered in the literature discussed above.

1.2 Main Results

The main theoretical contribution of the paper is the distributional limit of the kernel density estimator defined in (6). More precisely, we prove (cf. Theorem 2.12), given certain regularity conditions on $f_{\mathrm{d}^2_{\mathcal{X},m}}$, $\mathrm{d}^2_{\mathcal{X},m}(y)$ and $\mathcal{X}$, (see Condition 2.2 in Section 2.1) that for $n \to \infty$, $h = o\left(n^{-1/5}\right)$ and $nh \to \infty$

$$\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\int K^2(u)\,du\right). \tag{7}$$

This means that, although the kernel density estimator $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ is based on transformed, dependent random variables, asymptotically, it behaves precisely as the inaccessible kernel density estimator $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$ based on independent random variables. This entails that many methods which are feasible for kernel density estimators based on i.i.d. data, can be applied in this much more complex setting as well, with the same asymptotic justification.

1.3 Application

Chromosomes, which consist of chromatin fibres, are essential parts of cell nuclei in human beings and carry the genetic information important for heredity transmission. It is known by now that there are small scale self-interacting genomic regions, so called topologically associating domains (TADs) which are often associated with loops in the chromatin fibers [43]. As an application, we consider chromatin loop analysis, one aspect of which is to study the presence or absence of loops in the chromatin (see Section 4).
The local loop structure is very well characterized by local nearest neighbor means as illustrated on the right of Figure 2 and hence we propose to use DTM-signatures for tackling this issue. Figure 2 shows the pipeline for the data transformation (left) and the resulting kernel density estimators ($m = 1/250$, biweight kernel, bandwidth selection as in Section 4) for the four data sets shown in Figure 1 (right, same coloring). It shows that the kernel density estimators mainly differ between the different conditions and not between the corresponding chromatin fibers and that the differences between the conditions are clearly pronounced. This demonstrates that the transformation allows for a qualitative analysis of the data.

1.4 Related Work

The use of the DTM-signature for the purpose of pose invariant object discrimination was proposed by Brécheteau [7], who in particular established a relation between the DTM-signature and the Gromov-Wasserstein distance (see Mémoli [37] for a definition). In the

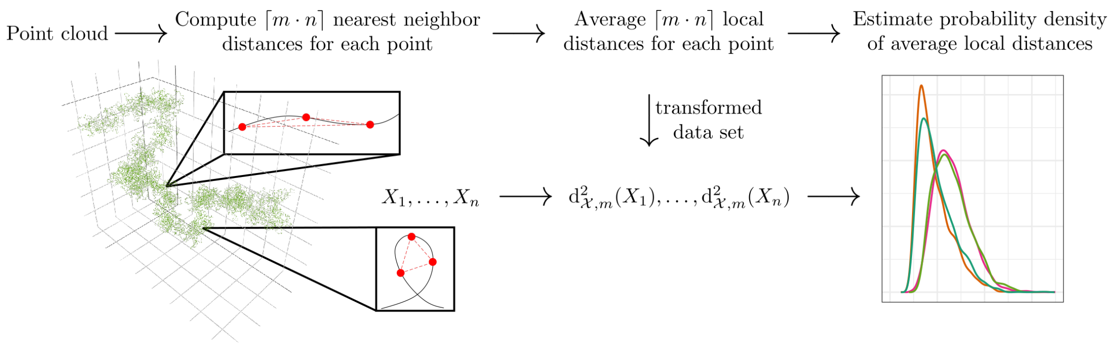

Fig. 2: Data analysis pipeline: Illustration of the different steps in the proposed data analysis (left). The red dots in the details of the image represent data points, the red lines show the point-to-point distances, whereas the underlying chromatin structure is depicted by a black line. Right: The resulting DTM-density estimates of the point clouds illustrated in Figure 1 (same coloring).

aforementioned work, the author considers the asymptotic behavior of the Wasserstein distance between sub-sampled estimates of the DTM-signatures for two different spaces. One big advantage of our method of estimating the DTM-densities over the former is that it does not require sub-sampling and all data points can be used for the analysis.
As illustrated in Section 1.1, the DTM-signature is based on the DTM-function (see (1)). This function has been thoroughly studied and applied in the context of support estimation and topological data analysis [9, 11, 12] and for its sample counterpart (see (3)) many consistency properties have been established in Chazal et al. [13, 14].

Distance based signatures for object discrimination have been applied and studied in a variety of settings [2, 3, 8, 21, 44, 47]. Recently, lower bounds of the Gromov-Wasserstein distance (see Mémoli [37]) have received some attention in applications [22] and in the investigation of their discriminating properties and their statistical behavior [38, 55].

Furthermore, it is noteworthy that nearest neighbor distributions are of great interest in various fields in biology [39, 57] as well as in physics [4, 30, 50]. In these fields it is quite common to consider the (mean of the) distribution of all nearest neighbors for data analysis. While this case corresponds to $m = 1/n$ and is not included in our theoretic analysis, we would like to emphasize that taking the mean over a certain percentage of nearest neighbors makes our method a lot more robust against noise, which is why it performs so well in the analysis of noisy point clouds.
In the analysis of SMLM images, methods from spatial statistics are often employed. Related to the global distribution of all distances is Ripley's K, which is used to infer on the amount and the degree of clustering in a given data set as compared to a point cloud generated by a homogeneous Poisson point process (see, e.g., Nicovich et al. [42] for the application of Ripley's K in this context). Despite the connection via certain distributions

of distances, the objectives and underlying models are quite different to the setting of this paper, such that a direct comparison is not straightforward.

Kernel density estimation from dependent data is a broad and well investigated topic. In addition to the references provided in Section 1.1, kernel density estimators of symmetric functions of the data and dyadic undirected data have been considered [20, 23]. In these settings, the summands of the corresponding kernel density estimators admit a "U-statistic like" dependency structure that has to be accounted for. While this is more closely related to the dependency structure which we are encountering in our analysis, the structure of the statistics that appear in the decomposition of the kernel density estimator (6) is quite different, such that those results cannot directly be transferred to our setting.

1.5 Organization of the Paper

In Section 2 we state the main results and are concerned with the derivation of (7) and the assumptions required for this.

Afterwards, in Section 3 we illustrate our findings via simulations. In Section 4, we apply our methodology to the classification within the framework of chromatin loop analysis.

Notation: Throughout the following, we denote the d-dimensional Lebesgue measure by λ^d and the $(d-1)$-dimensional surface measure in $\mathbb{R}^d$ by σ^{d-1}. We write $B(x,r)$ for the open ball in $\mathbb{R}^d$ (equipped with $||\cdot||$) with center x and radius r. Given a function f or a measure μ, we write $\operatorname{supp}(f)$ and $\operatorname{supp}(\mu)$ to denote their respective support. Let F be a distribution function with compact support $[a,b]$ and let F^{-1} denote the corresponding quantile function. As frequently done, we set $F^{-1}(0) = a$ and $F^{-1}(1) = b$. Let $U \subseteq \mathbb{R}^{d_1}$ be an open set. We denote by $C^k(U, \mathbb{R}^{d_2})$ the set of all k-times continuously differentiable functions from U to $\mathbb{R}^{d_2}$. Further, we denote by $C^{k,1}(U, \mathbb{R}^{d_2})$ the set of all k-times continuously differentiable functions from U to $\mathbb{R}^{d_2}$, whose k'th derivative is Lipschitz continuous. For $d_2 = 1$, we abbreviate this to $C^k(U)$ and $C^{k,1}(U)$. If the domain and range of a function g are clear from the context, we will usually write $g \in C^k$ or $g \in C^{k,1}$.

2 Distributional Limits

In this section, we state our main theoretical results, upon which our statistical methodology is based. We show that $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ is a reasonable estimator for the density of the DTM-signature by proving the distributional limit (7). Before we come to this, we recall the setting, establish the conditions required and ensure that they are met in some simple examples.

2.1 Setting and Assumptions

First of all, we summarize the setting introduced in Section 1.1.

Setting 2.1. *Let $(\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$ denote a regular Euclidean metric measure space. For $x \in \mathcal{X}$ let $\mathrm{d}^2_{\mathcal{X},m}(x)$ denote the corresponding Distance-to-Measure function with mass parameter $m \in (0,1]$. Let $X \sim \mu_{\mathcal{X}}$ and assume that the Distance-to-Measure Signature $\mathrm{d}^2_{\mathcal{X},m}(X)$ has a density $f_{\mathrm{d}^2_{\mathcal{X},m}}$. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$ and denote by $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$ and $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ the kernel density estimators defined in (6) and (5), respectively.*

It is noteworthy that the assumption that $\mathrm{d}^2_{\mathcal{X},m}(X)$ admits a Lebesgue density is slightly restrictive. The probability measure $\mu_{\mathrm{d}^2_{\mathcal{X},m}}$ of the DTM-signature can have a pure point component $\mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{pp}}$ in addition to the continuous component $\mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{cont}}$, if the spaces considered have very little local structure (for examples, see Section 2.2). That is,

$$\mu_{\mathrm{d}^2_{\mathcal{X},m}} = \mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{pp}} + \mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{cont}}.$$

If we define $f_{\mathrm{d}^2_{\mathcal{X},m}}$ to be the Radon-Nikodym derivative of the absolutely continuous component $\mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{cont}}$, i.e., $f_{\mathrm{d}^2_{\mathcal{X},m}}\,\mathrm{d}\lambda = \mathrm{d}\mu_{\mathrm{d}^2_{\mathcal{X},m},\mathrm{cont}}$, the pointwise asymptotic analysis of $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ performed in Section 2.3 (see Theorem 2.12) remains valid for all y with $\mu_{\mathrm{d}^2_{\mathcal{X},m}}(\{y\}) = 0$ that meet the corresponding assumptions. This guarantees that our analysis remains meaningful even if parts of our space do not provide local structure that is discriminative.

In order to derive the statement (7), we require certain regularity assumptions on the density $f_{\mathrm{d}^2_{\mathcal{X},m}}$, the DTM function $\mathrm{d}^2_{\mathcal{X},m}$ and the kernel K. For the sake of completeness, we first recall some facts about the relation of the level sets of a given function. Let $g : \mathbb{R}^d \to \mathbb{R}$ and let $y \in \mathbb{R}$ be such that $g^{-1}(\{y\}) \neq \emptyset$. Suppose that the function g is continuously differentiable in an open environment of $g^{-1}(\{y\})$. Assume further that $\nabla g \neq 0$ on the level set $g^{-1}(\{y\})$. Then, it follows by Cauchy-Lipschitz's theory that there exists a constant $h_0 > 0$, an open set $W \supset g^{-1}([y-h_0, y+h_0])$ and a canonical one parameter family of C^1-diffeomorphisms $\Phi : [-h_0, h_0] \times W \to \mathbb{R}^d$ with the following property:

$$\Phi(v, g^{-1}(\{y\})) = g^{-1}(\{y+v\})$$

for all $v \in [-h_0, h_0]$ (for the precise construction of Φ see the proof of Lemma D.4 in Section D.1). Throughout the following, the family $\{\Phi(v,\cdot)\}_{v\in[-h_0,h_0]}$ (also abbreviated to Φ) is referred to as *canonical level set flow of* $g^{-1}(\{y\})$.

Condition 2.2. *Let $f_{\mathrm{d}^2_{\mathcal{X},m}}$ be supported on $[D_1, D_2]$ and let $y \in [D_1, D_2]$. Assume that there exists $\epsilon > 0$ such that $f_{\mathrm{d}^2_{\mathcal{X},m}}$ is twice continuously differentiable on $(y-\epsilon, y+\epsilon)$. Further, suppose that the function $\mathrm{d}^2_{\mathcal{X},m} : \mathbb{R}^d \to \mathbb{R}$ is $C^{2,1}$ on an open neighborhood of the level set*

$$\Gamma_y \coloneqq {\mathrm{d}^2_{\mathcal{X},m}}^{-1}(\{y\}) = \{x \in \mathbb{R}^d : \mathrm{d}^2_{\mathcal{X},m}(x) = y\},$$

that $\nabla \mathrm{d}^2_{\mathcal{X},m} \neq 0$ on Γ_y and that there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\mathcal{I}_{\mathcal{X}}(y; v) \coloneqq \int_{\Gamma_y} \left| \mathbb{1}_{\{x \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}} \right| \, d\sigma^{d-1}(x) \leq C_y |v|, \tag{8}$$

where $\{\Phi(v, \cdot)\}_{v \in [-h_0, h_0]}$ denotes the canonical level set flow of Γ_y and C_y denotes a finite constant that depends on y and $\mathrm{d}^2_{\mathcal{X},m}$. Suppose that the kernel $K : \mathbb{R} \to \mathbb{R}_+$, is an even, twice continuously differentiable function with $\mathrm{supp}(K) = [-1, 1]$. If $m < 1$, we assume additionally that there are constants $\kappa > 0$ and $1 \leq b < 5$ such that for $u \in (0, 1)$ it holds

$$\omega_{\mathcal{X}}(u) \coloneqq \sup_{x \in \mathcal{X}} \sup_{t,t' \in (0,1)^2, |t-t'| < u} \left| F_x^{-1}(t) - F_x^{-1}(t') \right| \leq \kappa u^{1/b}. \tag{9}$$

The satisfiability of Condition 2.2 is an important issue that is difficult to address in general. Hence, in Section 2.2 we will verify that the requirements of Condition 2.2 are met in several simple examples. Nevertheless, in order to put Condition 2.2 into a broader perspective, we first gather some known regularity properties of $\mathrm{d}^2_{\mathcal{X},m}$ as well as $\{F_x^{-1}\}_{x \in \mathcal{X}}$ and discuss the technical requirement (8) afterwards.

2.1.1 Regularity of $\mathrm{d}^2_{\mathcal{X},m}$ and $\{F_x^{-1}\}_{x \in \mathcal{X}}$

We distinguish between the cases $m < 1$ and $m = 1$ for the presentation of known regularity results. For $m < 1$, the smoothness of $\mathrm{d}^2_{\mathcal{X},m}$ has been investigated in Chazal et al. [11], where the authors derived the following results.

Lemma 2.3.

1. *Let $(\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$ denote an Euclidean metric measure space. Then, the function $\mathrm{d}^2_{\mathcal{X},m} : \mathbb{R}^d \to \mathbb{R}$ is almost everywhere twice differentiable.*
2. *If $\mathcal{X} = (\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$ denotes a regular Euclidean metric measure space, then the function $\mathrm{d}^2_{\mathcal{X},m} : \mathbb{R}^d \to \mathbb{R}$ is differentiable with derivative*

$$\nabla \mathrm{d}^2_{\mathcal{X},m}(x) = \frac{2}{m} \int [x - y] \, d\bar{\mu}_x(y),$$

where $\bar{\mu}_x = \mu_{\mathcal{X}}|_{B(x, \gamma_{\mu_{\mathcal{X}},m}(x))}$ and $\gamma_{\mu_{\mathcal{X}},m}(x) = \inf\{r > 0 : \mu_{\mathcal{X}}\left(\bar{B}(x,r)\right) > m\}$.

Another important point for the case $m < 1$ is the verification of inequality (9). This corresponds to bounding a uniform modulus of continuity for the family $\{F_x^{-1}\}_{x\in\mathcal{X}}$. An application of Lemma 3 in Chazal et al. [13] immediately yields the subsequent result.

Lemma 2.4. *Let $(\mathcal{X}, ||\cdot||, \mu_\mathcal{X})$ be a regular Euclidean metric measure space. Suppose that there are constants $a, b > 0$ such that for all $r > 0$ and all $x \in \mathcal{X}$*

$$\mu_\mathcal{X}(B(x,r)) \geq 1 \wedge ar^b. \tag{10}$$

Then, it holds that

$$\omega_\mathcal{X}(u) \leq 2\left(\frac{h}{a}\right)^{1/b} \operatorname{diam}(\mathcal{X}).$$

Remark 2.5. Condition (10) is frequently assumed in the context of shape analysis. Measures that fulfill (10) are often called *(a,b)-standard* (see Chazal et al. [13], Cuevas [17], Fasy et al. [18] for a detailed discussion of (a,b)-standard measures). In particular, we observe that our assumption (9) is met, whenever $b < 5$.

In the case $m = 1$, it is important to observe that the DTM function admits the following specific form:

$$\mathrm{d}^2_{\mathcal{X},1}(x) = \int_0^1 F_x^{-1}(u)\, du = \mathbb{E}\left[||X - x||^2\right], \tag{11}$$

where $X \sim \mu_\mathcal{X}$. This identity allows us to derive the following lemma.

Lemma 2.6. *Let $\mathcal{X} = (\mathcal{X}, ||\cdot||, \mu_\mathcal{X})$ denote a regular Euclidean metric measure space and let $X \sim \mu_\mathcal{X}$. Then, it holds that:*

1. *The function $\mathrm{d}^2_{\mathcal{X},1} : \mathbb{R}^d \to \mathbb{R}$ is given as*

$$x = (x_1, \ldots, x_d) \mapsto (x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots + (x_d - c_d)^2 + \zeta, \tag{12}$$

   where $c = (c_1, \ldots, c_d)^T = \mathbb{E}[X]$ and ζ denotes a finite constant that can be made explicit.

2. *The function $\mathrm{d}^2_{\mathcal{X},1} : \mathbb{R}^d \to \mathbb{R}$ is three times continuously differentiable.*
3. *We have $\nabla \mathrm{d}^2_{\mathcal{X},1}(x) = 0$ if and only if $x = \mathbb{E}[X]$.*
4. *Consider the representation of $\mathrm{d}^2_{\mathcal{X},1}$ in (12). Set $\Gamma_y = {\mathrm{d}^2_{\mathcal{X},1}}^{-1}(\{y\})$ and suppose that $\left(y - \mathbb{E}\left[||X - \mathbb{E}[X]||^2\right]\right) > 2h_0 > 0$. The canonical level set flow $\{\Phi(v,\cdot)\}_{v\in[-h_0,h_0]}$ of Γ_y considered as function from $[-h_0, h_0] \times {\mathrm{d}^2_{\mathcal{X},1}}^{-1}([y - h_0, y + h_0])$ to $\mathbb{R}^d$ is for $x = (x_1, \ldots, x_d)$ given as*

$$(v, x) \mapsto \left((x_1 - c_1)\sqrt{1 + \frac{v}{||x - c||^2}} + c_1, \ldots, (x_d - c_d)\sqrt{1 + \frac{v}{||x - c||^2}} + c_d\right). \tag{13}$$

In order to increase the readability of this section, the proof of Lemma 2.6 is postponed to Section A in the Appendix.

2.1.2 Discussion of assumption (8) in Condition 2.2

To conclude this section, we consider the technical assumption (8). First of all, it is obvious (if $\mathrm{d}^2_{\mathcal{X},m}$ is nowhere constant) that the assumption only comes into play for $d \geq 2$. Furthermore, we observe that it is trivially fulfilled if there exists some $\epsilon > 0$ such that $\Gamma_{y-\epsilon} \subset \mathcal{X}$, $\Gamma_y \subset \mathcal{X}$ and $\Gamma_{y+\epsilon} \subset \mathcal{X}$. Only if this is not the case, there might be points $y \in \mathcal{X}$ for which (8) is not satisfied. However, the assumption will typically be satisfied for all points of regularity of the density $f_{\mathrm{d}^2_{\mathcal{X},m}}$. To provide some intuition on this matter, we will consider the following example.

Example 2.7. Let $\mathcal{X} = [0,1]^2$ and let $\mu_{\mathcal{X}}$ stand for the uniform distribution on $\mathcal{X}$. In this case, using relation (11), we obtain for $x = (x_1, x_2) \in \mathcal{X}$

$$\mathrm{d}^2_{\mathcal{X},1}(x) = \mathbb{E}\left[||X - x||^2\right] = \left(x_1 - \frac{1}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^2 + \frac{1}{6}.$$

The corresponding DTM-density is supported on $[1/6, 2/3]$ and it is smooth everywhere except for $y = 5/12$, where $f_{\mathrm{d}^2_{\mathcal{X},m}}$ has a kink (detailed computations are provided in Section B.3 in the appendix). The level sets Γ_y ($y \geq 1/6$), are concentric circles centered at $(1/2, 1/2)$ with radii $\sqrt{y - 1/6}$. For all $y < 5/12$ the level sets are fully contained in the open cube $(0,1)^2$. For all $y > 5/12$, we have $\mathbb{R}^2 \backslash [0,1]^2 \cap \Gamma_y \neq \emptyset$, i.e., the level sets are at least partly outside of the cube $[0,1]^2$. This means that $y = 5/12$ is, in a sense, a transition point. In order to check (8) for $y \geq 5/12$, we observe that Lemma 2.6 implies that for $v > 0$, $y \in [5/12, 2/3]$ and each $x \in \mathcal{X}$ the equality $\mathbb{1}_{\{x \in \Gamma_y \cap \mathcal{X}\}} = 0$ implies $\mathbb{1}_{\{\Phi(v,x))\Gamma_y \cap \mathcal{X}\}} = 0$, where Φ denotes the canonical level set flow of Γ_y. Consequently, it follows that

$$\mathcal{I}_{\mathcal{X}}(y; v) = \left| \int_{\Gamma_y} \mathbb{1}_{\{x \in \mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}} \, d\sigma^1(x) \right| = \left|\mathcal{L}(\Gamma_y \cap \mathcal{X}) - \mathcal{L}(\Gamma_{y+v} \cap \mathcal{X})\right|,$$

where $\mathcal{L}(\Gamma_y \cap \mathcal{X})$ stands for the length of the curve Γ_y in $\mathcal{X}$ and $\mathcal{L}(\Gamma_{y+v} \cap \mathcal{X})$ is defined analogously. Using the above equality, it is easy to verify that the requirement (8) is satisfied for all $y \in [1/6, 2/3] \backslash \{5/12\}$. Figure 3 exemplarily illustrates the behavior of the level sets in a neighborhood of $(1/2, 1)$. The figure shows the level sets $\Gamma_{5/12}$ in blue and $\Gamma_{5/12+v}$ for some $v > 0$ as dotted line in black. We observe that for $0 \leq v \leq h_0$ for some h_0 sufficiently small, the value of the integral $\mathcal{I}_{\mathcal{X}}(y; v)$ corresponds to $2\pi v$ minus four times

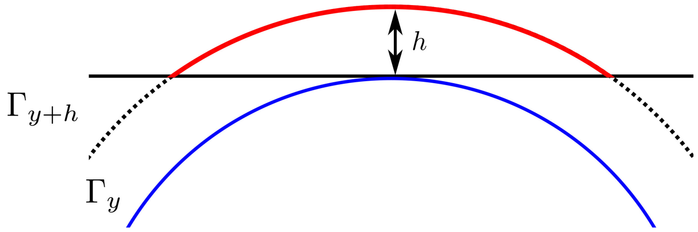

Fig. 3: Tangential Level Set: Illustration of the behavior of the levels sets in a neighborhood of tangential intersection point with the boundary of $\mathcal{X}$ in the setting of Example 2.7.

the length of the red line, which can be calculated explicitly using a well-known formula for circular segments:

$$\mathcal{I}_{\mathcal{X}}(y;v) = \left|2\pi v - 4\arcsin\left(2\sqrt{v-v^2}\right)\right| \geq \sqrt{v}.$$

This proves that for $y = 5/12$ the requirement (8) is not fulfilled.

We conclude this subsection by noting that the dimension of $\mathcal{X}$ heavily influences the regularity of (8). While it seems to be problematic, if Γ_y intersects tangentially with the boundary $\partial\mathcal{X}$ of $\mathcal{X}$ for $d = 2$, this is not necessarily the case for $d \geq 3$. In particular, if we consider $\mathcal{X} = [0,1]^3$ equipped with the uniform distribution, we find that for $y = 3/4$ the level set Γ_y tangentially touches $\partial\mathcal{X}$ at 6 points. However, here, it does not cause any problems. Following our considerations from Example 2.7, one can show that condition (8) holds for all points y in the support of $f_{\mathrm{d}^2_{\mathcal{X},m}}$.

2.2 Examples of DTM-Densities

In the following, we will derive $\mathrm{d}^2_{\mathcal{X},m}$ as well as $f_{\mathrm{d}^2_{\mathcal{X},m}}$ in several simple examples explicitly and verify that in these settings Condition 2.2 is met almost everywhere. Since calculating $\mathrm{d}^2_{\mathcal{X},m}$ and $f_{\mathrm{d}^2_{\mathcal{X},m}}$ explicitly is quite cumbersome (especially for $m < 1$), we concentrate on one- or two-dimensional examples. In order to increase the readability of this section, we postpone the the explicit, but lengthy representations of the derived DTM-functions and densities (as well as their derivation) to Section B.

We begin our considerations with the simplest case possible, the interval $[0,1]$ equipped with the uniform distribution.

Example 2.8. Let $\mathcal{X} = [0,1]$ and let $\mu_{\mathcal{X}}$ denote the uniform distribution on $\mathcal{X}$. Furthermore, we consider two values for m, namely $m_1 = 1$ and $m_2 = 0.1$. In Section B.1, we derive $\mathrm{d}^2_{\mathcal{X},1}$, $\mathrm{d}^2_{\mathcal{X},0.1}$, $f_{\mathrm{d}^2_{\mathcal{X},1}}$ (see Figure 4 for an illustration). For $m = 1$, the requirement (8) does not come into play as $\mathcal{X}$ is one-dimensional and $\mathrm{d}^2_{\mathcal{X},1}$ nowhere constant. Further, we point out that the density $f_{\mathrm{d}^2_{\mathcal{X},1}}$ is unbounded (however twice continuously differentiable on the interior of its support). In the case $m = 0.1$ things are quite different. The function $\mathrm{d}^2_{\mathcal{X},0.1}$ is constant on $[0.05, 0.95]$ and hence the random variable $\mathrm{d}^2_{\mathcal{X},0.1}(X)$, $X \sim \mu_{\mathcal{X}}$, does not have a Lebesgue density.

It is immediately clear that the DTM-signature can only admit a density with respect to the Lebesgue measure, if the DTM-function defined in (1) is almost nowhere constant. In Example 2.8 this is the case for $m_1 = 1$ but not for $m_2 = 0.1$. Recall that the DTM-function considers the quantile function of the random variable $||X - x||^2$, $\mathcal{X} \sim \mu_{\mathcal{X}}$, on $[0, m]$ for each $x \in \mathcal{X}$. In Example 2.8, $\mu_{\mathcal{X}}$ denotes the uniform measure on $\mathcal{X} = [0, 1]$. Hence, it is evident in this setting that the quantile functions of the random variables $\{||X - x||\}_{x\in[m/2,1-m/2]}$ agree on $[0, m]$. Consequently, the corresponding DTM-signature admits a Lebesgue density only for $m = 1$. In the next example, we equip $\mathcal{X}$ with another distribution, whose density is not constant on $\mathcal{X}$. In this case, we will find that also for $m < 1$ the corresponding DTM-signature admits a Lebesgue density.

Example 2.9. Let $\mathcal{X} = [0, 1]$ and let $\mu_{\mathcal{X}}$ denote the probability distribution on $[0, 1]$ with density $f(x) = 2x$. Let $m = 0.1$. In Section B.2, we derive $\mathrm{d}^2_{\mathcal{X},0.1}$ explicitly and demonstrate that the random variable $\mathrm{d}^2_{\mathcal{X},0.1}(X)$, $X \sim \mu_{\mathcal{X}}$, admits a Lebesgue density in this setting (see Figure 4 for an illustration). We observe that $\mathrm{d}^2_{\mathcal{X},0.1}$ is continuously differentiable everywhere and three times continuously differentiable almost everywhere. Further, the density $f_{\mathrm{d}^2_{\mathcal{X},0.1}}$ admits one discontinuity for $y = \frac{-683}{60} + 18\sqrt{\frac{2}{5}}$ and is C^2 almost everywhere.

We observe that the DTM-densities derived in Example 2.8 and Example 2.9 are both unbounded. This has a simple explanation. Let $(\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$ be a regular Euclidean metric measure space and denote the d-dimensional Lebesgue density of $\mu_{\mathcal{X}}$ by $g_{\mu_{\mathcal{X}}}$. Suppose that $f_{\mathrm{d}^2_{\mathcal{X},m}}$ exists. Then, one can show (see e.g. Appendix C of Weitkamp et al. [55]) that

$$f_{\mathrm{d}^2_{\mathcal{X},m}}(y) = \int_{\{x\in\mathcal{X}:\mathrm{d}^2_{\mathcal{X},m}(x)=y\}} \frac{g_{\mu_{\mathcal{X}}}(u)}{||\nabla \mathrm{d}^2_{\mathcal{X},m}(u)||}\, d\sigma^{d-1}(u). \tag{14}$$

Since $d\sigma^0$ corresponds to integration with respect to the counting measure, the DTM-density of a one-dimensional Euclidean metric measure space is unbounded if there are $u \in \mathcal{X}$ with $|\nabla \mathrm{d}^2_{\mathcal{X},m}(u)| = 0$ (this is the case in Example 2.8 and Example 2.9). However, it is important to note that this behavior mainly occurs for one-dimensional Euclidean metric measure spaces. For higher dimensional spaces, the area (w.r.t. $d\sigma^{d-1}$) of the set $A = \{x \in \mathcal{X} : ||\nabla \mathrm{d}^2_{\mathcal{X},m}(u)|| = 0\}$, is usually a null set. Hence, it is possible that the density $f_{\mathrm{d}^2_{\mathcal{X},m}}$ defined in (14) remains bounded even if A is non-empty (see Example 2.7 and Example 2.10).

To conclude this section and in order to illustrate that the showcased regularity of the DTM-function $\mathrm{d}^2_{\mathcal{X},m}$ and the DTM-density $f_{\mathrm{d}^2_{\mathcal{X},m}}$ does not only hold for one-dimensional settings, we consider two simple examples in $\mathbb{R}^2$ next. As the derivation of the family $(F_x^{-1})_{x\in\mathcal{X}}$ can be incredibly time consuming, we restrict ourselves in the following to the case $m = 1$.

Example 2.7 (Continued)**.** Recall that $\mathcal{X} = [0,1]^2$, $\mu_{\mathcal{X}}$ stands for the uniform distribution on $\mathcal{X}$ and that $m = 1$. Based on our previous considerations it is possible to derive $f_{\mathrm{d}^2_{\mathcal{X},1}}$ explicitly (see Section B.3 for the derivation). As illustrated in Figure 4, the density $f_{\mathrm{d}^2_{\mathcal{X},1}}$ is continuous. Moreover, it is twice continuously differentiable inside its support for $y \neq \frac{5}{12}$, which is also the only point where the requirements of (8) are not met, as discussed previously.

We note that the density $f_{\mathrm{d}^2_{\mathcal{X},1}}$ derived in Example 2.7 is constant on $[1/6, 5/12]$. This kind of behavior is also expressed when considering a disc in $\mathbb{R}^2$ equipped with the uniform distribution (it is easy to verify that $f_{\mathrm{d}^2_{\mathcal{X},1}}$ is a constant function in this case). It is well known that it is difficult for kernel density estimators to approximate constant pieces or a constant function. However, it is not reasonable to assume that the data stems from a uniform distribution over a compact set in many applications (such as chromatin loop analysis). More often, it is possible to assume that the data generating distribution is more concentrated in the center of the considered set. The final example of this section showcases that in such a case the corresponding DTM-signature admits a density without any constant parts even on the disk.

Example 2.10. Let $\mathcal{X}$ denote a disk in $\mathbb{R}^2$ centered at $(0,0)$ with radius 1 and let $\mu_{\mathcal{X}}$ denote probability measure with density

$$f(x_1, x_2) = \begin{cases} -\frac{2}{\pi}\left(x_1^2 + x_2^2 - 1\right) & x_1^2 + x_2^2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

In this framework, we derive $\mathrm{d}^2_{\mathcal{X},1}$ and $f_{\mathrm{d}^2_{\mathcal{X},1}}$ in Section B.4. We observe that the level sets Γ_y (of $\mathrm{d}^2_{\mathcal{X},1}$) are contained in $\mathcal{X}$ for any $y \in [1/3, 4/3]$, i.e., condition (8) is met for all $y \in (1/3, 4/3)$ in this setting. Further, we realize that $f_{\mathrm{d}^2_{\mathcal{X},1}}$ (see Figure 4 for an illustration) is smooth and nowhere constant on the interior of its support.

2.3 Theoretical Results

In this section we study the asymptotic behavior of the kernel estimator of the DTM-density (6). Clearly, standard methodology implies the following pointwise central limit theorem for the kernel estimator $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$ defined in (5).

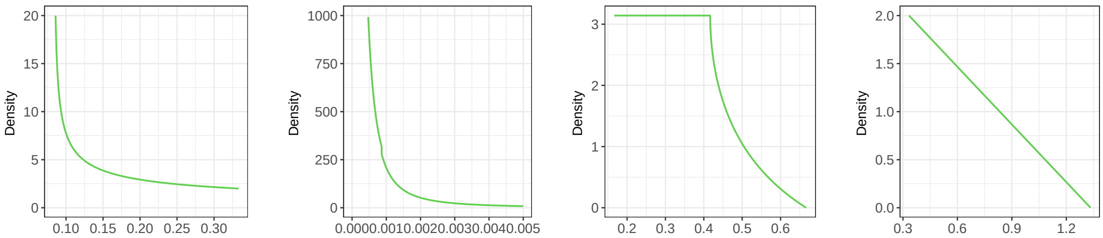

Fig. 4: Distance-to-Measure signature: Illustration of the densities calculated in Example 2.8-2.9, Example 2.7 and Example 2.10 (from left to right).

Theorem 2.11. *Assume Setting 2.1 and suppose that* $\mathrm{d}^2_{\mathcal{X},m}(X_1)$ *admits a density that is twice continuously differentiable in an environment of* y*. Suppose further that the kernel* $K : \mathbb{R} \to \mathbb{R}_+$*, is an even, twice continuously differentiable function with* $\mathrm{supp}(K) = [-1, 1]$*. Then, it holds for* $n \to \infty$*,* $h = o\left(n^{-1/5}\right)$ *and* $nh \to \infty$ *that*

$$\sqrt{nh}\left(\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\int K^2(u)\,du\right).$$

Surprisingly perhaps, despite the complicated stochastic dependence of the random variables $\delta^2_{\mathcal{X},m}(X_i)$, asymptotically, $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},m}}$ and $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ behave equivalently in the following sense.

Theorem 2.12. *Assume Setting 2.1 and let Condition 2.2 hold. Then, it holds for* $n \to \infty$*,* $h = o\left(n^{-1/5}\right)$ *and* $nh \to \infty$ *that*

$$\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\int K^2(u)\,du\right).$$

As the the proof of Theorem 2.12 is lengthy and quite technical, it has been deferred to Appendix C. There, we will write the density estimator $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ as a U-statistic plus remainder terms. Then, using a Hoeffding decomposition, the dependencies can be handled. However, showing that the remainder terms vanish is not trivial and requires the application of some tools from geometric measure theory.

3 Simulations

In the following, we investigate the finite sample behavior of $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ in Monte Carlo simulations. To this end, we illustrate the pointwise limit derived in Theorem 2.11 in the setting of Example 2.10 and exemplarily highlight the discriminating potential of $\widehat{f}_{\delta^2_{\mathcal{X},m}}$. All simulations were performed in R (R Core Team [36]).

3.1 Pointwise Limit

We start with the illustration of Theorem 2.11. To this end, we consider the Euclidean metric measure space $(\mathcal{X}, ||\cdot||, \mu_{\mathcal{X}})$ from Example 2.10. Recall that in this setting, $\mathcal{X}$ denotes a disk in $\mathbb{R}^2$ centered at $(0,0)$ with radius 1 and that $\mu_{\mathcal{X}}$ denotes the probability measure with density

$$f(x_1, x_2) = \begin{cases} -\frac{2}{\pi}\left(x_1^2 + x_2^2 - 1\right) & x_1^2 + x_2^2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

Now, we choose $m = 1$ and consider

$$\widehat{f}_{\delta^2_{\mathcal{X},1}}(y) = \frac{1}{nh_1}\sum_{i=1}^{n} K_{Bi}\left(\frac{\delta^2_{\mathcal{X},m}(X_i) - y}{h_1}\right),$$

where K_{Bi} denotes the Biweight kernel, i.e.,

$$K_{Bi}(u) = \begin{cases} \frac{15}{16}\left(1 - u^2\right)^2 & |u| \leq 1, \\ 0 & \text{else.} \end{cases} \tag{15}$$

Since we have calculated $\mathrm{d}^2_{\mathcal{X},1}$ explicitly (see (18)), it is of interest to compare the behavior of $\widehat{f}_{\delta^2_{\mathcal{X},1}}(y)$ to the one of

$$\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}(y) = \frac{1}{nh_2}\sum_{i=1}^{n} K_{Bi}\left(\frac{\mathrm{d}^2_{\mathcal{X},1}(X_i) - y}{h_2}\right).$$

As discussed previously, $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}$ is different from $\widehat{f}_{\delta^2_{\mathcal{X},1}}$ a kernel density estimator based on independent data, whose limit behavior is well understood (see Theorem 2.11). Nevertheless, for $y \in (1/3, 4/3)$, $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}(y)$ and $\widehat{f}_{\delta^2_{\mathcal{X},1}}(y)$ admit the same asymptotic behavior according to Theorem 2.12, whose requirements can be easily checked in this setting (see Example 2.10). In order to illustrate this, we generate two independent samples $\{X_i\}_{i=1}^n$ and $\{X_i'\}_{i=1}^n$ of $\mu_{\mathcal{X}}$ and calculate $\Delta_n = \{\delta^2_{\mathcal{X},1}(X_i)\}_{i=1}^n$ as well as $D_n = \{\mathrm{d}^2_{\mathcal{X},1}(X_i)\}_{i=1}^n$ for $n = 50, 500, 2500, 5000$. We set

$$h_1 = (1.06 \min\{s(\Delta_n), \mathrm{IQR}(\Delta_n)/1.34\})^{5/4} n^{-1/4}$$

and

$$h_2 = (1.06 \min\{s(D_n), \mathrm{IQR}(D_n)/1.34\})^{5/4} n^{-1/4},$$

where s is the usual sample standard deviation and IQR denotes the inter quartile range. Based on Δ_n and D_n, we choose a central value of y and calculate

$$\sqrt{nh_1}(\widehat{f}_{\delta^2_{\mathcal{X},1}}(y) - f_{\mathrm{d}^2_{\mathcal{X},1}}(y)) \text{ and } \sqrt{nh_2}(\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}(y) - f_{\mathrm{d}^2_{\mathcal{X},1}}(y)). \tag{16}$$

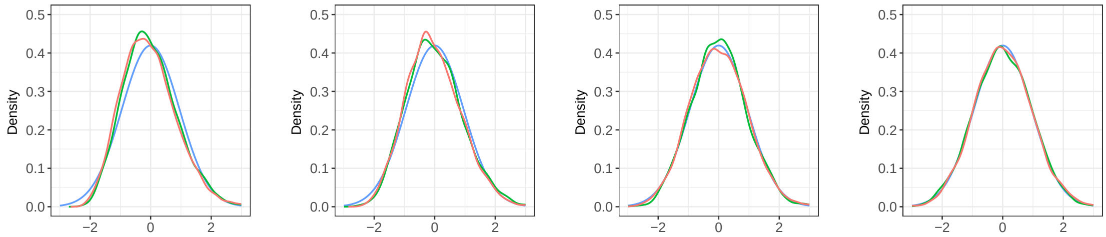

Fig. 5: Pointwise limit distribution: Kernel density estimators of $\sqrt{nh_1}(\widehat{f}_{\delta^2_{\mathcal{X},1}}(0.7) - f_{\mathrm{d}^2_{\mathcal{X},1}}(0.7))$ (in red) and $\sqrt{nh_2}(\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}(0.7) - f_{\mathrm{d}^2_{\mathcal{X},1}}(0.7))$ (in green) for $n = 50, 500, 2500, 5000$ (from left to right, sample size 5,000) and the normal limiting density (blue).

For each n, we repeat this process 5,000 times. The finite sample distributions of the quantities defined in (16) are compared to their theoretical normal counter part in Figure 5 (exemplarily for the specific choice of $y = 0.7$). The kernel density estimators displayed (Gaussian kernel with bandwidth given by Silverman's rule) highlight that the asymptotic behavior of $\widehat{f}_{\delta^2_{\mathcal{X},1}}(y)$ (red) matches that of $\widehat{f}_{\mathrm{d}^2_{\mathcal{X},1}}(y)$ (green). Further, we observe that even for small samples sizes both finite sample distributions strongly resemble their theoretical normal limit distribution (blue).

3.2 Discriminating Properties

In the remainder of this section, we will showcase empirically the potential of the DTM-signature for discriminating between different Euclidean metric measure spaces. To this end, let $\mu_{\mathcal{Y}_1}$ stand for the uniform distribution on a 3D-pentagon (inner pentagon side length: 1, Euclidean distance between inner and outer pentagon: 0.4, height: 0.4) and let $\mu_{\mathcal{Y}_7}$ denote the uniform distribution on a torus (center radius: 1.169, tube radius: 0.2) with the same center and orientation (see the plots for $t_1 = 0$ and $t_7 = 1$ in Figure 6). In order to interpolate between these measures, let $\Pi^{\mu_{\mathcal{Y}_7}}_{\mu_{\mathcal{Y}_1}}(t)$, $t \in [0, 1]$, denote the 2-Wasserstein geodesic between $\mu_{\mathcal{Y}_1}$ and $\mu_{\mathcal{Y}_7}$ (see e.g. Santambrogio [46, Sec. 5.4] for a formal definition). Figure 6 displays the Euclidean metric measure spaces $\mathcal{Y}_i$, $1 \leq i \leq 7$, corresponding to $\mu_{\mathcal{Y}_i} = \Pi^{\mu_{\mathcal{Y}_7}}_{\mu_{\mathcal{Y}_1}}(t_i)$ for $t_i \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$ (the geodesic has been approximated discretely based on 40,000 points with the the WSGeometry-package [28]).

In this example, we are not interested in only finding local changes, but we want to distinguish between Euclidean metric measure spaces that differ globally. Hence, $m = 1$ seems to be the most reasonable choice. At the end of this section, we will illustrate the influence of the parameter m in the present setting. We draw independent samples of size n from $\mu_{\mathcal{Y}_i}$, denoted as $\{Y_{j,n,i}\}_{j=1}^n$, and calculate $\Delta_{n,i} = \{\delta^2_{\mathcal{X},1}(Y_{j,n,i})\}_{j=1}^n$ and $\widehat{f}_{\delta^2_{\mathcal{Y}_i,1}}$

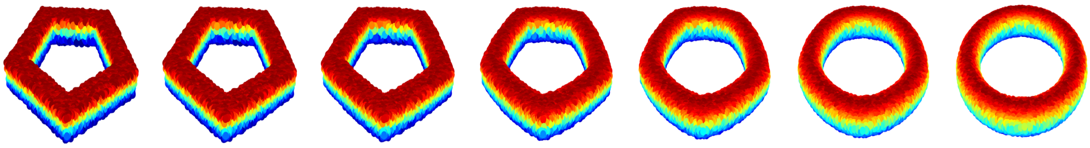

Fig. 6: Metric measure spaces: Graphical illustration of the metric measure spaces $\{\mathcal{Y}_i\}_{i=1}^7$.

based on each of these samples for $1 \leq i \leq 7$ and $n = 500, 2500, 5000, 10000$ (Biweight kernel, $h_i = 1.06 \min\{s(\Delta_n), \mathrm{IQR}(\Delta_n)/1.34)\} n^{-1/5}$). We repeat this procedure for each i and n 10 times and display the resulting kernel density estimators in the upper row of Figure 7. While it is not possible to reliably distinguish between the realizations of $\widehat{f}_{\delta^2_{\mathcal{Y}_0,1}}$ (blue-green), $\widehat{f}_{\delta^2_{\mathcal{Y}_1,1}}$ (orange) and $\widehat{f}_{\delta^2_{\mathcal{Y}_2,1}}$ (blue) by eye for $n = 500$, this is very simple for $n \geq 2500$. Now, that we have estimated the densities, we can choose a suitable notion of distance between densities (e.g. the L^1-distance) and perform a linkage clustering in order to showcase that the illustrations in the upper row are not deceptive and that it is indeed possible to discriminate between the Euclidean metric measure spaces considered based on the kernel density estimators of the respective DTM-densities. To this end, we calculate the L_1-distance between the kernel density estimators considered and perform an average linkage clustering on the resulting distance matrix for each n. The results are showcased in the lower row of Figure 7. The average linkage clustering confirms our previous observations.

To conclude this section, we illustrate the influence of the choice of m. For this purpose, we repeat the above procedure with $n = 10,000$ and $m = 0.2, 0.4, 0.6, 0.8$ (this means that we can use the alternative representation of $\mathrm{d}^2_{\mathcal{X},m}$ in (4) with $k = 2000, 4000, 6000, 8000$). The resulting kernel density estimators are displayed in the upper row and the corresponding average clustering in the bottom row of Figure 8 (same coloring as previously). As we consider the transformation of $\mu_{\mathcal{Y}_1}$ into $\mu_{\mathcal{Y}_7}$ along a 2-Wasserstein geodesic, it is intuitive that choosing m too small is not informative in this setting (the goal is to distinguish between the whole spaces). Indeed, this is exactly, what we observe. For $m = 0.2$ the kernel density estimators strongly resemble each other and in particular the Euclidean metric measure spaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are hardly distinguishable (see the corresponding dendrogram in the lower row of Figure 8). For $m \geq 0.4$ the kernel density estimators are better separated and the corresponding dendrograms highlight that it possible to discriminate between the spaces $\mathcal{Y}_i$ based on the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{Y}_i,m}}$, $i = 1, \ldots, 7$ and $m = 0.4, 0.6, 0.8$. It is noteworthy that although the form of the kernel density estimators drastically changes between $m = 0.4$ and $m = 1$, the quality of the corresponding clustering only increases slightly with increasing m.

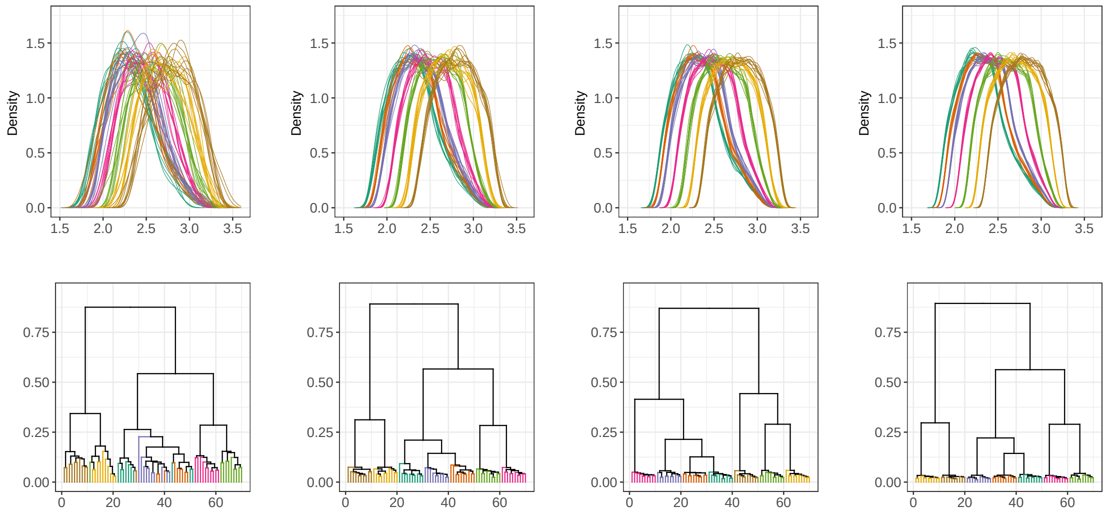

Fig. 7: Discriminating between Euclidean metric measure spaces: Upper row: Ten realizations of the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{Y}_1,1}}$ (blue-green), $\widehat{f}_{\delta^2_{\mathcal{Y}_2,1}}$ (orange), $\widehat{f}_{\delta^2_{\mathcal{Y}_3,1}}$ (blue), $\widehat{f}_{\delta^2_{\mathcal{Y}_4,1}}$ (pink), $\widehat{f}_{\delta^2_{\mathcal{Y}_5,1}}$ (green), $\widehat{f}_{\delta^2_{\mathcal{Y}_6,1}}$ (yellow) and $\widehat{f}_{\delta^2_{\mathcal{Y}_6,1}}$ (brown) for $n = 500, 2500, 5000, 10000$ (from left to right). Lower row: The results of an average linkage clustering of the considered kernel density estimators based on the L^1-distance (same coloring).

4 Chromatin Loop Analysis

In this section, we will highlight how to use the DTM-density-transformation for chromantin loop analysis. First, we briefly recall some important facts about chromatin fibers, state the goal of this analysis and precisely describe the data used here.

For human beings, chromosomes are essential parts of cell nuclei. They carry the genetic information important for heredity transmission and consist of chromatin fibers. Learning the topological 3D structure of the chromatin fiber in cell nuclei is important for a better understanding of the human genome. As discussed in Section 1.3, TADs are self-interacting genomic regions, which are often associated with loops in the chromatin fibers. These domains have been estimated to the range of 100–300 nm [43]. Hi-C data [33] allow to construct spatial proximity maps of the human genome and are often used to analyze genome-wide chromatin organization and to identify TADs. However, spatial size and form, and how frequently chromatin loops and domains exist in single cells, cannot directly be answered based on Hi-C data, whereas in 3D visualization of chromosomal regions via SMLM with a sufficiently high resolution, this information might be more easily accessible [26]. Therefore, in the above reference, such an approach is considered, in which two groups

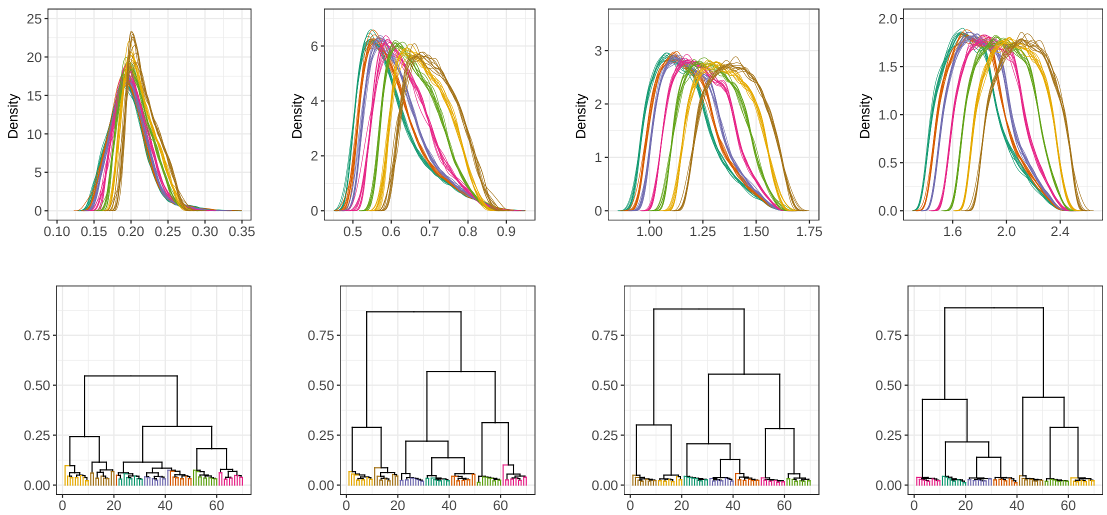

Fig. 8: The influence of m: Upper row: Ten realizations of the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{Y}_1,m}}$ (blue-green), $\widehat{f}_{\delta^2_{\mathcal{Y}_2,m}}$ (orange), $\widehat{f}_{\delta^2_{\mathcal{Y}_3,m}}$ (blue), $\widehat{f}_{\delta^2_{\mathcal{Y}_4,m}}$ (pink), $\widehat{f}_{\delta^2_{\mathcal{Y}_5,m}}$ (green), $\widehat{f}_{\delta^2_{\mathcal{Y}_6,m}}$ (yellow) and $\widehat{f}_{\delta^2_{\mathcal{Y}_7,m}}$ (brown) for $n = 10000$ and $m = 0.2, 0.4, 0.6, 0.8$ (from left to right). Lower row: The results of an average linkage clustering of the considered kernel density estimators based on the L^1-distance (same coloring).

of images of chromatin fibers were produced: Chromatin with supposedly fully intact loop structures and chromatin, which had been treated with auxin prior to imaging. Auxin is known to cause a degrading of the loops. Therefore, in the second set of images, the loops are expected to be mostly dissolved. The obtained resolution in these images was of the order of 150 nm, i.e., below the diffraction limit and comparable to the typical sizes of TADs. This means that the analysis of chromatin loops based on these images is tractable but difficult as we will not see detailed loops when zooming in.

In this paper we analyse simulated SMLM data of chromatin fibers that mimic the chromatin structure with loops as local features and compare them to simulated data that mimic the progressive degradation of loop structures in five steps. The simulated structures mimic the first chromosome (of 23 in total) of the human genome, which is the longest with approximately 249 megabases (Mb, corresponding to 249,000,000 nucleotides). Each step corresponds to a loop density with a different parameter, which we denote by c. The value of c is the number of loops per megabase. A value of $c = 25$ corresponds to a high loop density with 2490 loops in total and corresponds to the setting without the application of auxin. Values of $c = 10, 6, 4, 2$ correspond to decreasing states of resolved loops (1494, 996 and 498 loops) and $c = 0$ encodes the fully resolved state. These simulated images provide a controlled setting in which we can investigate the applicability of our methods and in which we can explore how small a difference in loop density our method can still

pick up and when it starts to break down. Here, we only consider classification into the different conditions based on the estimated DTM density. While it is clear from the results described below that information on loop size and frequency is encoded in these densities, a quantification of these parameters requires a deeper study of the proposed methods and is beyond the scope of this manuscript.
In our study, we consider 102 synthetic, noisy samples of size 49800 of 6 different loop densities each and denote the corresponding samples as $\mathcal{X}_{i,c}$, $c = 25, 10, 6, 4, 2, 0$, $1 \leq i \leq 102$. These samples are created by first discretizing the chromatin structure such that the distance between two points along the chromatin structure corresponds to 45 nm. Then, we add independent, centered Gaussian errors with covariance matrix

$$\Sigma = \begin{pmatrix} 45 & 0 & 0 \\ 0 & 45 & 0 \\ 0 & 0 & 90 \end{pmatrix}$$

to each point (see Figure 9 for an illustration of data obtained in this fashion). This high level of noise is chosen to match the experimental data obtained in Hao et al. [26]. Throughout the following, we consider the data on a scale of 1:45. We stress once again that the goal of our analysis is to distinguish between the respective loop conditions and not between chromatin fibers from which the points are sampled (the overall form of the chromatin fibers within one condition can be quite different). We demonstrate in the following that the corresponding DTM-signatures, or more precisely the corresponding kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$, $1 \leq i \leq 102$, (m chosen suitably small) represent a useful transformation of the data that allows discrimination between the different loop densities, while disregarding the overall shape of the chromatin fiber. To this end, we follow the strategy proposed in Section 1.3 and calculate $\Delta_{i,c} = \{\delta^2_{\mathcal{X},m}(X_{j,i,c}) : X_{j,i,c} \in \mathcal{X}_{i,c}\}$ for $1 \leq i \leq 102$, $c \in \{25, 10, 6, 4, 2, 0\}$ and $m \in \{1/9960, 1/4980, 1/1245\}$. These particular choices of m entail that in order to calculate $\delta^2_{\mathcal{X},m}(X_{j,i,c})$ we need to take the mean over the distances to the $k = 5, 10, 40$ nearest neighbors of $X_{j,i,c} \in \mathcal{X}_{i,c}$ (recall the representation of $\delta^2_{\mathcal{X},m}$ in (4)). We determine $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$ based on each of the samples $\Delta_{i,c}$ (Biweight kernel, $h = 1.06 \min\{s(\Delta_n), \mathrm{IQR}(\Delta_n)/1.34\}n^{-1/5}$). The resulting kernel density estimators are displayed in Figure 10. Generally, the kernel density estimators based on the different samples with the same loop density strongly resemble each other and it is possible to roughly distinguish between the different values of c. For all values of m considered, the DTM-density estimators based on $\mathcal{X}_{i,25}$, $1 \leq i \leq 102$, (here the respective chromatin fibers form many loops) are well separated from the other kernel density estimators and the estimators based on the samples $\Delta_{i,2}$ and $\Delta_{i,0}$ (which correspond to the lowest loop densities considered) are the most similar when comparing the different loop densities. In order to make a more qualitative comparison between the estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$, we use the strategy developed in Section 3.2 and perform an average linkage clustering based on the L_1-distance between the estimated densities. For clarity, we restrict ourselves to the

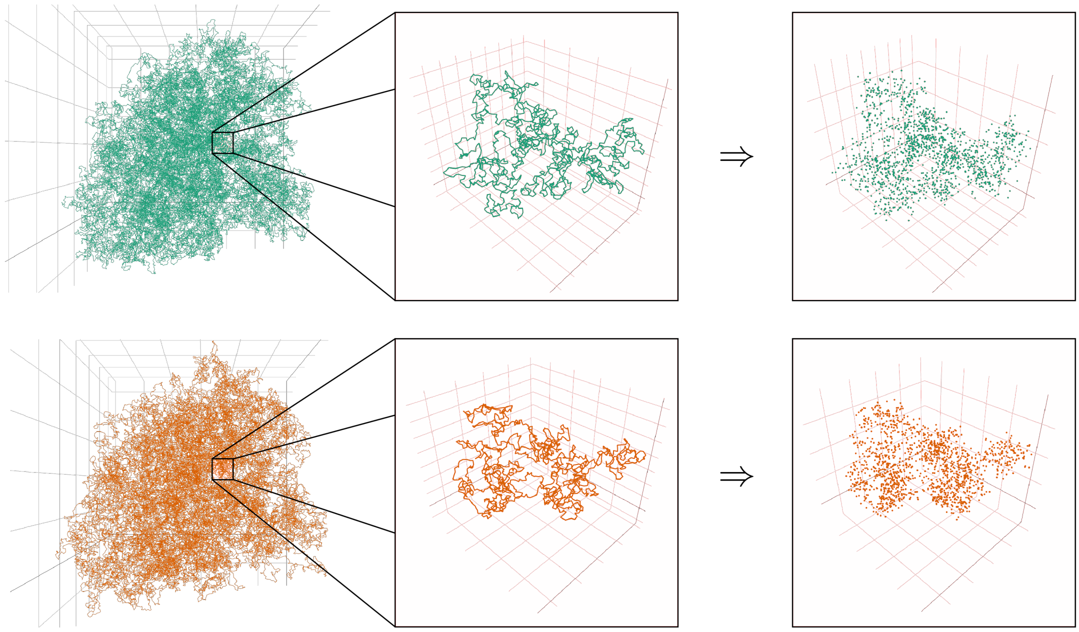

Fig. 9: Chromatin Loops: Upper row: Illustration of a chromatin structure and the corresponding sample with loop density $c = 25$. Lower row: Illustration of the same chromatin and the corresponding sample with loop density $c = 10$.

comparison of the loop density $c = 25$ against $c = 10$ as well as $c = 2$ against $c = 0$ and point out that the comparison between the setting $c = 2$ against $c = 0$ is very difficult as the loop frequencies are very low. The dendrograms in the upper row of Figure 11 illustrate the comparison of $c = 25$ and $c = 10$. It is remarkable that for each m the correct clusters are obtained. The lower row of Figure 11 showcases the dendrograms for the comparison of the estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,2},m}}$ and $\widehat{f}_{\delta^2_{\mathcal{X}_{i,0},m}}$, $1 \leq i \leq 102$ and $m \in \{1/9960, 1/4980, 1/1245\}$. For $m \in \{1/9960, 1/4980\}$, we obtain (up to one exception) the correct clusters, although they are much closer (w.r.t. the L^1-distance) than the clusters for the previous comparisons. However, for $m = 1/1245$, it is no longer possible to reliably identify two clusters that correspond to $c = 2$ and $c = 0$. It seems that in this case m is too large to yield a perfect discrimination.

To conclude this section, we investigate whether classification based on the DTM-density estimates $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$ is possible. Here, we restrict ourselves once again to the comparison of $c = 25$ with $c = 10$ as well as of $c = 2$ with $c = 0$. For each comparison, we randomly select 5%/10% (rounded up) of the density estimates for each considered loop density and classify the remaining ones according to the the majority of the labels of their $k = 1, 3, 5$ nearest neighbors in the randomly selected sample. We repeat this procedure for both comparisons

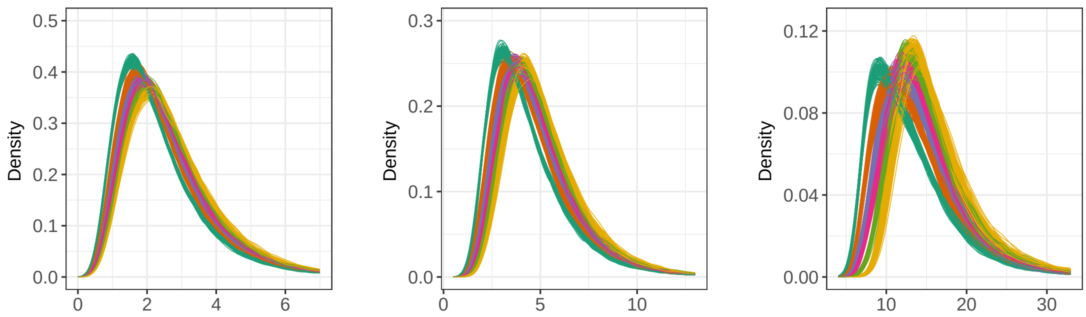

Fig. 10: Chromatin Loop Analysis I: Illustration of the DTM-density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$, $1 \leq i \leq 102$, for $c = 25$ (blue-green), $c = 10$ (orange), $c = 6$ (blue), $c = 4$ (pink), $c = 2$ (green) and $c = 0$ (yellow) and $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right).

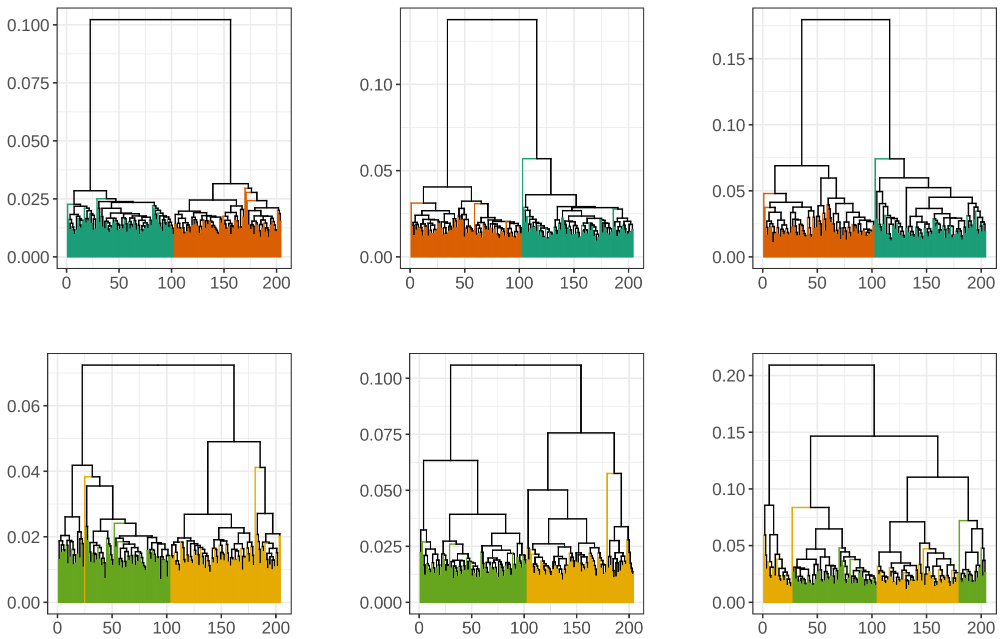

Fig. 11: Chromatin Analysis II: Upper row: The results of an average linkage clustering of the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,25},m}}$ (blue-green) and $\widehat{f}_{\delta^2_{\mathcal{X}_{i,10},m}}$ (orange), $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right) based on the L^1-distance. Lower row: The results of an average linkage clustering of the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,2},m}}$ (green) and $\widehat{f}_{\delta^2_{\mathcal{X}_{i,0},m}}$ (yellow), $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right) based on the L^1-distance.

10,000 times and report the relative number of misclassifications in Table 1. The upper row of said table highlights that in the comparison of $c = 25$ and $c = 10$ the DTM-density estimates are always classified correctly. Things change in the comparison of $c = 2$ with $c = 0$. While for all m at least 90% of the classifications are correct, there is a noticeable difference between the individual values of m. We observe that $m = 1/4980$ yields by far the best performance in this setting. It is clear that the loop distributions of the respective chromatin fibers for these two parameters are extremely similar (the chromatin admits few to no loops). Hence, choosing m too large incorporates too much global (non-loop) structure and makes it difficult to discriminate between these two loop densities. On the other hand, choosing m too small seems to incorporate too little structure.

To conclude, we find that it is possible for a suitable choice of m to clearly distinguish between the different loop densities based on the DTM-density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,c},m}}$. We have illustrated that these estimators yield a good summary of the data and can be used to approach the (already quite difficult) problem of chromatin loop analysis for noisy synthetic data.

	$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$
5%	0.000	0.000	0.000	5%	0.000	0.000	0.000	5%	0.000	0.000	0.000
10%	0.000	0.000	0.000	10%	0.000	0.000	0.000	10%	0.000	0.000	0.000

	$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$		$k = 1$	$k = 3$	$k = 5$
5%	0.029	0.049	0.069	5%	0.012	0.019	0.039	5%	0.029	0.064	0.100
10%	0.020	0.033	0.043	10%	0.001	0.007	0.012	10%	0.008	0.021	0.025

Tab. 1: Chromatin Analysis III: Upper row: The relative number of missclassifications of a k-nearest neighbor classification (w.r.t. the L^1-distance) based on the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,25},m}}$ and $\widehat{f}_{\delta^2_{\mathcal{X}_{i,10},m}}$, $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right). Lower row: The relative number of missclassifications of a k-nearest neighbor classification (w.r.t. the L^1-distance) based on the kernel density estimators $\widehat{f}_{\delta^2_{\mathcal{X}_{i,2},m}}$ and $\widehat{f}_{\delta^2_{\mathcal{X}_{i,0},m}}$, $1 \leq i \leq 102$, for $m \in \{1/9960, 1/4980, 1/1245\}$ (from left to right).

Acknowledgements

C. Weitkamp gratefully acknowledges support by the DFG Research Training Group 2088.

A Proof of Lemma 2.6

In this section, we state the full proof of Lemma 2.6.

Proof of Lemma 2.6. Let $X = (X_1, \dots, X_d) \sim \mu_{\mathcal{X}}$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

1. We observe that

$$\begin{aligned} \mathrm{d}^2_{\mathcal{X},1}(x) = \mathbb{E}\left[||X - x||^2\right] &= \sum_{i=1}^{d} \left(\mathbb{E}\left[X_i^2\right] - 2x_i\mathbb{E}\left[X_i\right] + x_i^2\right) \\ &= \sum_{i=1}^{d} \Big((x_i - \mathbb{E}\left[X_i\right])^2 + \mathbb{E}\left[X_i^2\right] - (\mathbb{E}\left[X_i\right])^2 \Big). \end{aligned}$$

   Setting $c_i = \mathbb{E}\left[X_i\right]$ and $\zeta = \sum_{i=1}^d \left(\mathbb{E}\left[X_i^2\right] - (\mathbb{E}\left[X_i\right])^2\right)$ yields the claim.

2. This follows directly from the fist statement.

3. The fist statement implies that

$$\nabla \mathrm{d}^2_{\mathcal{X},1}(x) = 2(x - \mathbb{E}\left[X\right]).$$

   Clearly, this is zero if and only if $x = \mathbb{E}\left[X\right]$.

4. By the second and third statement $\mathrm{d}^2_{\mathcal{X},1}$ is three times continuously differentiable and $\nabla \mathrm{d}^2_{\mathcal{X},1} > 0$ on ${\mathrm{d}^2_{\mathcal{X},1}}^{-1}([y - 2h_0, y + 2h_0])$. In consequence, there exists an open set $U \supset {\mathrm{d}^2_{\mathcal{X},1}}^{-1}([y - h_0, y + h_0])$ such that the function

$$\varphi : U \subset \mathbb{R}^d \to \mathbb{R}^d;\ x \mapsto \frac{\nabla \mathrm{d}^2_{\mathcal{X},1}(x)}{||\nabla \mathrm{d}^2_{\mathcal{X},1}(x)||^2}$$

   is $C^2(U, \mathbb{R}^d)$. By Theorem 2 in Chapter 15 of Hirsch and Smale [29] there is a unique flow $\Phi^* : [-h_0, h_0] \times W \to \mathbb{R}^d$ with

$$\begin{cases} \frac{\partial}{\partial v}\Phi^*(v, x) = \frac{\nabla \mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v,x))}{||\nabla \mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v,x))||^2} \\ \Phi^*(0, x) = x, \end{cases} \tag{17}$$

   where $W \subset \mathbb{R}^d$ is an open set that contains ${\mathrm{d}^2_{\mathcal{X},1}}^{-1}([y-h_0, y+h_0])$. Differentiating the function $v \mapsto \mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v, x))$ immediately shows that $\mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v, x)) = \mathrm{d}^2_{\mathcal{X},1}(x) + v$. This implies that $\Phi^*(v, {\mathrm{d}^2_{\mathcal{X},1}}^{-1}(\{y\})) = {\mathrm{d}^2_{\mathcal{X},1}}^{-1}(\{y + v\})$. In consequence, it only remains prove that Φ defined in the statement is a solution of the ordinary differential equation (17). For this purpose, we observe that $\Phi(0, x) = x$ for all x. Furthermore, we derive that

$$\frac{\partial}{\partial v}\Phi(v, x) = \left(\frac{x_1 - c_1}{2||x - c||^2\sqrt{\frac{v}{||x-c||^2} + 1}}, \dots, \frac{x_d - c_d}{2||x - c||^2\sqrt{\frac{v}{||x-c||^2} + 1}} \right)^T$$

By the first statement, it follows immediately that

$$\frac{\nabla \mathrm{d}^2_{\mathcal{X},1}(x)}{||\nabla \mathrm{d}^2_{\mathcal{X},1}(x)||^2} = \left(\frac{x_1 - c_1}{2||x-c||^2}, \dots, \frac{x_d - c_d}{2||x-c||^2}\right)^T.$$

In consequence, we find that

$$\frac{\nabla \mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v,x))}{||\nabla \mathrm{d}^2_{\mathcal{X},1}(\Phi^*(v,x))||^2} = \left(\frac{x_1 - c_1}{2||x-c||^2\sqrt{\frac{v}{||x-c||^2}+1}}, \dots, \frac{x_d - c_d}{2||x-c||^2\sqrt{\frac{v}{||x-c||^2}+1}}\right)^T,$$

which proves the fourth statement.

□

B Additional Details on Example 2.7-2.10

In this section, we will provide additional details on the examples considered in Section 2.2. For each example considered, we first briefly recall the setting and derive the corresponding DTM-function and DTM-density.

B.1 Example 2.8

Let $\mathcal{X} = [0,1]$ and let $\mu_{\mathcal{X}}$ denote the uniform distribution on $\mathcal{X}$. Furthermore, we consider two values for m, namely $m_1 = 1$ and $m_2 = 0.1$.

First, we derive $\mathrm{d}^2_{\mathcal{X},1}$ and $f_{\mathrm{d}^2_{\mathcal{X},1}}$. To this end, we observe that for $x \in \mathcal{X}$ and $X \sim \mu_{\mathcal{X}}$

$$\mathrm{d}^2_{\mathcal{X},1}(x) = \int_0^1 F_x^{-1}(t)\,dt = \mathbb{E}\left[(X-x)^2\right] = \frac{1}{3} - x + x^2.$$

This immediately gives that

$$F_{\mathrm{d}^2_{\mathcal{X},1}}(t) = \mathbb{P}\left(\mathrm{d}^2_{\mathcal{X},1}(X) \le t\right) = \begin{cases} 0 & t \le \frac{1}{12}, \\ \sqrt{4t - \frac{1}{3}} & \frac{1}{12} \le t \le \frac{1}{3}, \\ 1 & t > \frac{1}{3}. \end{cases}$$

Hence, $f_{\mathrm{d}^2_{\mathcal{X},1}}$ is given as

$$f_{\mathrm{d}^2_{\mathcal{X},1}}(t) = \begin{cases} \frac{2\sqrt{3}}{\sqrt{12t-1}} & \frac{1}{12} \le t \le \frac{1}{3}, \\ 0 & \text{else}. \end{cases}$$

Next, we come to $m_2 = 0.1$. In order to calculate $\mathrm{d}^2_{\mathcal{X},0.1}$ and $f_{\mathrm{d}^2_{\mathcal{X},0.1}}$, it is necessary to derive the family $\left(F_x^{-1}\right)_{x\in\mathcal{X}}$ explicitly. A short calculation yields that for $0 \le x \le 1/2$

$$F_x^{-1}(y) = \begin{cases} \frac{y^2}{4} & 0 < y \le 2x, \\ (y-x)^2 & 2x < y < 1, \end{cases}$$

and for $1/2 < x \le 1$

$$F_x^{-1}(y) = \begin{cases} \frac{y^2}{4} & 0 < y \le 2(1-x), \\ (y-(1-x))^2 & 2(1-x) < y < 1. \end{cases}$$

Therefore, we find that

$$\mathrm{d}^2_{\mathcal{X},0.1}(x) = \begin{cases} x^2 + 0.1x + \frac{1}{300} & 0 \le x < 0.05, \\ \frac{1}{1200} & 0.05 \le x \le 0.95, \\ (1-x)^2 + 0.1(1-x) + \frac{1}{300} & 0.95 \le x \le 1. \end{cases}$$

Since $\mathrm{d}^2_{\mathcal{X},0.1}$ is constant for $x \in [0.05, 0.95]$ it is immediately clear that the corresponding distribution function $F_{\mathrm{d}^2_{\mathcal{X},0.1}}$ is not continuous. Indeed, we find that

$$F_{\mathrm{d}^2_{\mathcal{X},0.1}}(y) = \mathbb{P}\left(\mathrm{d}^2_{\mathcal{X},0.1}(X) \le y\right) = \begin{cases} 0 & y < \frac{1}{1200}, \\ \frac{1}{20}\sqrt{400y - \frac{1}{3}} + 0.9 & \frac{1}{1200} \le y \le \frac{1}{300}, \\ 1 & y \ge \frac{1}{300}. \end{cases}$$

B.2 Example 2.9

Let $\mathcal{X} = [0,1]$ and let $\mu_{\mathcal{X}}$ denote the probability distribution on $[0,1]$ with density $f(x) = 2x$. Let $m = 0.1$. As previously, we have to explicitly calculate the family $\left(F_x^{-1}\right)_{x\in\mathcal{X}}$. A short calculation shows that for $0 \le x \le 1/2$ we have that

$$F_x^{-1}(y) = \begin{cases} \frac{y^2}{16x^2} & 0 < y \le 4x^2, \\ (\sqrt{y}-x)^2 & 4x^2 < y < 1, \end{cases}$$

and for $1/2 < x \le 1$

$$F_x^{-1}(y) = \begin{cases} \frac{y^2}{16x^2} & 0 < y \le 4x(1-x), \\ (x-\sqrt{1-y})^2 & 4x(1-x) < y < 1. \end{cases}$$

The integration of these function on $[0, 0.1]$, shows that the corresponding DTM-function is given as

$$\mathrm{d}^2_{\mathcal{X},0.1}(x) = \begin{cases} x^2 - \frac{2}{3}\sqrt{\frac{2}{5}}x + \frac{1}{20} & 0 \le x \le \frac{\sqrt{0.1}}{2}, \\ \frac{1}{4800x^2} & \frac{\sqrt{0.1}}{2} \le x \le \frac{1}{2} + \frac{3}{2\sqrt{10}}, \\ x^2 + \left(18\sqrt{\frac{2}{5}} - \frac{40}{3}\right)x + \frac{19}{20} & \frac{1}{2} + \frac{3}{2\sqrt{10}} < x \le 1. \end{cases}$$

It is obvious that in this case $\mathrm{d}^2_{\mathcal{X},0.1}$ is almost nowhere constant. Furthermore, we can now show that

$$F_{\mathrm{d}^2_{\mathcal{X},0.1}}(y) = \begin{cases} 0 & y \le \frac{4320\sqrt{10}-13661}{180}, \\ \frac{2}{45}(20\sqrt{5} - 27\sqrt{2})\sqrt{13661 - 4320\sqrt{10} + 180y} & \frac{4320\sqrt{10}-13661}{180} < y \le \frac{19-6\sqrt{10}}{120}, \\ y - \frac{3}{5}\sqrt{360y - 8640\sqrt{10} + 27322} - \frac{1}{4800y} - 48\sqrt{10} + \frac{27493}{180} \\ \quad + \frac{4}{9}\sqrt{900y - 21600\sqrt{10} + 68305} & \frac{19-6\sqrt{10}}{120} < y \le \frac{1080\sqrt{\frac{2}{5}}-683}{60}, \\ 1 - \frac{1}{4800y} & \frac{1080\sqrt{\frac{2}{5}}-683}{60} < y \le \frac{1}{120} \\ -y + \frac{1}{45}\sqrt{360y - 2} + \frac{173}{180} & \frac{1}{120} < y \le \frac{1}{20} \\ 1 & y > \frac{1}{20}. \end{cases}$$

This allows us to derive that

$$f_{\mathrm{d}^2_{\mathcal{X},0.1}}(y) = \begin{cases} \frac{80\sqrt{5}-108\sqrt{2}}{\sqrt{13661-4320\sqrt{10}+180y}} & \frac{4320\sqrt{10}-13661}{180} < y \le \frac{19-6\sqrt{10}}{120}, \\ 1 + \frac{40\sqrt{5}-54\sqrt{2}}{\sqrt{13661-4320\sqrt{10}+180y}} + \frac{1}{4800y^2} & \frac{19-6\sqrt{10}}{120} < y \le \frac{1080\sqrt{\frac{2}{5}}-683}{60}, \\ \frac{1}{4800y^2} & \frac{1080\sqrt{\frac{2}{5}}-683}{60} < y \le \frac{1}{120} \\ -1 + \frac{2}{\sqrt{-\frac{1}{2}+90y}} & \frac{1}{120} < y \le \frac{1}{20} \\ 0 & \text{else.} \end{cases}$$

B.3 Example 2.7

Let $\mathcal{X} = [0, 1]^2$ and let $\mu_{\mathcal{X}}$ stand for the uniform distribution on $\mathcal{X}$. Choose $m = 1$ and let $\mathcal{X} \sim \mu_{\mathcal{X}}$. Then, it is possible to derive that for $x = (x_1, x_2) \in \mathcal{X}$

$$\mathrm{d}^2_{\mathcal{X},1}(x) = \mathbb{E}\left[||X - x||^2\right] = x_1^2 + x_2^2 - x_1 - x_2 + \frac{2}{3}.$$

Hence, we find that

$$F_{\mathrm{d}^2_{\mathcal{X},1}}(y) = \begin{cases} 0 & y \leq \frac{1}{6}, \\ \pi\left(y-\frac{1}{6}\right) & \frac{1}{6} < y \leq \frac{5}{12}, \\ \frac{1}{3}\Big(\sqrt{36y-15}+(6y-1)\operatorname{arccot}\Big(2\sqrt{y-\frac{5}{12}}\Big) \\ \quad +(1-6y)\arctan\Big(\sqrt{4y-\frac{5}{3}}\Big)\Big) & \frac{5}{12} < y \leq \frac{2}{3}, \\ 1 & y > \frac{2}{3}. \end{cases}$$

The corresponding density is given as

$$f_{\mathrm{d}^2_{\mathcal{X},1}}(y) = \begin{cases} \pi & \frac{1}{6} \leq y \leq \frac{5}{12}, \\ 2\operatorname{arccot}\left(2\sqrt{y-\frac{5}{12}}\right) - 2\arctan\left(\sqrt{4y-\frac{5}{3}}\right) & \frac{5}{12} < y \leq \frac{2}{3}, \\ 0 & \text{else.} \end{cases}$$

B.4 Example 2.10

Let $\mathcal{X}$ denote a disk in $\mathbb{R}^2$ centered at $(0,0)$ with radius 1 and let $\mu_{\mathcal{X}}$ denote probability measure with density

$$f(x_1,x_2) = \begin{cases} -\frac{2}{\pi}\left(x_1^2+x_2^2-1\right) & x_1^2+x_2^2 \leq 1, \\ 0 & \text{else.} \end{cases}$$

In this case, it is for $m=1$ straight forward to derive that for $x=(x_1,x_2)\in\mathcal{X}$

$$\mathrm{d}^2_{\mathcal{X},1}(x) = x_1^2+x_2^2+\frac{1}{3}. \tag{18}$$

In consequence, we find that for $X \sim \mu_{\mathcal{X}}$

$$F_{\mathrm{d}^2_{\mathcal{X},1}}(y) = \begin{cases} 0 & y \leq \frac{1}{3}, \\ -y^2+\frac{8}{3}y-\frac{7}{9} & \frac{1}{3} < y \leq \frac{4}{3}, \\ 1 & \text{else.} \end{cases}$$

The corresponding density is given as

$$f_{\mathrm{d}^2_{\mathcal{X},1}}(y) = \begin{cases} -2y+\frac{8}{3} & \frac{1}{3} < y \leq \frac{4}{3}, \\ 0 & \text{else.} \end{cases} \tag{19}$$

C Proof of Theorem 2.12

In this section, we give the full proof of Theorem 2.12. The proof is composed of four steps, each of which formulated as an independent lemma (see Section C.1).

Step 1: Replacement of $\delta^2_{\mathcal{X},m}(X_i)$ by $\mathrm{d}^2_{\mathcal{X},m}(X_i)$ (Lemma C.1).
We provide a decomposition of $\widehat{f}_{\delta^2_{\mathcal{X},m}}$ in a sum of two leading terms in which $\delta^2_{\mathcal{X},m}(X_i)$ is replaced by $\mathrm{d}^2_{\mathcal{X},m}(X_i)$ in the argument of the kernel K and we show that the remainder terms are negligible.

Step 2: Introducing U-statistics (Lemma C.2).
It is shown that the leading terms obtained in Step 1 can be written as a (sum of two) U-statistic(s) asymptotically.

Step 3: Hoeffding decomposition (Lemma C.4).
Applying a Hoeffding decomposition allows to derive a representation of the (sum of two) U-statistic(s) of step 2 as a sum of a deterministic term (expectation), a stochastic leading term consisting of a sum of independent random variables and a remainder term.

Step 4: CLT for the leading term of Step 3 (Lemma C.7).
Since the leading term of Step 3 is a sum of centered independent random variables, we can apply a standard CLT to show its asymptotic normality.

C.1 Auxiliary lemmas representing Step 1 - Step 4

Before we come to the proof of Theorem 2.12, we will establish several auxiliary results. In order to highlight the overall proof strategy, the corresponding proofs are deferred to Section C.3. We begin this section, by addressing Step 1.

Lemma C.1 (Step 1). *Assume that Setting 2.1 holds and let Condition 2.2 be met. Then, it follows that*

$$\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right) + \frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)A_n(X_i)\right] + \mathcal{O}_P\left(\frac{1}{nh^3}\right) + o_P\left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h}\right),$$

where

$$A_n(x) := \frac{1}{m}\int_0^{F_x^{-1}(m)} F_x(t) - \widehat{F}_{x,n}(t)\,dt.$$

As a direct consequence of Lemma C.1, we find that for $h \in o\left(1/n^{1/5}\right)$ the statistic

$$V_n(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right) + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)A_n(X_i)$$
$$=: V_n^{(1)}(y) + V_n^{(2)}(y) \tag{20}$$

drives the limit behavior of $\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right)$. Next, we will establish that the statistic $V_n(y)$ can, up to asymptotically negligible terms, be written as a U-statistic (see e.g. Van der Vaart [52] for more information on U-statistics).

Lemma C.2 (Introduction of U-statistics, Step 2)**.** *Assume Setting 2.1 and let* $V_n^{(1)}$ *and* $V_n^{(2)}$ *be as defined in* (20)*. Then, we have*

$$V_n^{(1)}(y) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^{(1)}_{y,h}(X_i, X_j),$$

where

$$g^{(1)}_{y,h}(x_1,x_2) = \frac{1}{2h}\left(K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right) + K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_2)-y}{h}\right)\right).$$

Furthermore,

$$V_n^{(2)}(y) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^{(2)}_{y,h}(X_i, X_j) + \mathcal{O}_P\left(\frac{1}{nh^2}\right),$$

where

$$g^{(2)}_{y,h}(x_1,x_2) = \frac{1}{2mh^2}\Bigg[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)\int_0^{F^{-1}_{x_1}(m)} F_{x_1}(t) - \mathbb{1}_{\{||x_1-x_2||^2\le t\}}\,dt$$
$$+ K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_2)-y}{h}\right)\int_0^{F^{-1}_{x_2}(m)} F_{x_2}(t) - \mathbb{1}_{\{||x_1-x_2||^2\le t\}}\,dt\Bigg].$$

Remark C.3. It is important to note that $g^{(1)}_{y,h}$ and $g^{(2)}_{y,h}$ are symmetric by definition, i.e., $V_n^{(1)}(x)$ is a U-statistic and $V_n^{(2)}(x)$ can be decomposed into a U-statistic and an asymptotically negligible remainder term.

Combining Lemma C.1 and Lemma C.2, we see that

$$\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) = U_n + \mathcal{O}_P\left(\frac{1}{nh^3}\right) + o_P\left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h}\right), \tag{21}$$

where $U_n = U_n(y)$ denotes the U-statistic with kernel function $g_{y,h}(x_1, x_2) \coloneqq g^{(1)}_{y,h}(x_1, x_2) + g^{(2)}_{y,h}(x_1, x_2)$. Before we use (21) to finalize the proof of Theorem 2.12, we establish two further auxiliary results. Next, we rewrite U_n using the Hoeffding decomposition (see Van der Vaart [52, Sec. 11.4]), which is the key ingredient to handling the stochastic dependencies introduced by the terms $A_n(X_i)$.

Lemma C.4 (Hoeffding decomposition, Step 3)**.** *Assume Setting 2.1. Let U_n be the U-statistic with kernel function $g_{y,h}(x_1, x_2) = g^{(1)}_{y,h}(x_1, x_2) + g^{(2)}_{y,h}(x_1, x_2)$. Then, it follows that*

$$U_n = \Theta_{y,h} + \frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i) + \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} g_{y,h,2}(X_i, X_j).$$

Here, we have that

$$\Theta_{y,h} = \int K\left(v\right) f_{\mathrm{d}^2_{\mathcal{X},m}}(x + vh)\, dv.$$

Furthermore, let $Z_1, Z_2 \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$. Then, it holds that

$$g_{y,h,1}(x_1) = \frac{1}{2h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1) - y}{h}\right) - \frac{1}{2}\Theta_{y,h} + \mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1) - y}{h}\right)\Psi(x_1, Z_1)\right],$$

where

$$\Psi(x_1, x_2) \coloneqq ||x_1 - x_2||^2 \wedge F^{-1}_{x_2}(m) - \mathbb{E}_{Z_2}\left[||Z_2 - x_2||^2 \wedge F^{-1}_{x_2}(m)\right], \tag{22}$$

and

$$g_{y,h,2}(x_1, x_2) = g_{y,h}(x_1, x_2) - g_{y,h,1}(x_1) - g_{y,h,1}(x_2) - \Theta_{y,h}. \tag{23}$$

Remark C.5. It is well known that the mean zero random variables $(g_{y,h,2}(X_i, X_j))_{1\leq i<j\leq n}$ are uncorrelated (see Van der Vaart [52, Sec. 11.4]).

For our later considerations it is important to derive a certain regularity for the function Ψ defined in (22).

Lemma C.6. *Assume Setting 2.1 and let Ψ be the function defined in (22). Then, the function $z \mapsto \Psi(x_1, z)$ is Lipschitz continuous for all $x_1 \in \mathcal{X}$ and the corresponding Lipschitz constant does not depend on the choice of x_1, i.e., it holds*

$$|\Psi(x_1, z_1) - \Psi(x_1, z_2)| \leq C||z_1 - z_2||,$$

for all $z_1, z_2 \in \mathcal{X}$, where the constant $0 < C < \infty$ does not depend on x_1.

The next step in the proof of Theorem 2.12 is to derive for $n \to \infty$ and $h \to 0$ the limit distribution of $\sqrt{nh}\big(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i)\big)$.

Lemma C.7. *Assume Setting 2.1 and let Condition 2.2 be met. Let $n \to \infty, h \to 0$ such that $nh \to \infty$ and recall that*

$$g_{y,h,1}(x_1) = \frac{1}{2h} K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1) - y}{h}\right) - \frac{1}{2}\Theta_{y,h} + \mathbb{E}_{Z_1}\left[\frac{1}{2mh^2} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1) - y}{h}\right)\Psi(x_1, Z_1)\right],$$

It holds

$$\frac{2\sqrt{h}}{\sqrt{n}} \sum_{i=1}^{n} g_{y,h,1}(X_i) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y) \int K^2(u)\, du\right). \tag{24}$$

With all auxiliary results required established, we can finally come to the proof of Theorem 2.12. The proof strategy is to demonstrate that the limit of $\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right)$ coincides with the limit of $\frac{2\sqrt{h}}{\sqrt{n}} \sum_{i=1}^{n} g_{y,h,1}(X_i)$.

C.2 Proof of Theorem 2.12

The proof of Theorem 2.12 is now a consequence of the lemmas provided in the previous subsection.

Proof of Theorem 2.12. We find that

$$\begin{aligned}
&\left|\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) - \sqrt{nh}\left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i)\right)\right| \\
\leq \sqrt{nh}&\left|\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - \left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i) + \Theta_{y,h}\right)\right| + \sqrt{nh}\left|f_{\mathrm{d}^2_{\mathcal{X},m}}(y) - \Theta_{y,h}\right|.
\end{aligned} \tag{25}$$

In the following, we consider both summands separately.

First summand: For the first summand, we obtain that

$$\begin{aligned}
\mathcal{S}_n(y) &= \sqrt{nh}\left|\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - \left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i) + \Theta_{y,h}\right)\right| \\
&\leq \sqrt{nh}\left(\left|\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - V_n(y)\right| + \left|V_n(y) - U_n(y)\right| + \left|U_n(y) - \left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i) + \Theta_{y,h}\right)\right|\right),
\end{aligned}$$

where $V_n(y)$ and $U_n(y)$ are defined in(20) and (21) respectively. By Lemma C.1 and $h = o\left(n^{-1/5}\right)$ we obtain that

$$\sqrt{nh}\left|\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - V_n(y)\right| = \mathcal{O}_P\left(\frac{\sqrt{nh}}{nh^3}\right) + o_P\left(\frac{\sqrt{nh}\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}h}\right) = o_P(1).$$

Similarly, we get by Lemma C.2 and $h = o\left(n^{-1/5}\right)$ that

$$\sqrt{nh}\left|V_n(y) - U_n(y)\right| = \mathcal{O}_P\left(\frac{\sqrt{nh}}{nh^2}\right) = o_P(1).$$

Hence, it remains to consider

$$\sqrt{nh}\left|U_n(y) - \left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i) + \Theta_{y,h}\right)\right| = \sqrt{nh}\left|\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g_{y,h,2}(X_i, X_j)\right|,$$

where the last equality follows by Lemma C.4. Considering the definition of $g_{y,h,2}(x_1, x_2)$ in (23), we recognize that $g_{y,h,2}(x_1, x_2) \in \mathcal{O}\left(\frac{1}{h^2}\right)$, as $h \to 0$. Let now $g^*_{y,2}(x_1, x_2) = h^2 g_{y,h,2}(x_1, x_2)$. Then, $g^*_{y,h,2}(x_1, x_2) = \mathcal{O}(1)$, as $h \to 0$. Furthermore, we have by Remark C.5 that the random variables $\{g_{y,h,2}(X_i, X_j)\}_{1\le i<j\le n}$ are uncorrelated, whence the same holds for the random variables $\{g^*_{y,h,2}(X_i, X_j)\}_{1\le i<j\le n}$. In consequence, we obtain that

$$\operatorname{Var}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^*_{y,2}(X_i, X_j)\right) = \mathcal{O}\left(n^{-2}\right).$$

This in turn implies by Chebyshev's inequality that

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^*_{y,2}(X_i, X_j) = \mathcal{O}_P(n^{-1}).$$

Therefore, we obtain with $h = o\left(n^{-1/5}\right)$ that

$$\begin{aligned}\frac{2\sqrt{h}}{\sqrt{n}(n-1)}\sum_{1\le i<j\le n} g_{y,h,2}(X_i, X_j) =& \frac{\sqrt{nh}}{h^2}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^*_{y,2}(X_i, X_j)\right)\\ =&\mathcal{O}_P\left(\frac{1}{n^{1/2}h^{1.5}}\right) = o_P(1).\end{aligned}$$

Thus, we have shown that $\mathcal{S}(y) = o_P(1)$.
Second summand: Finally, we come to the second summand in (25). First of all, we observe that

$$\Theta_{y,h} = \int K(v)\, f_{\mathrm{d}^2_{\mathcal{X},m}}(x + vh)\, dv = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right)\right],$$

where $\{\mathrm{d}^2_{\mathcal{X},m}(X_i)\}_{i=1}^n$ is a collection of i.i.d. random variables with density $f_{\mathrm{d}^2_{\mathcal{X},m}}$. Since $f_{\mathrm{d}^2_{\mathcal{X},m}}$ is assumed to be twice differentiable on $(y-\epsilon, y+\epsilon)$ and K is symmetric, i.e., $\int uK(u)\,du = 0$, it follows by a straight forward adaptation of Proposition 1.2 of Tsybakov [51] that

$$\left|\Theta_{y,h} - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right| \leq Ch^2.$$

Here, C denotes a constant independent of n and h. We get that

$$\sqrt{nh}\left|\Theta_{y,h} - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right| = \mathcal{O}_P(\sqrt{nh^5}) = o_P(1),$$

as $h = o\left(n^{-1/5}\right)$.

In conclusion, we have shown that

$$\left|\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) - \sqrt{nh}\left(\frac{2}{n}\sum_{i=1}^n g_{y,h,1}(X_i)\right)\right| = o_P(1),$$

which yields that

$$\sqrt{nh}\left(\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) - f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\right) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\int_{\mathbb{R}} K^2(u)\,du\right)$$

as claimed. □

C.3 Proofs of the Auxiliary Lemmas from Section C.1

In this section, we gather the full proofs of Lemma C.1-C.7.

C.3.1 Proof of Lemma C.1

In the course of this proof we have to differentiate between the cases $0 < m < 1$ and $m = 1$.

The case $0 < m < 1$*:* By assumption the kernel K is twice continuously differentiable. Using a Taylor series approximation, we find that

$$K\left(\frac{\delta^2_{\mathcal{X},m}(X_i) - y}{h}\right) = K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right) + \frac{1}{h}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right)\left(\delta^2_{\mathcal{X},m}(X_i) - \mathrm{d}^2_{\mathcal{X},m}(X_i)\right)$$
$$+ \frac{1}{2h^2}K''\left(\frac{\zeta - y}{h}\right)\left(\delta^2_{\mathcal{X},m}(X_i) - \mathrm{d}^2_{\mathcal{X},m}(X_i)\right)^2,$$

for some ζ_i between $\mathrm{d}^2_{\mathcal{X},m}(X_i)$ and $\delta^2_{\mathcal{X},m}(X_i)$. By Theorem 9 in Chazal et al. [14] (whose conditions are met by assumption) it holds

$$\sup_{x\in\mathcal{X}} \left|\delta^2_{\mathcal{X},m}(x) - \mathrm{d}^2_{\mathcal{X},m}(x)\right| = \mathcal{O}_P(1/\sqrt{n}). \tag{26}$$

In consequence, we obtain that

$$\begin{aligned}\widehat{f}_{\delta^2_{\mathcal{X},m}}(y) =& \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\delta^2_{\mathcal{X},m}(X_i)-y}{h}\right)\\ =& \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)+\frac{1}{h}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\left(\delta^2_{\mathcal{X},m}(X_i)-\mathrm{d}^2_{\mathcal{X},m}(X_i)\right)\right]\\ &+\mathcal{O}_P(1/(nh^3)).\end{aligned}$$

Furthermore, it has been shown (see the proof of Theorem 5 in Chazal et al. [14]) that for any $x \in \mathcal{X}$

$$\begin{aligned}\delta^2_{\mathcal{X},m}(x) - \mathrm{d}^2_{\mathcal{X},m}(x) = \frac{1}{m}\int_0^{F_x^{-1}(m)} F_x(t) - \widehat{F}_{x,n}(t)\,dt + \frac{1}{m}\int_{F_x^{-1}(m)}^{\widehat{F}_{x,n}^{-1}(m)} m - \widehat{F}_{x,n}(t)\,dt\\ =: A_n(x) + R_n(x).\end{aligned} \tag{27}$$

In consequence, it remains to estimate $\sup_{x\in\mathcal{X}}|R_n(x)|$. Clearly, we have that

$$|R_n(x)| \leq \frac{1}{m}|S_n(x)|\,|T_n(x)|, \tag{28}$$

where

$$S_n(x) = \left|F_x^{-1}(m) - \widehat{F}_{x,n}^{-1}(m)\right| \text{ and } T_n(x) = \sup_t \left|F_x(t) - \widehat{F}_{x,n}(t)\right|.$$

Claim 1: It holds that

$$\sup_{x\in\mathcal{X}}|S_n(x)| = o_P\left(\left(\frac{\log(n)}{n}\right)^{1/(2b)}\right) \text{ as well as } \sup_{x\in\mathcal{X}}|T_n(x)| = \mathcal{O}_P\left(\sqrt{\frac{d}{n}}\right),$$

where $1 \leq b < 5$.

Combining Claim 1 with (28) yields $\sup_{x\in\mathcal{X}}|R_n(x)| = o_P\left(\frac{\log(n)^{1/(2b)}}{n^{1/2+1/(2b)}}\right)$, which gives the statement for $0 < m < 1$.

Proof of Claim 1: It has already been established in the proof of Theorem 9 in Chazal et al. [14] that under the assumptions made

$$\sup_{x\in\mathcal{X}}|T_n(x)| = \mathcal{O}_P\left(\sqrt{\frac{d}{n}}\right).$$

Hence, it only remains to demonstrate the first equality. For this purpose, let $\xi_i \overset{i.i.d.}{\sim}$ Uniform(0,1), $1 \leq i \leq n$, and denote by H_n their empirical distribution function. Define $k = mn$. Then, it holds that $\widehat{F}_{x,n}^{-1}(m) \overset{\mathcal{D}}{=} F_x^{-1}(\xi_{(k)}) = F_x^{-1}\left(H_n^{-1}(m)\right)$. Here, $\xi_{(k)}$ is the k-th order statistic and $\overset{\mathcal{D}}{=}$ denotes equality in deistribution. Hence, we have for any $m > 0$ and $x \in \mathcal{X}$ that

$$\begin{aligned}\mathbb{P}\left(|S_n(x)| > \epsilon\right) =& \mathbb{P}\left(\left|F_x^{-1}\left(H_n^{-1}(m)\right) - F_x^{-1}(m)\right| > \epsilon\right)\\ \leq& \mathbb{P}\left(\omega_x\left(\left|m - H_n^{-1}(m)\right|\right) > \epsilon\right),\end{aligned}$$

where ω_x denote the modulus of continuity for F_x^{-1}. This means that for $u \in (0,1)$

$$\omega_x(u) \coloneqq \sup_{t,t' \in (0,1)^2, |t-t'|<u} \left|F_x^{-1}(t) - F_x^{-1}(t')\right|.$$

By assumption, there exists a constant $\kappa \in \mathbb{R}$ such that $\omega_{\mathcal{X}}(u) = \sup_{x\in\mathcal{X}} \omega_x(u) \leq \kappa u^{1/b}$ for all $u \in (0,1)$. Hence, we find that

$$\begin{aligned}\mathbb{P}\left(\omega_x\left(\left|m - H_n^{-1}(m)\right|\right) > \epsilon\right) \leq& \mathbb{P}\left(\left|m - H_n^{-1}(m)\right| > \left(\frac{\epsilon}{\kappa}\right)^b\right)\\ \leq& 2\exp\left(-\frac{n\left(\frac{\epsilon}{\kappa}\right)^{2b}}{m}\frac{1}{1+\frac{2\left(\frac{\epsilon}{\kappa}\right)^b}{3m}}\right),\end{aligned} \tag{29}$$

where the last line follows from Shorack and Wellner [48] (Inequality 1 on Page 453 and Proposition 1, page 455). Next, we observe that

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} |S_n(x)| > \epsilon\right) \leq \mathbb{P}\left(\sup_{x\in\mathcal{X}} \omega_x\left(|m - H_n^{-1}(m)|\right) > \epsilon\right) \leq \mathbb{P}\left(\kappa|m - H_n^{-1}(m)|^{1/b} > \epsilon\right).$$

Using (29), we find that

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} |S_n(x)| > \epsilon\right) \leq 2\exp\left(-\frac{n\left(\frac{\epsilon}{\kappa}\right)^{2b}}{m}\frac{1}{1+\frac{2\left(\frac{\epsilon}{\kappa}\right)^b}{3m}}\right) \leq 2 \quad \exp\left(-\frac{n\left(\frac{\epsilon}{\kappa}\right)^{2b}}{m}\frac{1}{1+\frac{2}{3m}}\right),$$

since $\epsilon/\kappa < 1$ for epsilon small enough. As $m \leq 1$, we find that

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} |S_n(x)| > \epsilon\right) \leq 2\exp\left(-\frac{3n}{5}\left(\frac{\epsilon}{\kappa}\right)^{2b}\right).$$

Let now $\epsilon = \tau\left(\frac{\log(n)}{n}\right)^{1/(2b)}$. It follows that

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}} |S_n(x)| > \epsilon\right) \leq 2\exp\left(-\frac{3}{5}\left(\frac{\tau}{\kappa}\right)^{2b}\log(n)\right),$$

and thus

$$\sup_{x\in\mathcal{X}} |S_n(x)| = o_P\left(\frac{\log(n)}{n}\right)^{1/(2b)},$$

which yields Claim 1.

The case $m = 1$*:* Similar as for $0 < m < 1$, we find that

$$K\left(\frac{\delta^2_{\mathcal{X},1}(X_i) - y}{h}\right) = K\left(\frac{\mathrm{d}^2_{\mathcal{X},1}(X_i) - y}{h}\right) + \frac{1}{h}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},1}(X_i) - y}{h}\right)\left(\delta^2_{\mathcal{X},1}(X_i) - \mathrm{d}^2_{\mathcal{X},1}(X_i)\right)$$
$$+\frac{1}{2h^2}K''\left(\frac{\zeta_i - y}{h}\right)\left(\delta^2_{\mathcal{X},1}(X_i) - \mathrm{d}^2_{\mathcal{X},1}(X_i)\right)^2,$$

for some ζ_i between $\mathrm{d}^2_{\mathcal{X},1}(X_i)$ and $\delta^2_{\mathcal{X},1}(X_i)$. By Lemma E.1 we obtain

$$\sup_{x\in\mathcal{X}} \left|\delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x)\right| = \mathcal{O}_P(1/\sqrt{n}). \tag{30}$$

Furthermore, we note that for $x \in \mathcal{X}$

$$\delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x) = \int_0^{D_x} F_x(t) - \widehat{F}_{x,n}(t)\,dt, \tag{31}$$

where $[0, D_x]$ denotes the support of F_x. In combination with our previous considerations, we find that

$$\widehat{f}_{\delta^2_{\mathcal{X},1}}(y) = \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{\mathrm{d}^2_{\mathcal{X},1}(X_i) - y}{h}\right) + \frac{1}{h}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},1}(X_i) - y}{h}\right)A_n(x)\right] + \mathcal{O}_P\left(\frac{1}{nh^3}\right),$$

which yields the claim.

■

C.3.2 Proof of Lemma C.2

First, we consider $V_n^{(1)}(x)$. Clearly, we have

$$\begin{aligned}
V_n^{(1)}(y) =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right)\\
=& \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\frac{1}{2h}\left(K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i) - y}{h}\right) + K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_j) - y}{h}\right)\right)\\
=& \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} g^{(1)}_{y,h}(X_i, X_j).
\end{aligned}$$

Next, we come to $V_n^{(2)}(y)$. We have that

$$\begin{aligned}
V_n^{(2)}(y) =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)A_n(X_i)\\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\frac{1}{m}\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\widehat{F}_{X_i,n}(t)\,dt\\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\frac{1}{m}\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\{||X_i-X_j||^2\le t\}}\,dt\\
=& \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\frac{1}{m}\sum_{j=1}^{n}\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\mathbb{1}_{\{||X_i-X_j||^2\le t\}}\,dt.
\end{aligned}$$

Further, we obtain that

$$\begin{aligned}
V_n^{(2)}(y) =& \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\mathbb{1}_{\{||X_i-X_j||^2\le t\}}\,dt\\
=& \frac{1}{n^2}\sum_{1\le i<j\le n}\frac{1}{mh^2}\left[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\mathbb{1}_{\{||X_i-X_j||^2\le t\}}\,dt\right.\\
&\left.+K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_j)-y}{h}\right)\int_0^{F_{X_j}^{-1}(m)}F_{X_j}(t)-\mathbb{1}_{\{||X_j-X_i||^2\le t\}}\,dt\right]\\
&+\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{mh^2}\left[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-1\,dt\right].
\end{aligned}$$

We note that K is twice differentiable and $\mathcal{X}$ is compact, i.e.,

$$\left|\int_0^{F_{x_1}^{-1}(m)}F_{x_2}(t)-\mathbb{1}_{\{||x_1-x_2||^2\le t\}}\,dt\right|\le \operatorname{diam}(\mathcal{X})<\infty$$

for all $x_1,x_2\in\mathcal{X}$. This yields that

$$\begin{aligned}
V_n^{(2)}(y)=& \frac{2}{n^2}\sum_{1\le i<j\le n}\frac{1}{2mh^2}\left[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_i)-y}{h}\right)\int_0^{F_{X_i}^{-1}(m)}F_{X_i}(t)-\mathbb{1}_{\{||X_i-X_j||^2\le t\}}\,dt\right.\\
&\left.+K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_j)-y}{h}\right)\int_0^{F_{X_j}^{-1}(m)}F_{X_j}(t)-\mathbb{1}_{\{||X_j-X_i||^2\le t\}}\,dt\right]+\mathcal{O}_P\left(\frac{1}{nh^2}\right)\\
=& \frac{n-1}{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n}g_{y,h}^{(2)}(X_i,X_j)\right)+\mathcal{O}_P\left(\frac{1}{nh^2}\right)
\end{aligned}$$

$$=\frac{2}{n(n-1)}\sum_{1\le i<j\le n} g^{(2)}_{y,h}(X_i,X_j)+\mathcal{O}_P\left(\frac{1}{nh^2}\right).$$

■

C.3.3 Proof of Lemma C.4

Since $X_1,\dots X_n \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$, the claim follows by the Hoeffding decomposition [52, Lemma 11.11] once we have shown that

1. $\Theta_{y,h}=\mathbb{E}\left[U_n\right]$
2. $g_{y,h,1}(x_1)=\mathbb{E}\left[g_{y,h}(x_1,Z_1)\right]-\Theta_{y,h}$, where $Z_1\sim\mu_{\mathcal{X}}$.

First equality: We start by verifying the first equality. Clearly,

$$\mathbb{E}\left[U_n\right]=\mathbb{E}\left[g^{(1)}_{y,h}(X_1,X_2)\right]+\mathbb{E}\left[g^{(2)}_{y,h}(X_1,X_2)\right].$$

Since $X_1,X_2 \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$, we obtain that

$$\begin{aligned}\mathbb{E}\left[g^{(1)}_{y,h}(X_1,X_2)\right]=&\mathbb{E}\left[\frac{1}{2h}\left(K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)+K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_2)-y}{h}\right)\right)\right]\\ =&\mathbb{E}\left[\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\right]=\int\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(z)-y}{h}\right)d\mu_{\mathcal{X}}(z)\\ =&\int\frac{1}{h}K\left(\frac{u-y}{h}\right)d(\mathrm{d}^2_{\mathcal{X},m}\#\mu_{\mathcal{X}})(u).\end{aligned}$$

Here, the last equality follows by the change-of-variables formula ($\mathrm{d}^2_{\mathcal{X},m}\#\mu_{\mathcal{X}}$ denotes the pushforward measure of $\mu_{\mathcal{X}}$ with respect to $\mathrm{d}^2_{\mathcal{X},m}$). By assumption the measure $\mathrm{d}^2_{\mathcal{X},m}\#\mu_{\mathcal{X}}$ possesses a density $f_{\mathrm{d}^2_{\mathcal{X},m}}$ with respect to the Lebesgue measure. Hence,

$$\begin{aligned}\mathbb{E}\left[g^{(1)}_{y,h}(X_1,X_2)\right]&=\int\frac{1}{h}K\left(\frac{u-y}{h}\right)f_{\mathrm{d}^2_{\mathcal{X},m}}(u)\,du\\ &=\int K\left(v\right)f_{\mathrm{d}^2_{\mathcal{X},m}}(y+vh)\,dv.\end{aligned}$$

As $X_1, X_2 \overset{i.i.d.}{\sim} \mu_{\mathcal{X}}$, we obtain for the second summand that

$$\begin{aligned}\mathbb{E}\left[g^{(2)}_{y,h}(X_1,X_2)\right] =&\mathbb{E}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\int_0^{F^{-1}_{X_1}(m)} F_{X_1}(t)-\mathbb{1}_{\{||X_1-X_2||^2\leq t\}}\,dt\right]\\ &+\mathbb{E}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_2)-y}{h}\right)\int_0^{F^{-1}_{X_2}(m)} F_{X_2}(t)-\mathbb{1}_{\{||X_2-X_1||^2\leq t\}}\,dt\right]\\ =&\mathbb{E}\left[\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\int_0^{F^{-1}_{X_1}(m)} F_{X_1}(t)-\mathbb{1}_{\{||X_1-X_2||^2\leq t\}}\,dt\right]\end{aligned}$$

Since X_1 and X_2 are independent, we further find that

$$\begin{aligned}\mathbb{E}\left[g^{(2)}_{y,h}(X_1,X_2)\right]=&\mathbb{E}_{X_1}\left[\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\mathbb{E}_{X_2}\left[\int_0^{F^{-1}_{X_1}(m)} F_{X_1}(t)-\mathbb{1}_{\{||X_1-X_2||^2\leq t\}}dt\right]\right]\\ \overset{(i)}{=}&\mathbb{E}_{X_1}\left[\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\int_0^{F^{-1}_{X_1}(m)} F_{X_1}(t)-\mathbb{E}_{X_2}\left[\mathbb{1}_{\{||X_1-X_2||^2\leq t\}}\right]dt\right]\\ =&\mathbb{E}_{X_1}\left[\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right)\int_0^{F^{-1}_{X_1}(m)} F_{X_1}(t)-F_{X_1}(t)\,dt\right]\\ =&0,\end{aligned}$$

where (i) follows by the Theorem of Tonelli/Fubini [6, Thm. 18]. Combining our results, we find that $\mathbb{E}[U_n] = \Theta_{y,h}$.

Second equality: Recall that $Z_1 \sim \mu_{\mathcal{X}}$. We demonstrate that

$$g_{y,h,1}(x_1) = \mathbb{E}\left[g_{y,h}(x_1,Z_1)\right] - \Theta_{y,h} = \mathbb{E}\left[g^{(1)}_{y,h}(x_1,Z_1)\right] + \mathbb{E}\left[g^{(2)}_{y,h}(x_1,Z_1)\right] - \Theta_{y,h}.$$

Once again, we consider the two summands separately. We observe that

$$\begin{aligned}\mathbb{E}\left[g^{(1)}_{y,h}(x_1,Z_1)\right] =&\mathbb{E}\left[\frac{1}{2h}\left(K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)+K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\right)\right]\\ =&\frac{1}{2h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)+\frac{1}{2}\Theta_{y,h}.\end{aligned}$$

Here, the last equality follows by our previous considerations for $\mathbb{E}[U_n]$. For the second summand, it follows that

$$\begin{aligned}\mathbb{E}\left[g_{y,h}^{(2)}(x_1,Z_1)\right] =&\mathbb{E}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)\int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t)-\mathbb{1}_{\{||x_1-Z_1||^2\le t\}}\,dt\right]\\ &+\mathbb{E}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\int_0^{F_{Z_1}^{-1}(m)} F_{Z_1}(t)-\mathbb{1}_{\{||Z_1-x_1||^2\le t\}}\,dt\right]\\ =:&T_1+T_2.\end{aligned}$$

The Theorem of Tonelli/Fubini [6, Thm. 18] shows that

$$\begin{aligned}T_1 &= \frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)\int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t)-\mathbb{E}\left[\mathbb{1}_{\{||x_1-Z_1||^2\le t\}}\right]\,dt\\ &= \frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)\int_0^{F_{x_1}^{-1}(m)} F_{x_1}(t)-F_{x_1}(t)\,dt=0.\end{aligned}$$

Furthermore, we obtain for $Z_2\sim\mu_{\mathcal{X}}$ independent of Z_1 that

$$\begin{aligned}T_2 =&\mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\int_0^{F_{Z_1}^{-1}(m)} \mathbb{E}_{Z_2}\left[\mathbb{1}_{\{||Z_1-Z_2||^2\le t\}}\right]-\mathbb{1}_{\{||Z_1-x_1||^2\le t\}}\,dt\right]\\ =&\mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\mathbb{E}_{Z_2}\left[\int_0^{F_{Z_1}^{-1}(m)} \mathbb{1}_{\{||Z_1-Z_2||^2\le t\}}-\mathbb{1}_{\{||Z_1-x_1||^2\le t\}}\,dt\right]\right],\end{aligned}$$

where the last step follows by the theorem of Tonelli/Fubini [6, Thm. 18]. Moreover, Lemma F.1 yields that

$$\begin{aligned}T_2 =&\mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\mathbb{E}_{Z_2}\left[||x_1-Z_1||^2\wedge F_{Z_1}^{-1}(m)-||Z_1-Z_2||^2\wedge F_{Z_1}^{-1}(m)\right]\right]\\ =&\mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\left(||x_1-Z_1||^2\wedge F_{Z_1}^{-1}(m)-\mathbb{E}_{Z_2}\left[||Z_1-Z_2||^2\wedge F_{Z_1}^{-1}(m)\right]\right)\right].\end{aligned}$$

Combining all of our results, we finally get that

$$g_{y,h,1}(x_1) =\frac{1}{2h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(x_1)-y}{h}\right)-\frac{1}{2}\Theta_{y,h}+\mathbb{E}_{Z_1}\left[\frac{1}{2mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\Psi(x_1,Z_1)\right],$$

as claimed.

■

C.3.4 Proof of Lemma C.6

Let $x_1 \in \mathcal{X}$ be arbitrary. We observe that for any $z_1, z_2 \in \mathcal{X}$

$$\begin{aligned}|\Psi(x_1,z_1)-\Psi(x_1,z_2)| =& \Big| ||x_1-z_1||^2 \wedge F_{z_1}^{-1}(m) - \mathbb{E}_{Z_2}\left[||Z_2-z_1||^2 \wedge F_{z_1}^{-1}(m)\right] \\ & - ||x_1-z_2||^2 \wedge F_{z_2}^{-1}(m) + \mathbb{E}_{Z_2}\left[||Z_2-z_2||^2 \wedge F_{z_2}^{-1}(m)\right] \Big| \\ \leq & \Big| ||x_1-z_1||^2 \wedge F_{z_1}^{-1}(m) - ||x_1-z_2||^2 \wedge F_{z_2}^{-1}(m) \Big| \\ & + \Big| \mathbb{E}_{Z_2}\left[||Z_2-z_1||^2 \wedge F_{z_1}^{-1}(m)\right] - \mathbb{E}_{Z_2}\left[||Z_2-z_2||^2 \wedge F_{z_2}^{-1}(m)\right] \Big| \\ =: & \Psi_1(z_1,z_2) - \Psi_2(z_1,z_2).\end{aligned}$$

In the following, we consider Ψ_1 and Ψ_2 separately. We have that for $z_1, z_2 \in \mathcal{X}$

$$\begin{aligned}\Psi_1(z_1,z_2) =& \Big| ||x_1-z_1||^2 \wedge F_{z_1}^{-1}(m) - ||x_1-z_2||^2 \wedge F_{z_2}^{-1}(m) \Big| \\ \leq & \Big| ||x_1-z_1||^2 - ||x_1-z_2||^2 \Big| + \Big| F_{z_1}^{-1}(m) - F_{z_2}^{-1}(m) \Big| \\ \leq & D \Big| ||x_1-z_1|| - ||x_1-z_2|| \Big| + 2\sqrt{D}||z_1-z_2||,\end{aligned}$$

where the last inequality follows with $D = \operatorname{diam}(\mathcal{X}) < \infty$ and Lemma 8 in Chazal et al. [14]. In particular, note that in the current setting we have that

$$\sup_{t\in(0,1)} \sup_{x\in\mathcal{X}} F_x^{-1}(t) \leq D < \infty.$$

In consequence, we obtain that for $z_1, z_2 \in \mathcal{X}$

$$\Psi_1(z_1,z_2) \leq D\Big| ||z_1-z_2|| \Big| + 2\sqrt{D}||z_1-z_2|| \leq C||z_1-z_2||,$$

where C denotes a constant that only depends on $\mathcal{X}$.

Next, we consider

$$\begin{aligned}\Psi_2(z_1,z_2) =& \Big| \mathbb{E}_{Z_2}\left[||Z_2-z_1||^2 \wedge F_{z_1}^{-1}(m)\right] - \mathbb{E}_{Z_2}\left[||Z_2-z_2||^2 \wedge F_{z_2}^{-1}(m)\right] \Big| \\ \leq & \mathbb{E}_{Z_2}\left[\Big| ||Z_2-z_1||^2 \wedge F_{z_1}^{-1}(m) - ||Z_2-z_2||^2 \wedge F_{z_2}^{-1}(m) \Big|\right].\end{aligned}$$

Considering our previous calculation, we immediately obtain that

$$\Psi_2(z_1,z_2) \leq \mathbb{E}_{Z_2}\left[C||z_1-z_2||\right] = C||z_1-z_2||,$$

where C denotes the same constant as previously. Combining our results, we find that

$$|\Psi(x_1,z_1)-\Psi(x_1,z_2)| \leq C||z_1-z_2||,$$

where the constant C only depends on $\mathcal{X}$ and not on x_1. This yields the claim.

■

C.3.5 Proof of Lemma C.7

Next, we derive (24) using Lyapunov's Central Limit Theorem for triangular arrays [5, Sec. 27]. To this end, we define $z_{in} := 2g_{y,h,1}(X_i)$, $\bar{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_{in}$ and $\sigma_{in}^2 := \operatorname{Var}(z_{in})$. Clearly, for n fixed the z_{in}'s are independent and identically distributed. In order to check the assumptions of Lyapunov's Central Limit Theorem, it remains to find $r > 2$ such that

$$\rho_{1n} := \mathbb{E}\left[|z_{1n} - \mathbb{E}\left[z_{1n}\right]|^r\right] < \infty \tag{32}$$

and

$$\frac{n\rho_{1n}}{(n\sigma_{1n}^2)^{r/2}} \to 0, \tag{33}$$

as $n \to \infty$.

Calculation of σ_{1n}^2: The next step is to consider σ_{1n}^2. As $\mathbb{E}\left[z_{1n}\right] = 0$ by construction, we find that

$$\sigma_{1n} = \mathbb{E}\left[|z_{1n}|^2\right] = \mathbb{E}\left[|2g_{y,h,1}(X_1)|^2\right] = \mathbb{E}\left[|T_3 + T_4|^2\right],$$

where

$$T_3 := \frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1) - y}{h}\right) - \Theta_{y,h} \tag{34}$$

and

$$T_4 := \mathbb{E}_{Z_1}\left[\frac{1}{mh^2}K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1) - y}{h}\right)\Psi(X_1, Z_1)\right]. \tag{35}$$

Here, $\Psi(x_1, x_2)$ is the function defined in (22). Obviously, we obtain that

$$\sigma_{1n}^2 = \mathbb{E}\left[T_3^2\right] + \mathbb{E}\left[T_4^2\right] + 2\mathbb{E}\left[T_3 T_4\right].$$

In the following, we treat each of these summands separately.

First summand: Considering the first term, we see that

$$\mathbb{E}\left[T_3^2\right] = \mathbb{E}\left[\left|\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1) - y}{h}\right) - \Theta_{y,h}\right|^2\right].$$

which is essentially the variance of the kernel density estimator of the real valued random variable $\mathrm{d}^2_{\mathcal{X},m}(X_1)$. Hence, one can show using standard arguments (see e.g. Silverman [49, Sec. 3.3]) that

$$\mathbb{E}\left[|T_3|^2\right] = \frac{f_{\mathrm{d}^2_{\mathcal{X},m}}(y)}{h}\int |K(u)|^2\,du + o\left(\frac{1}{h}\right)$$

as $h \to 0$.

Second summand: Next, we consider $\mathbb{E}\left[|T_4|^2\right]$. We have that

$$\mathbb{E}\left[|T_4|^2\right] = \mathbb{E}\left[\left|\mathbb{E}_{Z_1}\left[\frac{1}{mh^2} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1) - y}{h}\right) \Psi(X_1, Z_1)\right]\right|^2\right].$$

Recall that $Z_1 \sim \mu_{\mathcal{X}}$ and that $\mu_{\mathcal{X}}$ has, by assumption, a Lipschitz continuous Lebesgue density. Denote this density by $g_{\mu_{\mathcal{X}}}$. Then, it follows that

$$\begin{aligned} &\mathbb{E}_{Z_1}\left[\frac{1}{mh^2} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1) - y}{h}\right) \Psi(X_1, Z_1)\right] \\ =&\frac{1}{mh^2}\int_{\mathcal{X}} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(z) - y}{h}\right) \Psi(X_1, z) g_{\mu_{\mathcal{X}}}(z)\, d\lambda^d(z) \end{aligned} \tag{36}$$

Next, we realize that

$$\sup_{x_1, x_2} |\Psi(x_1, x_2)| \le D$$

and hence there is a constant $0 < C < \infty$ such that

$$\max\left\{\sup_{x\in\mathcal{X}} g_{\mu_{\mathcal{X}}}(x), \sup_{x_1,x_2} |\Psi(x_1, x_2)|\right\} < C. \tag{37}$$

Further, it follows by Lemma C.6 that the function $\Psi^*_{x_1} : \mathcal{X} \to \mathbb{R}$, $z \mapsto \Psi(x_1, z)$ is Lipschitz continuous for all $x_1 \in \mathcal{X}$ with a Lipschitz constant that does not depend on x_1. This in combination with the Lipschitz continuity of $g_{\mu_{\mathcal{X}}}$ and (37) implies that the function

$$\psi_{x_1} : \mathcal{X} \to \mathbb{R}, \quad z \mapsto \Psi(x_1, z) g_{\mu_{\mathcal{X}}}(z)$$

is Lipschitz continuous for all $x_1 \in \mathcal{X}$ with a Lipschitz constant that does not depend on x_1. We have that the function $x \mapsto \mathrm{d}^2_{\mathcal{X},m}(x)$ is coercive, that $\mathrm{d}^2_{\mathcal{X},m}$ is $C^{2,1}$ on an open neighborhood of $\Gamma_y = {\mathrm{d}^2_{\mathcal{X},m}}^{-1}(y)$ and that $\nabla \mathrm{d}^2_{\mathcal{X},m} \neq 0$ on Γ_y by assumption. By Condition 2.2, there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\int_{\Gamma_y} \left|\mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}}\right| d\mathscr{H}^{d-1}(x) \le C_y |v|,$$

where Φ denotes the canonical level set flow of Γ_y and C_y denotes a finite constant that depends on y and $\mathrm{d}^2_{\mathcal{X},m}$. Furthermore, the kernel K is twice continuously differentiable and $\mathrm{supp}(K) = [-1, 1]$. Since K is also even, by assumption, it follows that K' is odd, i.e.

$$\int_{-1}^{1} K'(z)\, dz = 0.$$

Thus, we find by Theorem D.1 that there exists some constants $c_y > 0$ and $h_0 > 0$ (depending on $\mathrm{d}^2_{\mathcal{X},m}$, y and $\mathcal{X}$) such that for any $h < h_0$ we obtain that

$$\left|\mathbb{E}_{Z_1}\left[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\Psi(X_1,Z_1)\right]\right| \le c_y h^2.$$

In consequence, we find that for h small enough

$$\mathbb{E}\left[|T_4|^2\right] \le \mathbb{E}\left[\frac{1}{m^2h^4}\left|\mathbb{E}_{Z_1}\left[K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(Z_1)-y}{h}\right)\Psi(X_1,Z_1)\right]\right|^2\right] \le \frac{c_y}{m^2}.$$

This in particular shows that $\mathbb{E}\left[|T_4|^2\right] = \mathcal{O}(1)$ as $h \to 0$.

Third summand:. By Hölder's inequality, we obtain that

$$\mathbb{E}\left[T_3T_4\right] \le \mathbb{E}\left[|T_3T_4|\right] \le \left(\mathbb{E}\left[|T_3|^2\right]\right)^{1/2}\left(\mathbb{E}\left[|T_4|^2\right]\right)^{1/2}.$$

Plugging in our previous findings, we find that

$$\mathbb{E}\left[T_3T_4\right] = \mathcal{O}\left(\frac{1}{\sqrt{h}}\right)\cdot\mathcal{O}\left(1\right) = \mathcal{O}\left(\frac{1}{\sqrt{h}}\right).$$

as $h \to 0$. In consequence, we find that

$$\sigma^2_{1n} = \frac{f_{\mathrm{d}^2_{\mathcal{X},m}}(x)}{h}\int |K(u)|^2\,du + o\left(\frac{1}{h}\right).$$

This concludes our consideration of σ^2_{1n}.

Calculation of third moments: We choose $r = 3$ and consider $\rho_{1n} = \mathbb{E}\left[|z_{1n} - \mathbb{E}\left[z_{1n}\right]|^r\right]$. By construction $\mathbb{E}\left[z_{1n}\right] = 0$. Thus, we obtain

$$\rho_{1n} = \mathbb{E}\left[|z_{1n}|^3\right] = \mathbb{E}\left[|g_{y,h,1}(X_1)|^3\right] = \mathbb{E}\left[|T_3+T_4|^3\right],$$

where T_3 and T_4 denote the terms defined in (34) and (35), respectively. Furthermore, it follows that

$$\rho_{1n} \le \mathbb{E}\left[(|T_3|+|T_4|)^3\right] \le 8\mathbb{E}\left[|T_3|^3\right] + 8\mathbb{E}\left[|T_4|^3\right].$$

Considering the first summand, this yields that

$$\mathbb{E}\left[|T_3|^3\right] = \mathbb{E}\left[\left|\frac{1}{h}K\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(X_1)-y}{h}\right) - \Theta_{y,h}\right|^3\right],$$

which is the third moment of the kernel density estimator of the real valued random variable $d^2_{\mathcal{X},m}(X_1)$. In particular, one can show using standard arguments that

$$\mathbb{E}\left[|T_3|^3\right] \leq \frac{8 f_{d^2_{\mathcal{X},m}}(x)}{h^2} \int |K(u)|^3 \, du + o\left(\frac{1}{h^2}\right).$$

It remains to consider

$$\mathbb{E}\left[|T_4|^3\right] = \mathbb{E}\left[\left|\frac{1}{2mh^2}\mathbb{E}_{Z_1}\left[K'\left(\frac{d^2_{\mathcal{X},m}(Z_1) - y}{h}\right)\Psi(X_1, Z_1)\right]\right|^3\right].$$

We have already shown that for $h \to 0$

$$\left|\frac{1}{2mh^2}\mathbb{E}_{Z_1}\left[K'\left(\frac{d^2_{\mathcal{X},m}(Z_1) - y}{h}\right)\Psi(X_1, Z_1)\right]\right| = \mathcal{O}(1).$$

Consequently, this implies that

$$\mathbb{E}\left[|T_4|^3\right] = \mathcal{O}(1).$$

Hence, we obtain that

$$\rho_{1n} \leq \frac{8 f_{d^2_{\mathcal{X},m}}(y)}{h^2} \int_{-1}^{1} |K(u)|^3 \, du + o\left(\frac{1}{h^2}\right).$$

Applying Lyapunov's CLT: Now that we have calculated ρ_{1n} and σ^2_{1n}, we can verify the remaining assumption of Lyapunov's Central Limit Theorem for triangular array's [5, Sec. 26]. First of all, we observe that $\rho_{1n} < \infty$, since K, K' and Ψ are continuous and compactly supported. Furthermore, we obtain

$$\begin{aligned}\frac{n\rho_{1n}}{\left(n\sigma^2_{1n}\right)^{3/2}} \leq & \frac{\frac{8n f_{d^2_{\mathcal{X},m}}(x)}{h^2} \int |K(u)|^3 \, du + o\left(\frac{n}{h^2}\right)}{\left(\frac{n f_{d^2_{\mathcal{X},m}}(x)}{2h} \int |K(u)|^2 \, du + o\left(\frac{n}{h}\right)\right)^{3/2}} \\ =& \mathcal{O}\left(\frac{n}{h^2}\right) \cdot \mathcal{O}\left(\frac{n^{-3/2}}{h^{-3/2}}\right) \\ =& \mathcal{O}\left((nh)^{-1/2}\right) \to 0,\end{aligned}$$

if $nh \to \infty$. In consequence, Lyapunov's Central Limit Theorem for triangular arrays is applicable. It yields that

$$\frac{\bar{z}_n - \mathbb{E}\left[\bar{z}_n\right]}{\sqrt{\operatorname{Var}\left(\bar{z}_n\right)}} \xrightarrow{D} N(0,1).$$

This in turn implies that

$$\sqrt{nh}\left(\frac{2}{n}\sum_{i=1}^{n} g_{y,h,1}(X_i)\right) \Rightarrow N\left(0, f_{\mathrm{d}^2_{\mathcal{X},m}}(y)\int_{-1}^{1}|K(u)|^2\,du\right)$$

which gives the claim.

■

D Some Geometric Measure Theory

In the proof of Lemma C.7, we need to bound the term (36):

$$\mathcal{I}(y) := \int_{\mathcal{X}} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(z)-y}{h}\right)\Psi(X_1,z)g_{\mu_{\mathcal{X}}}(z)\,d\lambda^d(z),$$

where $g_{\mu_{\mathcal{X}}}$ denotes the Lebesgue density of $\mu_{\mathcal{X}}$ (which exists by assumption) and $X_1 \sim \mu_{\mathcal{X}}$. Since the kernel K and thus also its derivative K' are supported on $[-1,1]$, we obtain

$$\mathcal{I}(y) = \int_{A_h(y)} K'\left(\frac{\mathrm{d}^2_{\mathcal{X},m}(z)-y}{h}\right)\Psi(X_1,z)g_{\mu_{\mathcal{X}}}(z)\,d\lambda^d(z),$$

where

$$A_h(y) := \left\{z\in\mathcal{X}\,|\, y-h \le \mathrm{d}^2_{\mathcal{X},m}(z) \le y+h\right\} = \left(\mathrm{d}^2_{\mathcal{X},m}\right)^{-1}[y-h,y+h]\cap\mathcal{X}. \tag{38}$$

In the following, we will show how to control such integrals over thickened level sets such as $A_h(y)$ for small h. More precisely, we prove the subsequent theorem that has already been applied to bound the term $\mathcal{I}(y)$ in the proof of Lemma C.7.

Theorem D.1. *Let $\mathcal{X}\subset\mathbb{R}^d$ be a compact set. Let $g\colon \mathcal{X}\to[-\alpha,\alpha]$ be α-Lipschitz continuous and suppose that $k\colon\mathbb{R}\to[-\alpha,\alpha]$ for some $\alpha>0$. Assume that* $\mathrm{supp}(k)=[-1,1]$ *and $\int k(s)\,\mathrm{d}s = 0$. Let $\mathbb{d}\colon\mathbb{R}^d\to\mathbb{R}$ be a coercive function, i.e., $\lim_{||x||\to\infty}\mathbb{d}(x)=\infty$, with level sets $\Gamma_y=\mathbb{d}^{-1}\{y\}$ for $y\in\mathbb{R}$. Call $y\in\mathbb{R}$ a $C^{2,1}$*-regular bounded value *of d with respect to $\mathcal{X}$ if*

C.1 *Γ_y has an open neighborhood on which $\mathbb{d}$ is $C^{2,1}$,*

C.2 *$\nabla\mathbb{d}\neq 0$ on Γ_y.*

C.3 *There exists $h_0^* > 0$ and such that for all $-h_0^* < v < h_0^*$*

$$\int_{\Gamma_y} \left|\mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}}\right| \, d\mathscr{H}^{d-1}(x) \leq C_y|v|,$$

where Φ denotes the canonical level set flow of Γ_y and C_y denotes a constant that only depends on the function $\mathbb{d}$, the variable y and the underlying space $\mathcal{X}$.

If y is a $C^{2,1}$-regular bounded value of d with respect to $\mathcal{X}$, then

$$\left|\int_{\mathcal{X}} k\left(\frac{\mathbb{d}(x)-y}{h}\right) g(x)\,\mathrm{d}\lambda^d(x)\right| \leq c_y h^2 \tag{39}$$

for some $c_y > 0$ and any $0 < h < h_0$, where c_y and $h_0 > 0$ only depend on d, y, α and $\mathcal{X}$ (and not on k and g explicitly).

The proof of Theorem D.1 consists of three steps, each of which is formulated as an independent lemma (see Section D.1).

Step 1: Splitting the integration (Lemma D.2).
We first note that integration over $A_h(y)$ can be split into integrating first over the surface $(\mathrm{d}^2_{\mathcal{X},m})^{-1}(v) \cap \mathcal{X}$ with respect to the $(d-1)$-dimensional Hausdorff measure $\mathscr{H}^{d-1}$ (see Federer [19], Morgan [41] for an introduction) and afterwards over $v \in [y-h, y+h]$.

Step 2: Flow regularity (Lemma D.3).
Let $W \subset \mathbb{R}^d$ be open and let $\varphi : W \to \mathbb{R}^d$ be $C^{2,1}$. We show that the flow Φ corresponding to the initial value problem

$$u' = \varphi(u) \tag{40}$$

is $C^{2,1}$ on its domain (see Hirsch and Smale [29] for information about initial value problems and flows).

Step 3: Local Lipschitz continuity (Lemma D.4).
We prove that the integral of a bounded, α-Lipschitz function $g : \mathcal{X} \subset \mathbb{R}^d \to [-\alpha, \alpha]$ over the level set of a $C^{2,1}$-regular bounded value $\mathbb{d}$ with respect to $\mathcal{X}$, denoted as y, is locally Lipschitz continuous in y. More precisely, we prove that there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$ it holds that

$$\left|\int_{\Gamma_{y+v}\cap\mathcal{X}} g(x)\,d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y\cap\mathcal{X}} g(x)\,d\mathscr{H}^{d-1}(x)\right| \leq C_y|v|,$$

where $C_y > 0$ denotes a constant that only depends on $\mathbb{d}, y, \alpha$ and $\mathcal{X}$.

D.1 Auxiliary Lemmas Representing Step 1 - Step 3

Lemma D.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function. Let $h > 0$, $\mathcal{X} \subset \mathbb{R}^d$ a compact space and $g : \mathbb{R}^d \to \mathbb{R}$ such that the function*

$$x \mapsto \frac{|g(x)|}{||\nabla f(x)||} \mathbb{1}_{\{x \in \mathcal{X} \,:\, |f(x)| \leq h\}} \tag{41}$$

is integrable with respect to λ^d. Then, it follows that

$$\int_{\{x \in \mathcal{X} \,:\, |f(x)| \leq h\}} g(x)\, d\lambda^d(x) = \int_{-h}^{h} \int_{f^{-1}(v) \cap \mathcal{X}} \frac{g(x)}{||\nabla f(x)||}\, d\mathscr{H}^{d-1}(x)\, dv,$$

where $\mathscr{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure.

Proof. First of all, we observe that

$$\int_{\{x \in \mathcal{X} \,:\, |f(x)| \leq h\}} g(x)\, d\lambda^d(x) = \int_{\mathbb{R}^d} \frac{g(x)}{||\nabla f(x)||} \mathbb{1}_{\{x \in \mathcal{X} \,:\, |f(x)| \leq h\}} ||\nabla f(x)||\, d\lambda^d(x).$$

Since the function defined in (41) is integrable, it follows by the *co-area formula* (see Federer [19, Thm. 3.2.12], where the *k-dimensional Jacobian* of f is $||\nabla f||$ in this setting) that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{g(x)}{||\nabla f(x)||} \mathbb{1}_{\{x \in \mathcal{X} \,:\, |f(x)| \leq h\}} ||\nabla f(x)||\, d\lambda^d(x) \\
&= \int_{-\infty}^{\infty} \int_{f^{-1}(v)} \frac{g(x)}{||\nabla f(x)||} \mathbb{1}_{\left\{x \in \mathbb{R}^d \,:\, -h \leq f(x) \leq h\right\}} \mathbb{1}_{\{x \in \mathcal{X}\}}\, d\mathscr{H}^{d-1}(x)\, dv \\
&= \int_{-h}^{h} \int_{f^{-1}(v) \cap \mathcal{X}} \frac{g(x)}{||\nabla f(x)||}\, d\mathscr{H}^{d-1}(x)\, dv.
\end{aligned}$$

This yields the claim. ◻

Lemma D.3. *Let $W \subset \mathbb{R}^d$ be an open set and let $\varphi : W \to \mathbb{R}^d$ be $C^{r,1}$, $1 \leq 0 < \infty$. Consider the initial value problem*

$$\frac{\partial}{\partial t} u(t) = \varphi(u), \quad u(0) = y \tag{42}$$

Let $\Omega \subset \mathbb{R} \times W$ be defined as

$$\Omega = \{(t, y) \in \mathbb{R} \times W \,:\, t \in J(y)\},$$

where $J(y) \subset \mathbb{R}$ denotes the maximal open interval on which the ODE defined in (42) admits a solution. Then, the flow $\Phi : \Omega \to \mathbb{R}^d$ is also in $C^{r,1}$.

Proof. This statement essentially follows by a combination of Theorem 8.3 and Theorem 10.3 in Amann [1] with the idea of proof of Theorem 2 in Chapter 15 of Hirsch and Smale [29]. For the sake of completeness, we give the full argument here.

To prove the claim, we induct on r. The case $r = 0$ follows by combining Theorem 8.3 of Amann [1] with Theorem 10.3 of the same reference. Suppose as induction hypothesis, that $r \geq 1$ and that the flow of every differential equation

$$\frac{\partial}{\partial t}\zeta(t) = \varphi(\zeta)$$

with $\varphi \in C^{r-1,1}$ is $C^{r-1,1}$. Consider the differential equation on $\mathbb{R}^d \times \mathbb{R}^d$ defined by the vector field $\varphi^* : W \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$, $\varphi^*(u, v) = (\varphi(u), D\varphi(u)v)$, i.e.,

$$\frac{\partial}{\partial t}(u, v) = \varphi^*(u, v)$$

or equivalently,

$$u' = \varphi(u), \qquad v' = D\varphi(u)v. \tag{43}$$

Since φ^* is in $C^{r-1,1}$, the flow Φ^* of (43) is $C^{r-1,1}$ by the induction hypothesis. But this flow is just

$$\Phi^*(t, (u, v)) = (\Phi(t, u), D\Phi_t(u)v)$$

since the second equation in (43) is the *variational equation* (see Hirsch and Smale [29, Chap. 15] for a definition) of the first equation. Therefore, $\partial\Phi/\partial u$ is a $C^{r-1,1}$ function of (t, u), since $\partial\Phi/\partial u = D\Phi_t(u)$. Moreover, $\partial\Phi/\partial t$ is in $C^{r-1,1}$ since

$$\frac{\partial}{\partial t}\Phi = \varphi(\Phi(t, u)).$$

It follows that Φ is $C^{r,1}$, since its first partial derivatives are $C^{r-1,1}$. □

Lemma D.4. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact set. Let $g : \mathcal{X} \to [-\alpha, \alpha]$ be an α-Lipschitz function for some $\alpha > 0$. Let $\mathbb{d} \colon \mathbb{R}^d \to \mathbb{R}$ be a coercive function and $y \in \mathbb{R}$ a $C^{2,1}$*-regular bounded value *of $\mathbb{d}$ with respect to $\mathcal{X}$. Let $\Gamma_y = \mathbb{d}^{-1}\{y\}$. Then, it holds that*

$$\left| \int_{\Gamma_{y+v}\cap\mathcal{X}} g(x)\, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y\cap\mathcal{X}} g(x)\, d\mathscr{H}^{d-1}(x) \right| \leq C_y|v|, \tag{44}$$

for some $C_y > 0$ and any $-h_0 < v < h_0$, where C_y and $h_0 > 0$ only depend on $\mathbb{d}$, y, α and $\mathcal{X}$ (and not on g explicitly).

Proof. Before we prove (44), we ensure that the statement is well defined and prove that under the assumptions made

$$\int_{\Gamma_y \cap \mathcal{X}} |g(x)| \, d\mathscr{H}^{d-1}(x) \leq \alpha \int_{\Gamma_y \cap \mathcal{X}} d\mathscr{H}^{d-1}(x) < \infty.$$

To this end, we observe that $\mathscr{H}^{d-1}(\mathbb{d}^{-1}(\{y\} \cap \mathcal{X}) \leq \mathscr{H}^{d-1}(\mathbb{d}^{-1}(\{y\}))$. As $\mathbb{d}$ is coercive it follows that the set $\mathbb{d}^{-1}([0, y])$ is bounded. Hence, the same holds true for $\mathbb{d}^{-1}(\{y\})$. Furthermore, as $\mathbb{d}$ is $C^{2,1}$ in an open environment of the level set Γ_y and $\nabla \mathbb{d} \neq 0$ on Γ_y, it follows that Γ_y is a compact C^1-manifold of dimension $d-1$ [53, Thm. 9], which obviously has finite volume (and hence finite $(d-1)$-dimensional Hausdorff measure [19, 41]).

Now, we focus on proving the statement (44). By assumption, $\mathbb{d}$ is $C^{2,1}$ on an open environment U of Γ_y with $||\nabla \mathbb{d}|| > 0$ on Γ_y. In consequence, there exists $h_0' > 0$ such that $\mathbb{d}^{-1}([y - h_0', y + h_0']) \subset U$ and $||\nabla \mathbb{d}|| > 0$ on $\mathbb{d}^{-1}([y - h_0', y + h_0'])$. This means that the function

$$\varphi(u) : \mathbb{R}^d \to \mathbb{R}^d, \; u \mapsto \frac{\nabla \mathbb{d}(u)}{||\nabla \mathbb{d}(u)||^2}$$

is $C^{1,1}(\mathbb{d}^{-1}((y - h_0', y + h_0')), \mathbb{R}^d)$. By Lemma D.3 (or more generally by Cauchy-Lipschitz's theory [1, 29]) there exists $0 < h_0 \leq h_0'$ such that one can construct a flow $\Phi : [-h_0, h_0] \times W \to \mathbb{R}^d$ with

$$\begin{cases} \frac{\partial}{\partial t} \Phi(t, x) = \frac{\nabla \mathbb{d}(\Phi(t,x))}{||\nabla \mathbb{d}(\Phi(t,x))||^2} \\ \Phi(0, x) = x, \end{cases}$$

where $W \subset \mathbb{R}^d$ is an open set that contains $\mathbb{d}^{-1}([y - h_0, y + h_0])$. Differentiating the function $t \mapsto \mathbb{d}(\Phi(t, x))$ immediately shows that $\mathbb{d}(\Phi(t, x)) = \mathbb{d}(x) + t$. This implies that $\Phi(t, \mathbb{d}^{-1}(\{y\})) = \mathbb{d}^{-1}(\{y + t\})$. In particular, Lemma D.3 yields that Φ is in $C^{1,1}$. Consequently, we find that

$$\begin{aligned}
&\left| \int_{\Gamma_{y+v} \cap \mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) \right| \\
&= \left| \int_{\mathbb{d}^{-1}(\{y\})} g(x) \mathbb{1}_{\{x \in \mathcal{X}\}} d\mathscr{H}^{d-1}(x) - \int_{\Phi(v, \mathbb{d}^{-1}(\{y\}))} g(x) \mathbb{1}_{\{x \in \mathcal{X}\}} \, d\mathscr{H}^{d-1}(x) \right| \\
&\leq \int_{\mathbb{d}^{-1}(\{y\})} \left| g(\Phi(0, x)) J_{\Phi(0, \cdot)}(x) \mathbb{1}_{\{\Phi(0,x) \in \mathcal{X}\}} - g(\Phi(v, x)) J_{\Phi(v, \cdot)}(x) \mathbb{1}_{\{\Phi(v,x) \in \mathcal{X}\}} \right| \, d\mathscr{H}^{d-1}(x),
\end{aligned}$$

where $J_{\Phi(v,\cdot)}$ denotes the *Jacobian determinant* of $\Phi(v, \cdot)$. The last line follows by a change of variables (see e.g. Merigot and Thibert [40, Thm. 56]) and the fact that $\Phi(0, \cdot)$ is the identity. By Kirszbraun's Theorem [19, Thm. 2.10.43] we can extend $g : \mathcal{X} \to [-\alpha, \alpha]$

to a Lipschitz continuous function $\tilde{g} : \mathbb{R}^d \to \mathbb{R}$, that has the same Lipschitz constant α. Obviously, it holds that

$$\begin{aligned}&\int_{\mathbb{d}^{-1}(\{y\})} \left| g(\Phi(0,x)) J_{\Phi(0,\cdot)}(x) \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - g(\Phi(v,x)) J_{\Phi(v,\cdot)}(x) \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| \, d\mathscr{H}^{d-1}(x) \\ =& \int_{\mathbb{d}^{-1}(\{y\})} \left| \tilde{g}(\Phi(0,x)) J_{\Phi(0,\cdot)}(x) \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \tilde{g}(\Phi(v,x)) J_{\Phi(v,\cdot)}(x) \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| \, d\mathscr{H}^{d-1}(x).\end{aligned}$$

Therefore, we find that

$$\begin{aligned}&\left| \int_{\Gamma_{y+v}\cap\mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y\cap\mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) \right| \\ \le& \int_{\mathbb{d}^{-1}(\{y\})} \left| \tilde{g}(\Phi(0,x)) J_{\Phi(0,\cdot)}(x) - \tilde{g}(\Phi(v,x)) J_{\Phi(v,\cdot)}(x) \right| \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} \, d\mathscr{H}^{d-1}(x) \\ &+ \int_{\mathbb{d}^{-1}(\{y\})} \left| \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| \left| \tilde{g}(\Phi(v,x)) J_{\Phi(v,\cdot)}(x) \right| \, d\mathscr{H}^{d-1}(x).\end{aligned}$$

Since Φ is in $C^{1,1}([-h_0,h_0]\times W)$, it follows that $(v,x)\mapsto \tilde{g}(\Phi(v,x))$ and $(v,x)\mapsto J_{\Phi(v,\cdot)}(x)$ are Lipschitz continuous functions. We observe that for $(v,x)\in[-h_0,h_0]\times\mathcal{X}$

$$\begin{aligned}|\tilde{g}(\Phi(v,x))| \le |\tilde{g}(\Phi(0,x))| + |\tilde{g}(\Phi(v,x)) - \tilde{g}(\Phi(0,x))| &\le \alpha + \alpha||\Phi(v,x) - \Phi(0,x)|| \\ &\le \alpha + \alpha L_\Phi h_0,\end{aligned}$$

where L_Φ denotes the Lipschitz constant of Φ. This implies immediately that the function $(v,x) \mapsto \tilde{g}(\Phi(v,x)) J_{\Phi(v,\cdot)}(x)$ is Lipschitz continuous on $[-h_0,h_0]\times\mathcal{X}$ with a Lipschitz constant that only depends on $\mathbb{d}, y, \alpha$ and $\mathcal{X}$. Further, we realize that

$$\left| \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| > 0 \tag{45}$$

implies that either $x\in\mathcal{X}$ or $\Phi(v,x)\in\mathcal{X}$ (but not both). Given (45), our previous calculations show that

$$\left| \tilde{g}(\Phi(v,x)) J_{\Phi(v,\cdot)}(x) \right| \le C_y,$$

where C_y denotes a finite constant that depends only on $\mathbb{d}, y, \alpha$ as well as $\mathcal{X}$. For the remainder of this proof, this constant may vary from line to line. We obtain that

$$\begin{aligned}&\left| \int_{\Gamma_{y+v}\cap\mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y\cap\mathcal{X}} g(x) \, d\mathscr{H}^{d-1}(x) \right| \\ \le& \int_{\mathbb{d}^{-1}(\{y\})} C_y \, |v| \, \mathbb{1}_{\{x\in\mathcal{X}\}} \, d\mathscr{H}^{d-1}(x) + C_y \int_{\mathbb{d}^{-1}(\{y\})} \left| \mathbb{1}_{\{\Phi(0,x)\in\mathcal{X}\}} - \mathbb{1}_{\{\Phi(v,x)\in\mathcal{X}\}} \right| \, d\mathscr{H}^{d-1}(x).\end{aligned}$$

Since y is a $C^{2,1}$-regular value of $\mathbb{d}$ with respect to $\mathcal{X}$, we find that (by potentially adjusting h_0) there exists $h_0 > 0$ such that for all $-h_0 < v < h_0$

$$\left| \int_{\Gamma_{y+v}\cap\mathcal{X}} g(x)\, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y\cap\mathcal{X}} g(x)\, d\mathscr{H}^{d-1}(x) \right| \le C_y|v|\mathscr{H}^{d-1}(\Gamma_y) + C_y|v| \le C_y|v|.$$

This gives the claim. □

D.2 Proof of Theorem D.1

By assumption, we have that $\nabla\mathbb{d} \neq 0$ on the level set Γ_y. Furthermore, we have assumed that the function $\mathbb{d}$ is $C^{2,1}$ on an open neighborhood of Γ_y. Thus, there exists $h_0 > 0$ such that $||\nabla\mathbb{d}|| > 0$ on

$$\mathbb{d}^{-1}[y - h_0, y + h_0] = \{x \in \mathbb{R}^d \,:\, y - h_0 \le \mathbb{d}(x) \le y + h_0\}. \tag{46}$$

Throughout the following let $0 < h < h_0$. We get that

$$\begin{aligned}\left| \int_{\mathcal{X}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(x)\, \mathrm{d}\lambda^d(x) \right| &= \left| \int_{\{x\in\mathcal{X}\,:\, y-h\le\mathbb{d}(x)\le y+h\}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(x)\, \mathrm{d}\lambda^d(x) \right| \\ &= \left| \int_{\{x\in\mathcal{X}\,:\, |\mathbb{d}(x)-y|\le h\}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(x)\, \mathrm{d}\lambda^d(x) \right|.\end{aligned}$$

Since $||\nabla\mathbb{d}(x)|| > 0$ for $x \in \mathbb{d}^{-1}[y - h_0, y + h_0]$ and $|g(x)| \le \alpha$ for all x, we obtain that

$$\sup_{x\in\mathbb{R}^d} \left| \frac{g(x)}{||\nabla\mathbb{d}(x)||} \mathbb{1}_{\{x\in\mathcal{X}\,:\,|\mathbb{d}(x)-y|\le h\}} \right| < C_y, \tag{47}$$

where C_y denotes a constant that only depends on $\mathbb{d}, y, \alpha$ and $\mathcal{X}$ (in particular it can be chosen independently from h). In the following, C_y may vary from line to line. Clearly, (47) implies that the function

$$x \mapsto \frac{|g(x)|}{||\nabla\mathbb{d}(x)||} \mathbb{1}_{\{x\in\mathcal{X}\,:\,|\mathbb{d}(x)-y|\le h\}}$$

is λ^d-integrable for any $0 \le h \le h_0$. Therefore, it follows by Lemma D.2 in combination with (47) that

$$\begin{aligned}&\left| \int_{\mathcal{X}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(x)\, \mathrm{d}\lambda^d(x) \right| \\ &= \left| \int_{-h}^{h} \int_{\{x\in\mathcal{X}\,:\,\mathbb{d}(x)-y=v\}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) \frac{g(x)}{||\nabla\mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x)\, dv \right| \\ &= \left| \int_{-h}^{h} k\left(\frac{v}{h}\right) \int_{\{x\in\mathcal{X}\,:\,\mathbb{d}(x)-y=v\}} \frac{g(x)}{||\nabla\mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x)\, dv \right|.\end{aligned}$$

We note that

$$\{x \in \mathcal{X} : \mathbb{d}(x) - y = v\} = \mathbb{d}^{-1}(y+v) \cap \mathcal{X} = \Gamma_{y+v} \cap \mathcal{X}.$$

This yields that

$$\begin{aligned}
&\left| \int_{\mathcal{X}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(x)\, \mathrm{d}\lambda^d(x) \right| \\
&\leq \left| \int_{-h}^{h} k\left(\frac{v}{h}\right) \int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x)\, dv \right| \\
&+ \left| \int_{-h}^{h} k\left(\frac{v}{h}\right) \left(\int_{\Gamma_{y+v} \cap \mathcal{X}} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x) \right) dv \right| =: T_5 + T_6.
\end{aligned}$$

Next, we consider both summands separately. First of all, we observe that the integral

$$\int_{\Gamma_y \cap \mathcal{X}} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x)$$

does not depend on v. Consequently, we obtain that

$$T_5 \overset{(i)}{\leq} C_y \left| \int_{-h}^{h} k\left(\frac{v}{h}\right) dv \right| \left| \int_{\Gamma_y \cap \mathcal{X}} d\mathscr{H}^{d-1}(x) \right| \overset{(ii)}{\leq} C_y \left| \int_{-h}^{h} k\left(\frac{v}{h}\right) dv \right|.$$

Here, (i) follows by (47) and (ii) follows since $\mathscr{H}^{d-1}(\Gamma_y \cap \mathcal{X}) \leq C < \infty$ for some constant C, as already argued in the proof of of Lemma D.4. Setting $u = v/h$ and using $\int k(u)\, du = 0$ gives that

$$T_5 \leq C_y h \left| \int_{-1}^{1} k(u)\, du \right| = 0.$$

Hence, it only remains to consider the second summand T_6. Let $\mathcal{X}^* = \mathcal{X} \cap \mathbb{d}^{-1}([y - h_0, y + h_0])$. Since $h \leq h_0$, we obtain that

$$T_6 \leq \int_{-h}^{h} \left| k\left(\frac{v}{h}\right) \right| \left| \int_{\Gamma_{y+v} \cap \mathcal{X}^*} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x) - \int_{\Gamma_y \cap \mathcal{X}^*} \frac{g(x)}{||\nabla \mathbb{d}(x)||}\, d\mathscr{H}^{d-1}(x) \right| dv$$

We realize that the function

$$g^* : \mathcal{X}^* \to \mathbb{R}, \quad x \mapsto \frac{g(x)}{||\nabla \mathbb{d}(x)||}$$

is Lipschitz continuous, as $||\nabla \mathbb{d}(x)|| > 0$ for $x \in \mathbb{d}^{-1}([y - h_0, y + h_0])$, the function $||\nabla \mathbb{d}(x)||$ is in $C^{1,1}(\mathbb{d}(^{-1}(y - h_0, y + h_0))$ and g is Lipschitz continuous and bounded by assumption. As y is a $C^{2,1}$-regular bounded value of $\mathbb{d}$ with respect to $\mathcal{X}$, it is straight forward to verify that it is also one with respect to $\mathcal{X}^*$. Thus, the requirements of Lemma D.4 are met. By potentially decreasing h_0, we find for all h small enough that

$$T_6 \leq C_y \int_{-h}^{h} \left| k\left(\frac{v}{h}\right) \right| v \, dv.$$

Setting $u = v/h$ gives that

$$T_6 \leq C_y h^2 \int_{-1}^{1} |k(u)| \, u \, du \leq c_y h^2,$$

where $c_y > 0$ depends only on $\mathbb{d}, y, \alpha$ and $\mathcal{X}$.

All in all, this gives

$$\left| \int_{\mathcal{X}} k\left(\frac{\mathbb{d}(x) - y}{h}\right) g(z) \, \mathrm{d}\lambda^d(z) \right| \leq T_5 + T_6 \leq c_y h^2,$$

which yields the claim. ■

E The Distance-to-Measure-Function

In this section, we derive further properties of the DTM-function.

First of all, we ensure that

$$\sup_{x \in \mathcal{X}} \left| \delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x) \right| = \mathcal{O}_P(1/\sqrt{n})$$

This is a minor extension of Theorem 9 in Chazal et al. [14], which considers only $0 < m < 1$. For this purpose, we need to introduce some notation. For a compact set $A \subset \mathbb{R}^d$ we define the radius of the *smallest enclosing ball* of centered at zero as

$$r(A) = \inf\{r > 0 : A \subset \bar{B}(0, r)\},$$

where $\bar{B}(0, r)$ denotes a closed ball with radius r centered at the origin.

Lemma E.1. *Let P be a measure with compact support and let $\mathcal{X}$ be a compact domain on $\mathbb{R}^d$. Further, suppose that for any $x \in \mathcal{X}$ the pushforward measure of P by $||x - \cdot||^2$, whose distribution function is denoted by F_x, is supported on a finite closed interval $[0, D_x]$ with*

$$\sup_{x \in \mathcal{X}} D_x \leq D < \infty.$$

Suppose that there is a constant $C_{\mathcal{X}}$ such that $r(\mathcal{X}) < C_{\mathcal{X}}$. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ and denote the corresponding empirical measure by P_n. Then, it follows that

$$\sup_{x\in\mathcal{X}} \left|\delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x)\right| = \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Let $\widehat{F}_{x,n}$ be defined as in (2). Recalling (31), we find

$$\sqrt{n}\left(\delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x)\right) = \sqrt{n}\left(\int_0^{D_x} F_x(t) - \widehat{F}_{x,n}(t)\, dt\right)$$
$$= \sqrt{n}\left(\int_0^{D_x}\int \mathbb{1}_{\{||x-z||^2\le t\}}\, dP(z) \;-\int \mathbb{1}_{\{||x-z||^2\le t\}}\, dP_n(z)\, dt\right).$$

Since P is compactly supported and $\sup_{x\in\mathcal{X}}|D_x| < \infty$ by assumption, the Theorem of Tonelli/ Fubini [6, Thm. 18] yields

$$\begin{aligned}
&\sqrt{n}\left(\delta^2_{\mathcal{X},1}(x) - \mathrm{d}^2_{\mathcal{X},1}(x)\right)\\
=&\sqrt{n}\left(\int\int_0^{D_x} \mathbb{1}_{\{||x-z||^2\le t\}}\, dt\, dP(z) \;-\int\int_0^{D_x} \mathbb{1}_{\{||x-z||^2\le t\}}\, dt\, dP_n(z)\right)\\
=&-\nu_n(g_x),
\end{aligned}$$

where $\nu_n = \sqrt{n}\,(P_n - P)$ denotes the empirical process and

$$g_x(z) = \int_0^{D_x} \mathbb{1}_{\{||x-z||^2\le t\}}\, dt = D_x - ||x-z||^2.$$

Hence, the claim follows once we have shown that $\mathcal{G} = \{g_x : x \in \mathcal{X}\}$ is a Donsker class. To this end, we observe that by Chazal et al. [14, Lemma 8] it holds that for $x, x' \in \mathcal{X}$

$$\begin{aligned}
|D_x - D_{x'}| \le \sup_{t\in(0,1)} |F_x^{-1}(t) - F_{x'}^{-1}(t)| &\le 2 \sup_{t\in(0,1)}\left[\sup_{x\in\mathcal{X}} F_x^{-1}(t)\right]||x-x'||\\
&\le 2D||x-x'||.
\end{aligned}$$

Now, we have for any $(x, x') \in \mathcal{X}^2$ and any $z \in \mathrm{supp}(P)$ that

$$\begin{aligned}
|g_x(z) - g_{x'}(z)| &\le |D_x - D_{x'}| + ||x-x'||\left(||x|| + ||x'|| + 2||z||\right)\\
&\le 2\left(C_{\mathcal{X}} + D + ||z||\right)||x-x'||.
\end{aligned}$$

As P is compactly supported, it follows that $\mathcal{G}$ is a Donsker class (see Example 19.7 in Van der Vaart [52]). As already argued this yields the claim. □

F Miscellaneous

Lemma F.1. *Let x, y and κ denote non-negative real numbers. Then, it holds that*

$$\int_0^\kappa \mathbb{1}_{\{x\leq t\}} - \mathbb{1}_{\{y\leq t\}}\,dt = y \wedge \kappa - x \wedge \kappa.$$

Proof. In order to show the claim, we have to distinguish several cases:

1. $x \leq y \leq \kappa$: In this case we have that

$$\int_0^\kappa \mathbb{1}_{\{x\leq t\}} - \mathbb{1}_{\{y\leq t\}}\,dt = y - x = y \wedge \kappa - x \wedge \kappa.$$

2. $x \leq \kappa \leq y$: Here, obtain that

$$\int_0^\kappa \mathbb{1}_{\{x\leq t\}} - \mathbb{1}_{\{y\leq t\}}\,dt = \kappa - x = y \wedge \kappa - x \wedge \kappa.$$

3. $y \leq x \leq \kappa$: It follows that

$$\int_0^\kappa \mathbb{1}_{\{x\leq t\}} - \mathbb{1}_{\{y\leq t\}}\,dt = -(x - y) = y \wedge \kappa - x \wedge \kappa.$$

4. $y \leq \kappa \leq x$: We get

$$\int_0^\kappa \mathbb{1}_{\{x\leq t\}} - \mathbb{1}_{\{y\leq t\}}\,dt = -(\kappa - y) = y \wedge \kappa - x \wedge \kappa.$$

5. $\kappa \leq x$ and $\kappa \leq y$: In this case, the claim is trivial. □

References

[1] H. Amann. *Ordinary differential equations: An introduction to nonlinear analysis*, volume 13. Walter de Gruyter, 2011.

[2] S. Belongie, G. Mori, and J. Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006.

[3] J. R. Berrendero, A. Cuevas, and B. Pateiro-López. Shape classification based on interpoint distance distributions. *Journal of Multivariate Analysis*, 146:237–247, 2016.

[4] B. Bhattacharjee. nth-nearest-neighbor distribution functions of an interacting fluid from the pair correlation function: A hierarchical approach. *Physical Review E*, 67:041208, 2003.

[5] P. Billingsley. *Probability and measure.* John Wiley & Sons, 1979.

[6] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, 2013.

[7] C. Brécheteau. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature. *Electronic Journal of Statistics*, 13(1):795–849, 2019.

[8] D. Brinkman and P. J. Olver. Invariant histograms. *The American Mathematical Monthly*, 119(1): 4–24, 2012.

[9] M. Buchet, F. Chazal, T. K. Dey, F. Fan, S. Y. Oudot, and Y. Wang. Topological analysis of scalar fields with outliers, 2014. https://arxiv.org/abs/1412.1680.

[10] J. V. Castellana and M. R. Leadbetter. On smoothed probability density estimation for stationary processes. *Stochastic Processes and their Applications*, 21(2):179–193, 1986.

[11] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733–751, 2011.

[12] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):1–38, 2013.

[13] F. Chazal, P. Massart, B. Michel, et al. Rates of convergence for robust geometric inference. *Electronic journal of statistics*, 10(2):2243–2286, 2016.

[14] F. Chazal, B. Fasy, F. Lecci, B. Michel, A. Rinaldo, A. Rinaldo, and L. Wasserman. Robust topological inference: Distance to a measure and kernel distance. *The Journal of Machine Learning Research*, 18 (1):5845–5884, 2017.

[15] S. Chen and M. Pokojovy. Modern and classical k-sample omnibus tests. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(1):e1418, 2018.

[16] S. Csorgo and J. Mielniczuk. Density estimation under long-range dependence. *The Annals of Statistics*, 23(3):990–999, 1995.

[17] A. Cuevas. Set estimation: Another bridge between statistics and geometry. *Boletin de Estadistica e Investigacion Operativa*, 25(2):71–85, 2009.

[18] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.

[19] H. Federer. *Geometric measure theory.* Springer, 1969.

[20] E. W. Frees. Estimating densities of functions of observations. *Journal of the American Statistical Association*, 89(426):517–525, 1994.

[21] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In *Symposium on geometry processing*, volume 2, page 5. Vienna, Austria, 2005.

[22] M. Gellert, M. F. Hossain, F. J. F. Berens, L. W. Bruhn, C. Urbainsky, V. Liebscher, and C. H. Lillig. Substrate specificity of thioredoxins and glutaredoxins–towards a functional classification. *Heliyon*, 5 (12):e02943, 2019.

[23] B. S. Graham, F. Niu, and J. L. Powell. Kernel density estimation for undirected dyadic data, 2019. https://arxiv.org/abs/1907.13630.

[24] M. Hallin, Z. Lu, and L. Tran. Kernel density estimation for spatial processes: the l1 theory. *Journal of Multivariate Analysis*, 88(1):61–75, 2004.

[25] A. B. Hamza and H. Krim. Geodesic object representation and recognition. In *International conference on discrete geometry for computer imagery*, pages 378–387. Springer, 2003.
[26] X. Hao, J. Parmar, and B. Lelandais et al. Super-resolution visualization and modeling of human chromosomal regions reveals cohesin-dependent loop structures. *Genome Biology*, 22, 2021.
[27] Y. S. Hayakawa and T. Oguchi. Applications of terrestrial laser scanning in geomorphology. *Journal of Geography (Chigaku Zasshi)*, 125(3):299–324, 2016.
[28] F. Heinemann and N. Bonneel. *WSGeometry: Compute Wasserstein barycenters, geodesics, PCA and distances*, 2021. URL https://CRAN.R-project.org/package=WSGeometry. R package version 1.0.
[29] M. Hirsch and S. Smale. *Differential equations, dynamical systems, and linear algebra (Pure and Applied Mathematics, Vol. 60)*. Springer, 1974.
[30] S.-S. Hsiao, K.-T. Chen, and A. Y. Ite. Mean field theory of weakly-interacting rydberg polaritons in the eit system based on the nearest-neighbor distribution. *Optics Express*, 28(19):28414–28429, 2020.
[31] M. Lelek, M. Gyparaki, and G. Beliu et al. Single-molecule localization microscopy. *Nature Reviews Methods Primers*, 1, 2021.
[32] N. I. Libeskind et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473 (1):1195–1217, 08 2017.
[33] E. Lieberman-Aiden et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.
[34] E. Liebscher. Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications*, 65(1):69–80, 1996.
[35] Z. Lu. Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics*, 53(3):447–468, 2001.
[36] R Core Team. R*: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/.
[37] F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
[38] F. Mémoli and T. Needham. Distance distributions and inverse problems for metric measure spaces, 2018. https://arxiv.org/abs/1810.09646.
[39] H. Meng, Y. Gao, X. Yang, K. Wang, and J. Tian. K-nearest neighbor based locally connected network for fast morphological reconstruction in fluorescence molecular tomography. *IEEE transactions on medical imaging*, 39(10):3019–3028, 2020.
[40] Q. Merigot and B. Thibert. Optimal transport: discretization and algorithms. In *Handbook of Numerical Analysis*, volume 22, pages 133–212. Elsevier, 2021.
[41] F. Morgan. *Geometric measure theory: a beginner's guide*. Academic press, 2016.
[42] P. Nicovich, D. Owen, and K. Gaus. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nature Protocols*, 12:453–460, 2017.
[43] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, and L. A. Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706, 2018.
[44] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21(4):807–832, 2002.
[45] P. M. Robinson. Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3): 185–207, 1983.
[46] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 55. Birkäuser, NY, 2015.
[47] Y. Shi, P. M. Thompson, G. I. de Zubicaray, S. E. Rose, Z. Tu, I. Dinov, and A. W. Toga. Direct mapping of hippocampal surfaces with intrinsic shape context. *NeuroImage*, 37(3):792–807, 2007.
[48] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. SIAM, 2009.
[49] B. W. Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
[50] S. Torquato, B. Lu, and J. Rubinstein. Nearest-neighbor distribution functions in many-body systems. *Physical Review A*, 41(4):2059, 1990.

[51] A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer, 2008.

[52] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[53] A. Villanacci, L. Carosi, P. Benevieri, and A. Battinelli. *Differential topology and general equilibrium with complete and incomplete markets.* Springer Science & Business Media, 2002.

[54] G. Vosselman, B. G. Gorte, G. Sithole, and T. Rabbani. Recognising structure in laser scanner point clouds. *International archives of photogrammetry, remote sensing and spatial information sciences*, 46 (8):33–38, 2004.

[55] C. A. Weitkamp, K. Proksch, C. Tameling, and A. Munk. Gromov-wasserstein distance based object matching: Asymptotic inference, 2020. https://arxiv.org/abs/2006.12287.

[56] W. B. Wu and J. Mielniczuk. Kernel density estimation for linear processes. *The Annals of Statistics*, 30(5):1441 – 1459, 2002.

[57] G. Zou and H.-I. Wu. Nearest-neighbor distribution of interacting biological entities. *Journal of Theoretical Biology*, 172(4):347–353, 1995.