Sketching and Clustering Metric Measure Spaces

Facundo Mémoli* Anastasios Sidiropoulos† Kritika Singhal‡

Abstract

Two important optimization problems in the analysis of geometric data sets are *clustering* and *sketching*. Here, clustering refers to the problem of partitioning some input metric measure space into k clusters, minimizing some objective function f. Sketching, on the other hand, is the problem of approximating some metric measure space by a smaller one supported on a set of k points. Specifically, we define the k-sketch of some metric measure space M to be the nearest neighbor of M in the set of k-point metric measure spaces, under some distance function ρ on the set of metric measure spaces.

In this paper we demonstrate a duality between general classes of clustering and sketching problems. We present a general method for efficiently transforming a solution for a clustering problem to a solution for a sketching problem, and vice versa, with approximately equal cost.

More specifically, we obtain the following results. We define the sketching/clustering *gap* to be the supremum over all metric measure spaces of the ratio of the sketching and clustering objectives.

1. For metric spaces, we consider the case where f is the maximum cluster diameter, and ρ is the Gromov-Hausdorff distance. We show that the gap is constant for any compact metric space.
2. We extend the above results to obtain constant gaps for the case of metric measure spaces, where ρ is the p-Gromov-Wasserstein distance and the clustering objective involves minimizing various notions of the ℓ_p-*diameters* of the clusters.
3. We consider two competing notions of sketching for metric measure spaces, with one of them being more demanding than the other. These notions arise from two different definitions of p-Gromov-Wasserstein distance that have appeared in the literature. We then prove that whereas the gap between these can be arbitrarily large, in the case of doubling metric spaces the resulting sketching objectives are polynomially related.

*Department of Mathematics, The Ohio State University, memoli@math.osu,edu.
†Department of Computer Science, University of Illinois at Chicago, sidiropo@uic.edu.
‡Department of Mathematics, The Ohio State University, singhal.53@osu.edu.

1 Introduction

Clustering and sketching are two fundamental geometric primitives with numerous applications in science and engineering [29, 22]. In this paper we show a *duality* between clustering and sketching in metric spaces, and metric measure spaces. Specifically, we consider the problem of clustering a space into k blocks, and the problem of summarizing some space as a k-point space, for some small $k \in \mathbb{N}$. We show that for various natural clustering and sketching objectives, their optimal values are either within a constant factor of each other, or they are closely related. This duality is of interest since, typically, one can obtain approximation algorithms for computing the sketch of some space, using known approximation algorithms for the dual clustering problem.

Clustering metric spaces and metric measure spaces.

Several objective functions have been considered for clustering metric spaces, such as k-center [19, 27], k-median [29, 9, 11, 34, 33, 30, 44], k-means [18, 35, 14, 7, 45, 24, 25, 23, 38, 37, 1, 13, 15, 33, 8], and so on.

Many of these objectives can be extended to the case of metric measure spaces. For finite spaces, this is equivalent to the case where the input pointset is endowed with a non-negative weight function.

Sketching. Informally, *sketching* in computational geometry is the broad problem of obtaining some efficiently computable "summary" of some dataset x, which is viewed as a point in some metric space $(\mathcal{M}, \rho)$. For example, $\mathcal{M}$ could be some ℓ_p space (see [28, 32]), or some general normed space (see [5]), which is an important setting in streaming algorithms. Another example is when $\mathcal{M}$ is a set of strings, and ρ the edit distance (see [4]), which is a natural setting in string analysis and computational biology. Similarly, $\mathcal{M}$ could be the collection of all measures in the 2-dimensional grid, and ρ the earth-mover distance (see [3, 6]), which is a setting arising in image analysis and pattern recognition.

Projective sketching. We propose a specific type of sketching objective, which is natural for the case of summarizing metric spaces. We consider the case where $\mathcal{M}$ is a collection of metric spaces, or metric measure spaces, endowed with some distance ρ. We also let $\mathcal{M}_k$ be the collection of all k-point metric spaces, or metric measure spaces. Then, for some $M \in \mathcal{M}$, the *k-point sketch* of M is defined to be a nearest-neighbor of M in $\mathcal{M}_k$; that is, we define

$$s(M) = \inf_{M' \in \mathcal{M}_k} \rho(M, M'),$$

breaking ties arbitrarily.

We refer to the problem of computing this mapping as *projective sketching* (i.e. we project $\mathcal{M}$ onto $\mathcal{M}_k$).

1.1 Duality between clustering and sketching

We now give a high-level description of our results (see later in this Section for a formal description).

The case of metric spaces. We demonstrate a duality between clustering and projective sketching of metric spaces. Specifically, we consider the clustering objective of minimizing the maximum cluster diameter, which is within a factor of 2 from the k-center objective (see [19]). For the case where $\mathcal{M}$ is the collection of all compact metric spaces, and ρ is the Gromov-Hausdorff distance, we show that the ratio between the clustering and projective sketching objectives is a fixed constant.

The case of metric measure spaces. We further show a duality between clustering and projective sketching of metric measure spaces. We consider the clustering objective of minimizing the weighted ℓ_p-norm of the intra-cluster distances. This is within a factor of 2 of the k-median objective when $p = 1$, and the k-means objective when $p = 2$. When ρ is the p-Gromov-Wasserstein distance, we show that the ratio between the clustering and sketching objectives is bounded by some constant, for all $p \in [1, \infty]$. We also consider alternative definition of p-Gromov-Wasserstein distance due to Sturm [43]. We show that the two sketching objectives are polynomially related for metric measure spaces of bounded diameter and bounded doubling dimension.

Limits of duality. We also show that for some natural clustering objective, no analogous duality result is possible, even for the case of metric spaces. Specifically, when the objective is to maximize the minimum inter-cluster distance, we show that for any choice of ρ, the ratio between the clustering and sketching objectives is unbounded.

1.2 Formal statement of our results

We now formally state our main results.

1.2.1 The case of metric spaces

Let $\mathcal{M}$ denote the collection of all compact metric spaces. For $k \in \mathbf{N}$, let $\mathcal{M}_k \subset \mathcal{M}$ denote the collection of all finite metric spaces of cardinality at most k. For $(X, d_X) \in \mathcal{M}$ and $k \in \mathbf{N}$, let $\mathrm{Part}_k(X)$ denote the set of all partitions of X into k sets. A *clustering cost function* is a function $\Phi : \mathcal{M} \times \mathrm{Part}_k \to \mathbf{R}_+$ that takes as input some $(X, d_X) \in \mathcal{M}$ and some $P \in \mathrm{Part}_k(X)$, and returns the cost of clustering the space (X, d_X) according to P. Given such a Φ and $(X, d_X) \in \mathcal{M}$, we define

$$\mathrm{Shatter}_k^{\Phi}(X) = \inf_{P \in \mathrm{Part}_k(X)} \Phi(X, P).$$

A *sketching cost function* is a function $\Psi : \mathcal{M} \times \mathcal{M}_k \to \mathbf{R}_+$ that takes as input some $(X, d_X) \in \mathcal{M}$ and some $(M_k, d_k) \in \mathcal{M}_k$, and returns the cost of sketching (X, d_X) by (M_k, d_k). Given such a Ψ and $(X, d_X) \in \mathcal{M}$, we define

$$\mathrm{Sketch}_k^{\Psi}(X) = \inf_{M_k \in \mathcal{M}_k} \Psi(X, M_k).$$

A sketching cost function Ψ and a clustering cost function Φ are called *dual* if there exist constants $C_2 \geq C_1 > 0$ such that

$$C_1 \cdot \mathrm{Shatter}_k^{\Phi}(X) \leq \mathrm{Sketch}_k^{\Psi}(X) \leq C_2 \cdot \mathrm{Shatter}_k^{\Phi}(X) \tag{1}$$

for every $X \in \mathcal{M}$ and $k \in \mathbf{N}$. The duality is *strict* when $C_1 = C_2$. Interchangeably we will say that $\mathrm{Shatter}_k^{\Phi}$ and $\mathrm{Shatter}_k^{\Phi}$ are dual (resp. strictly dual) when the conditions above hold.

Given $X \in \mathcal{M}$, $k \in \mathbf{N}$ and $M_k \in \mathcal{M}_k$, we consider the choice of sketching cost function $\Psi(X, M_k) = d_{GH}(X, M_k)$. Here d_{GH} is the Gromov-Hausdorff distance between metric spaces, see section 2 for details. The resulting sketching objective is defined as

$$\mathrm{Sketch}_k(X) := \mathrm{Sketch}_k^{\Psi}(X) = \inf_{(M_k, d_k) \in \mathcal{M}_k} d_{GH}(X, M_k).$$

For any $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, we consider the choice of clustering cost function $\Phi(X,P) = \max_i \mathrm{diam}(B_i)$, where for any set $S \subseteq X$, its diameter is $\mathrm{diam}(S) = \sup_{x,x' \in S} d_X(x,x')$. The resulting clustering objective is defined as

$$\mathrm{Shatter}_k(X) := \mathrm{Shatter}_k^{\Phi}(X) = \inf_{P \in \mathrm{Part}_k(X)} \Phi(X,P).$$

Note that the above is within a factor of 2 from the k-center objective. Our main result in this setting can now be stated as follows.

Theorem 1.1 (Strict duality for metric spaces)**.** *For every $(X,d_X) \in \mathcal{M}$ and every $k \in \mathbf{N}$, we have*

$$\mathrm{Sketch}_k(X) = \frac{1}{2} \cdot \mathrm{Shatter}_k(X).$$

1.2.2 The case of metric measure spaces

A compact metric space $(X,d_X) \in \mathcal{M}$ is called a *metric measure space* (mm-space for short) if X is endowed with a probability measure μ_X such that its support is all of X: $\mathrm{supp}[\mu_X] = X$. We denote by $\mathcal{M}_w$ the collection of all mm-spaces and by $\mathcal{M}_{k,w}$ the collection of all metric measure spaces of cardinality at most k.

Given $X \in \mathcal{M}_w$, $k \in \mathbf{N}$, $1 \le p \le \infty$ and $M_k \in \mathcal{M}_{k,w}$, we consider the choice of sketching cost function $\Psi_p(X,M_k) = d_{GWp}(X,M_k)$. Here d_{GWp} is the p-Gromov-Wasserstein distance, see section 3.1 for details. The resulting sketching objective is defined as

$$\mathrm{Sketch}_{k,p}(X) = \mathrm{Sketch}_k^{\Psi_p}(X) = \inf_{(M_k,d_k,\mu_k)} d_{GWp}(X,M_k).$$

Given $X \in \mathcal{M}_w$, $1 \le p < \infty$ and $B \subseteq X$, define *p-diameter* of B as

$$\mathrm{diam}_p(B) := \left(\int_B \int_B d^p(x,x') \frac{d\mu_X(x)\, d\mu_X(x')}{(\mu_X(B))^2} \right)^{1/p}.$$

For $p = \infty$, define $\mathrm{diam}_\infty(B) := \sup_{x,x' \in B} d_X(x,x')$.

Given $X \in \mathcal{M}_w$, $k \in \mathbf{N}$, $1 \le p < \infty$, $1 \le q < \infty$ and $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, we consider the choice of clustering cost function $\Phi_{p,q}(X,P) = \left(\sum_i \mathrm{diam}_p^q(B_i)\, \mu_X(B_i)\right)^{1/q}$. The resulting clustering objective is defined as

$$\mathrm{Shatter}_{k,p,q}(X) = \inf_{P \in \mathrm{Part}_k(X)} \Phi_{p,q}(X,P) = \inf_{P \in \mathrm{Part}_k(X)} \left(\sum_i \mathrm{diam}_p^q(B_i)\, \mu_X(B_i) \right)^{1/q}.$$

For $1 \le p \le \infty$, $q = \infty$ and any $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, we consider $\Phi_{p,\infty}(X,P) = \max_i \mathrm{diam}_p(B_i)$ and define the clustering objective as

$$\mathrm{Shatter}_{k,p,\infty}(X) = \inf_{P \in \mathrm{Part}_k(X)} \Phi_{p,\infty}(X,P) = \inf_{P \in \mathrm{Part}_k(X)} \max_i \mathrm{diam}_p(B_i).$$

We remark that $\mathrm{Shatter}_{k,1,1}(X)$ is within a factor of 2 from the k-median objective, and $\mathrm{Shatter}_{k,2,2}(X)$ is within a factor of 2 from the k-means objective. Similarly, $\mathrm{Shatter}_{k,p,p}(X)$ is within a factor of 2 from the ℓ_p-facility location objective, where the goal is to find some subset K of k points that minimizes $\left(\sum_{x \in X} (d(x,K))^p w(x)\right)^{1/p}$, $w(x)$ being some weight function on X.

We show the following bound.

Theorem 1.2. $\text{Sketch}_{k,p}(X) \leq \text{Shatter}_{k,p,\infty}(X)$ *for all* $k \in \mathbf{N}$ *and* $1 \leq p < \infty$.

We also show that, in general, $\text{Sketch}_{k,p}$ and $\text{Shatter}_{k,p,\infty}$ are *not* dual (Theorem 3.6).

The above negative result motivates the following definition of a stronger sketching objective for metric measure spaces. Given $X \in \mathcal{M}_w$, $k \in \mathbf{N}$, $1 \leq p \leq \infty$ and $M_k \in \mathcal{M}_{k,w}$, we consider the choice of sketching objective $\Psi_p^S(X, M_k) = \zeta_p(X, M_k)$. Here $\zeta_p(X, M_k)$ is the Sturm's version of p-Gromov Wasserstein distance, see section 3.3 for details. The resulting sketching objective is defined as

$$\text{Sketch}_{k,p}^S(X) = \text{Sketch}_k^{\Psi_p^S}(X) = \inf_{(M_k, d_k, \mu_k)} \zeta_p(X, M_k).$$

Our main duality result for metric measure spaces can now be stated as follows.

Theorem 1.3 (Duality for mm-spaces)**.** *For all* $(X, d_X, \mu_X) \in \mathcal{M}_w$*,* $k \in \mathbf{N}$ *and* $1 \leq p < \infty$*, we have*

$$\frac{1}{2} \cdot \text{Shatter}_{k,p,p}(X) \leq \text{Sketch}_{k,p}^S(X) \leq \text{Shatter}_{k,p,p}(X).$$

We also show a relation between the weak and strong sketching objectives for the case of mm-spaces of bounded doubling dimension and bounded diameter. In this case, $\text{Sketch}_{k,p}^S(X)$ can be bounded by a function of $\text{Sketch}_{k,p}(X)$.

Theorem 1.4. *Let* $X \in \mathcal{M}_w$ *with doubling constant* $C > 0$ *and* $p \in [1, \infty)$*. Then,*

$$\delta \leq \text{Sketch}_{k,p}^S(X) < \left(8 \cdot \text{diam}(X) \cdot \delta^{1/(5 \log_2 C)} + \delta^{1/5}\right)^{1/p} \cdot M$$

whenever $\text{Sketch}_{k,p}(X) = \delta < 2^{-5}$*. Here* $M = 2 \cdot \text{diam}(X) + 45$*.*

1.2.3 Computational considerations

We also consider computational problems that arise in the context of optimizing the sketching objective, either exactly or approximately. Interestingly, we obtain polynomial-time approximation algorithms for various sketching objectives, by combining the duality theorems with known approximation algorithms for clustering. We obtain algorithms via this duality paradigm both for metric spaces [Section 2.3] and for mm-spaces [Section 3.6]. We complement the above results with NP-hardness proofs [Section 2.2].

1.2.4 Impossibility results

We finally show that a certain clustering objective does not admit any "natural" sketching dual. More precisely, we consider the clustering objective of maximizing the minimum inter-cluster distance. We define a notion of "natural" sketching objective (called *admissible*), that follows from a few simple axioms. We show that no natural sketching objective is dual to the above clustering objective. See Section 5 for details.

1.3 Other prior work

We remark that there are clustering problems that involve a different objective than the ones considered here. One such example is partitioning a metric space into clusters of diameter at most R, for some given $R > 0$, minimizing the number of clusters. This is known as the R-dominating set problem [36, 17]. It is not known whether similar duality results can be obtained in this case.

We also note that there are some deep results in the sketching literature that do not fit within our broad definition of sketching a point x in some metric space M. Such examples include sketching information divergences [20], and certain quadratic forms [2].

1.4 Organization

This paper is organized as follows. Section 2 proves strict duality between $\mathrm{Sketch}_k(X)$ and $\mathrm{Shatter}_k(X)$, and presents results on NP-hardness and 2-approximation of $\mathrm{Sketch}_k(X)$. Section 3 proves duality between $\mathrm{Sketch}_{k,p}(X)$ and $\mathrm{Shatter}_{k,p,\infty}(X)$ for some values of k and p, but shows that, in general, $\mathrm{Sketch}_{k,p}(X)$ and $\mathrm{Shatter}_{k,p,\infty}(X)$ are not dual to each other. This section then proves duality between $\mathrm{Sketch}^S_{k,p}(X)$ and $\mathrm{Shatter}_{k,p,p}(X)$ for all values of k and p. Section 4 proves a relation between $\mathrm{Sketch}_{k,p}(X)$ and $\mathrm{Sketch}^S_{k,p}(X)$ for doubling metric measure spaces. Section 5 presents examples of clustering cost functions that do not admit a dual sketching cost function.

2 The case of metric spaces

In this section, we prove that for all $X \in \mathcal{M}$ and $k \in \mathbf{N}$, $\mathrm{Sketch}_k(X)$ is strictly dual to $\mathrm{Shatter}_k(X)$. In section 2.2, we show that in general, the computation of $\mathrm{Sketch}_k(X)$ is NP-Hard. However, as a result of the duality, in section 2.3, we obtain a polynomial time approximation of $\mathrm{Sketch}_k(X)$. In section 2.4, we show that an optimal sketch of a metric space may not be any of its subspaces.

Given $(X, d_X), (Y, d_Y) \in \mathcal{M}$, a *correspondence* R between X and Y is a subset $R \subset X \times Y$ such that $\pi_1(R) = X$ and $\pi_2(R) = Y$. Here π_1 and π_2 are the projection maps. Let $\mathcal{R}(X, Y)$ denote the set of all correspondences between X and Y. Given a correspondence R between (X, d_X) and (Y, d_Y), its *distortion* is $\mathrm{dis}(R) := \sup_{(x,y),(x',y')\in R} \left|d_X(x, x') - d_Y(y, y')\right|$.

The *Gromov-Hausdorff distance* between $(X, d_X), (Y, d_Y) \in \mathcal{M}$ is defined as follows:

$$d_{GH}(X, Y) := \frac{1}{2} \inf_{R\in\mathcal{R}(X,Y)} \mathrm{dis}(R).$$

The Gromov-Hausdorff distance is a metric on $\mathcal{M}$ and satisfies triangle inequality, symmetry, positivity, and $d_{GH}(X, Y) = 0$ if and only if X and Y are isometric [10]. A fact is that in general the computation of GH distance leads to NP-hard problems. Furthermore, any $(1 + \epsilon)$-approximation is also NP-hard, see [40, 42, 16].

Given $(X, d_X) \in \mathcal{M}$, the *Hausdorff distance* between subsets C and D of X is defined as $d^X_H(C, D) := \inf\{\epsilon > 0 \mid C \subset D^\epsilon \,\mathrm{and}\, D \subset C^\epsilon\}$ where C^ϵ is the set of points in X such that their distance to some point in C is at most ϵ.

Definition 2.1 (Voronoi Map). *Let $(X, d_X) \in \mathcal{M}$ and $k \in \mathbf{N}$. We define a map $\mathcal{V}_{X,k} : \mathcal{M}_k \to \mathrm{Part}_k(X)$ as follows: for any $(M_k, d_k) \in \mathcal{M}_k$, let $R \in \mathcal{R}(X, M_k)$ be such that $d_{GH}(X, M_k) = \frac{1}{2} \cdot \mathrm{dis}(R)$, i.e. R achieves the smallest distortion [12]. Then, R naturally induces a surjective map,* *say $\phi : X \to M_k$. We consider the partition $P = \{B_1, \dots, B_k\} \in \mathrm{Part}_k(X)$ where, for each $i \in [k]$, we take $B_i = \phi^{-1}(i)$. We now set*

$$\mathcal{V}_{X,k}((M_k, d_k)) = \{B_i\}_{i=1}^k.$$

Definition 2.2 (**Hausdorff Map**). *Let $(X, d_X) \in \mathcal{M}$ and $k \in \mathbf{N}$. We define a map $\mathcal{H}_{X,k}$: $\mathrm{Part}_k(X) \to \mathcal{M}_k$ as follows: for any $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, we define a metric d_k on M_k by $d_k(i,j) = d_H^X(B_i, B_j)$ where $d_H^X(B_i, B_j)$ is the Hausdorff distance in X between sets B_i and B_j. We now set*

$$\mathcal{H}_{X,k}(P) = (M_k, d_k).$$

2.1 The strict duality between $\mathrm{Sketch}_k(X)$ and $\mathrm{Shatter}_k(X)$

In this section, we show that $\mathrm{Sketch}_k(X)$ and $\mathrm{Shatter}_k(X)$ are strictly dual to each other. In addition, we provide a scheme for obtaining an optimal sketch of a metric space from an optimal clustering of that metric space and vice-versa.

Lemma 2.3. *For every $(X, d_X) \in \mathcal{M}$, for every $k \in \mathbf{N}$ and for every $(M_k, d_k) \in \mathcal{M}_k$, we have*

$$\Phi(X, \mathcal{V}_{X,k}(M_k)) \le 2 \cdot d_{GH}(X, M_k).$$

Proof. Given $(X, d_X) \in \mathcal{M}$, $k \in \mathbf{N}$ and $(M_k, d_k) \in \mathcal{M}_k$, let $R \in \mathcal{R}(X, M_k)$ be such that $d_{GH}(X, M_k) = \frac{1}{2} \cdot \mathrm{dis}(R)$. Then, by definition, R induces a surjective map, say $\phi : X \to M_k$ where for every $x \in X$, $\phi(x) = i$ with $i \in M_k$ being such that $(x, i) \in R$. Note that given $x \in X$ there may be multiple i with $(x, i) \in R$. In such a case, we arbitrarily pick an i and assign it to $\phi(x)$. Clearly, $(X, \phi(X)) \subset R$. We have

$$\mathrm{dis}(\phi) = \sup_{x,x' \in X} \left| d_X(x, x') - d_k(\phi(x), \phi(x')) \right| \le \sup_{(x,i),(x',i') \in R} \left| d_X(x, x') - d_k(i, i') \right|.$$

Thus we get that $\mathrm{dis}(\phi) \le \mathrm{dis}(R) = 2 \cdot d_{GH}(X, M_k)$. We now consider the partition $P = \{B_1, \dots, B_k\}$ of X where $B_i = \phi^{-1}(i)$ for each $i \in [k]$. For every $i \in [k]$, we have

$$\begin{aligned}\mathrm{diam}(B_i) &= \max\nolimits_{x,x' \in B_i} d_X(x, x') \\ &= \sup_{x,x' \in B_i} \left| d_X(x, x') - d_k(\phi(x), \phi(x')) \right| \\ &\le \sup_{x,x' \in X} \left| d_X(x, x') - d_k(\phi(x), \phi(x')) \right|.\end{aligned}$$

Thus we get that $\mathrm{diam}(B_i) \le \mathrm{dis}(\phi) \le 2 \cdot d_{GH}(X, M_k)$. Therefore, $\max_i \mathrm{diam}(B_i) \le 2 \cdot d_{GH}(X, M_k)$. Since $\mathcal{V}_{X,k}(M_k) = P$, we get that $\Phi(X, \mathcal{V}_{X,k}(M_k)) \le 2 \cdot d_{GH}(X, M_k)$. □

Lemma 2.4. *For every $(X, d_X) \in \mathcal{M}$, for every $k \in \mathbf{N}$ and for every $P \in \mathrm{Part}_k(X)$, we have*

$$2 \cdot d_{GH}(X, \mathcal{H}_{X,k}(P)) \le \Phi(X, P).$$

Proof. Given $(X, d_X) \in \mathcal{M}$, $k \in \mathbf{N}$ and $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, let $\eta = \Phi(X, P) = \max_i \mathrm{diam}(B_i)$. We define a metric d_k on M_k as follows:

$$d_k(i, j) = d_H^X(B_i, B_j)$$

where $d_H^X(B_i, B_j)$ is the Hausdorff distance in X between sets B_i and B_j. We define a map $\phi : X \to M_k$ as $\phi(x) = i$, where i is such that $x \in B_i$. This map is surjective and its graph $R(\phi) := \{(x, \phi(x)) \mid x \in X\}$ is a correspondence between (X, d_X) and (M_k, d_k). This is because every $x \in X$ belongs to some cluster B_i and all clusters are non-empty.

For $(x, i), (y, j) \in R(\phi)$ with $i = j$, we have

$$\big|d_X(x, y) - d_k(i, j)\big| = d_X(x, y) \leq \mathrm{diam}(B_i) \leq \eta.$$

For $(x, i), (y, j) \in R(\phi)$ with $i \neq j$, we have

$$\big|d_X(x, y) - d_k(i, j)\big| = \big|d_X(x, y) - d_H^X(B_i, B_j)\big|.$$

For $1 \leq i, j \leq k$, we define

$$d_X(B_i, B_j) = \min_{x \in B_i} \min_{x' \in B_j} d_X(x, x').$$

Let $v_i \in B_i$ be such that $d_X(v_i, B_j)$ is the smallest and $v_j \in B_j$ be such that $d_X(v_j, B_i)$ is the smallest. Then we have $d_X(v_i, B_j) = d_X(v_j, B_i) = d_X(B_i, B_j)$.

Let $u_i \in B_i$ be such that $d_X(u_i, B_j)$ is the largest and $u_j \in B_j$ be such that $d_X(u_j, B_i)$ is the largest. Then we observe that

$$d_H^X(B_i, B_j) = \max\{d_X(u_i, B_j), d_X(u_j, B_i)\}.$$

By triangle inequality, we have

$$d_X(u_i, B_j) \leq d_X(u_i, v_i) + d_X(v_i, B_j) \leq \mathrm{diam}(B_i) + d_X(B_i, B_j) \leq \eta + d_X(B_i, B_j).$$

Similarly, $d_X(u_j, B_i) \leq \eta + d_X(B_i, B_j)$. This implies that $d_H^X(B_i, B_j) \leq \eta + d_X(B_i, B_j)$. Since $d_H^X(B_i, B_j) \geq d_X(B_i, B_j)$, we get that for $(x, i), (y, j) \in R(\phi)$ with $i \neq j$,

$$\big|d_H^X(B_i, B_j) - d_X(x, y)\big| \leq \big|\eta + d_X(B_i, B_j) - d_X(x, y)\big| \leq \eta.$$

The second inequality holds because

$$d_X(B_i, B_j) \leq d_X(x, y) \leq \eta + d_H^X(B_i, B_j).$$

The leftmost inequality holds because

$$d_X(x, y) \leq d_X(x, B_j) + \mathrm{diam}(B_j) \leq d_X(u_i, B_j) + \eta \leq d_H^X(B_i, B_j) + \eta.$$

Thus for the correspondence $R(\phi)$, we get $\mathrm{dis}(R(\phi)) \leq \eta$. This gives that $d_{GH}(X, M_k) \leq \frac{\eta}{2}$. Since $\mathcal{H}_{X,k}(P) = (M_k, d_k)$, we get that $2 \cdot d_{GH}(X, \mathcal{H}_{X,k}(P)) \leq \Phi(X, P)$. □

Proof of Theorem 1.1. Fix $(X, d_X) \in \mathcal{M}$ and $k \in \mathbf{N}$. Suppose $\mathrm{Sketch}_k(X) < \eta$. Then, there exists $(M_k, d_k) \in \mathcal{M}_k$ such that $d_{GH}(X, M_k) < \eta$. From Lemma 2.3, we have that

$$\Phi(X, \mathcal{V}_{X,k}(M_k)) \leq 2 \cdot d_{GH}(X, M_k) < 2\eta.$$

Since $\mathcal{V}_{X,k}(M_k) \in \mathrm{Part}_k(X)$ and $\mathrm{Shatter}_k(X) = \inf_{P \in \mathrm{Part}_k(X)} \Phi(X, P)$, we get that $\mathrm{Shatter}_k(X) < 2\eta$. Thus, we have shown that whenever $\mathrm{Sketch}_k(X) < \eta$, we get $\mathrm{Shatter}_k(X) < 2\eta$. We conclude that $\mathrm{Shatter}_k(X) \leq 2 \cdot \mathrm{Sketch}_k(X)$.

We now prove the opposite direction of the last inequality. Suppose $\text{Shatter}_k(X) < \eta$. Then, there exists $P \in \text{Part}_k(X)$ such that $\Phi(X, P) < \eta$. From Lemma 2.4, we have that

$$2 \cdot d_{GH}(X, \mathcal{H}_{X,k}(P)) \leq \Phi(X, P) < \eta.$$

Since $\mathcal{H}_{X,k}(P) \in \mathcal{M}_k$ and $\text{Sketch}_k(X) = \inf_{(M_k,d_k)\in\mathcal{M}_k} d_{GH}(X, M_k)$, we have that $2\cdot\text{Sketch}_k(X) < \eta$. Thus, we have shown that whenever $\text{Shatter}_k(X) < \eta$, we get $2 \cdot \text{Sketch}_k(X) < \eta$. We conclude that $2 \cdot \text{Sketch}_k(X) \leq \text{Shatter}_k(X)$. □

A direct corollary of Theorem 1.1 is the following.

Corollary 2.5. *We have that*

- *If* $(M_k, d_k) \in \mathcal{M}_k$ *is s.t.* $\text{Sketch}_k(X) = d_{GH}(X, M_k)$*, then* $\text{Shatter}_k(X) = \Phi(X, \mathcal{V}_{X,k}(M_k))$.
- *If* $P = \{B_i\}_{i=1}^k \in \text{Part}_k(X)$ *is s.t.* $\text{Shatter}_k(X) = \Phi(X, P)$*, then* $\text{Sketch}_k(X) = d_{GH}(X, \mathcal{H}_{X,k}(P))$.

Remark 2.1. *Note that Corollary 2.5 along with Lemmas 2.3 and 2.4 tells us that for any* $(X, d_X) \in \mathcal{M}$ *and* $k \in \mathbf{N}$*, given an optimal sketch* (M_k, d_k) *of* X *and a correspondence* R *between* X *and* M_k *that achieves the smallest distortion [12], we can determine an optimal clustering of* X*. On the other hand, given an optimal clustering* $P = \{B_i\}_{i=1}^d \in \text{Part}_k(X)$*, an optimal sketch of* X *can be obtained by calculating the Hausdorff distance between sets* B_i*,* $i \in [k]$*.*

2.2 The computation of $\text{Sketch}_k(X)$ is NP-hard

We show that an instance of the Set Cover Decision problem can be reduced to an instance of computing $\text{Shatter}_k(X)$. Since we have $\text{Sketch}_k(X) = \frac{1}{2} \cdot \text{Shatter}_k(X)$, computation of $\text{Sketch}_k(X)$ is NP-hard.

Definition 2.6 (Set Cover Decision Problem). *Given a universe of elements* $U = \{u_1, \ldots, u_n\}$*, a collection* $S = \{S_1, \ldots, S_m\}$ *of subsets of* U *such that* $\cup_{i=1}^m S_i = U$ *and* $k \in \mathbf{N}$*, the set cover decision problem asks if there is a collection* $\{S_{i_1}, \ldots, S_{i_l}\} \subseteq S$ *such that* $\cup_{j=1}^l S_{i_j} = U$ *and* $l \leq k$*.*

Given an instance of a set cover decision problem with input U, S and $k \in \mathbf{N}$, we construct a graph G as follows: For every $u_i \in U$, add a vertex to G. By abuse of notation, we refer to this vertex as u_i. For every $S_i \in S$, add a vertex to G and refer to this vertex as S_i. We add two more vertices r and r' to G. We now add edges to join the vertices of G. We add an edge of length 1 joining r and r'. For every $i \in [m]$, we add an edge of length 1 from r to every S_i. For every $i \in [n]$ and $j \in [m]$, if $u_i \in S_j$ then we add an edge of length 1 between vertices u_i and S_j in G. Let $V(G)$ denote the set of vertices of G and $E(G)$ denote the set of edges of G.

Theorem 2.7. *There is a set cover of* U *of size* k *if and only if there exists a partition of* G *into* $k+1$ *blocks with the diameter of every block being at most* 2*.*

Proof. Suppose there is a set cover of U of size k. Let $S' = \{S_{i_1}, \ldots, S_{i_k}\}$ be such a set cover. Then $\cup_{j=1}^k S_{i_j} = U$. We construct a partition of G as follows: For $j \in [k]$ and $S_{i_j} \in S'$, let $B_j = S_{i_j} \cup \{u \in U \mid u \in S_{i_j}\}$. Let $B_{k+1} = V(G) \setminus \cup_{j=1}^k B_j$. It is easy to check that the for every $j \in [k+1]$, $\text{diam}(B_j) \leq 2$. We observe that $\text{Shatter}_{k+1}(G) = 2$. Thus the partition $\{B_1, \ldots, B_{k+1}\}$ is an optimal partition.

Conversely, suppose that there exists a partition P of G into $k+1$ blocks such that the diameter of every block is at most 2. We may assume that all blocks are connected.

We first show that such a partition P can be modified to ensure that that there is no block that is a singleton containing some $u \in U$.

Claim 2.8. *There is no block in partition P that is a singleton containing some $u \in U$.*

Proof. Suppose P has a block B such that $B = \{u\}$ for some $u \in U$. Then we consider any set S_i such that $u \in S_i$. Let B' be the block such that $S_i \in B'$. There are five cases:

1. $B' = S_i$. In this case, we merge B and B' into one block.
2. B' contains $u' \in U$, $u' \neq u$, but B' neither contains r nor any other S_j, $j \neq i$. In this case, we merge B' and B into a single block.
3. B' contains both r and r'. In this case, if B' contains any $S_j \neq S_i$ then we remove S_i from B' and add S_i to B. If B' does not contain any other S_j, $j \neq i$ then we remove S_i from B' and add S_i to B.
4. B' does not contain r but contains some $S_j \neq S_i$. This happens when there is a $u' \in U$ such that $u' \in B'$ and $u' \in S_j \cap S_i$. In this case, we remove S_i from B' and add it to B.
5. B' contains r but not r'. This means that r' is in a block by itself. We remove r from B and add to the block of r'. Now we are in one of the above cases.

We observe that in all the above cases, we do not change the maximum diameter of the partition. Therefore we can assume that there is no block in P that is a singleton containing some $u \in U$. □

We now show that we can modify the partition P such that vertices r and r' belong to the same block.

Claim 2.9. *Vertices r and r' belong to the same block of partition P.*

Proof. Let $B_{r'}$ be the block containing the vertex r'. Suppose $B_{r'}$ does not contain r. Since the blocks are connected, we have $B_{r'} = \{r'\}$. Let B_r be the block containing the vertex r. If $B_r = \{r\}$, we merge B_r and $B_{r'}$ to a single block. If B_r contains some $u \in U$, then it also contains some $S_i \in S$. In this case, we remove r from B_r and add r to $B_{r'}$. Note that removing r from B_r does not disconnect B_r. If B_r does not contains any $u \in U$ but contains some $S_i \in S$, then we merge B_r and $B_{r'}$ to a single cluster. We observe that any of the above steps do not change the maximum diameter of partition P. □

We now show that the k blocks of P obtained by excluding the block containing vertices r and r' give a set cover of U size k. We observe that the block containing r does not contain any $u \in U$ because this block also contains r' and the diameter of the partition is at most 2. Therefore, all $u \in U$ are covered by the remaining k blocks. Since the blocks are connected and there is no block that is a singleton containing some $u \in U$, we get that every block contains some $S_i \in S$. Let L be the set of blocks of P containing exactly one $S_i \in S$ and M be the set of blocks of P containing more than one S_i. We observe that for every block $B \in M$, every $u \in U \cap B$ is contained in every $S_i \in B$. We define a set cover H as follows: for every block $B \in L$, add the unique $S_i \in B$ to H. For every block $B \in M$, pick some $S_i \in B$ arbitrarily and add to H. The set H contains k elements of S that cover U. Thus we get a set cover of U of size k. □

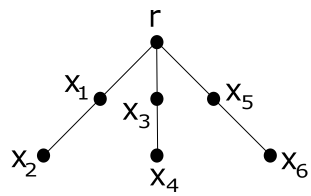

Figure 1: The space corresponding to $m = 1$ in Theorem 2.11.

2.3 A polynomial time approximation of $\mathrm{Sketch}_k(X)$

Theorem 2.10. *Given $X \in \mathcal{M}$ and $k \in \mathbf{N}$, there is a polynomial time algorithm that outputs a 2-approximation of* $\mathrm{Sketch}_k(X)$.

Proof. There is a polynomial time greedy algorithm by Gonzalez [19] that gives a 2-approximation of $\mathrm{Shatter}_k(X)$. Let $P \in \mathrm{Part}_k(X)$ be the partition of X obtained from this greedy algorithm. We apply the Hausdorff map (Definition 2.2) to P to obtain a k-point metric space. From Remark 2.1, we get that the k-point metric space obtained is a 2-approximation of $\mathrm{Sketch}_k(X)$. We note that computing the k-point metric space takes polynomial time, since the computation involves calculating Hausdorff distance between blocks of partition P, and this can be done in polynomial time. □

2.4 Study of minimizers of $d_{GH}(X, M_k)$

We show that for any $m \in \mathbf{N}$, there exists a metric space (X, d_X) such that the minimizer of $d_{GH}(X, M_{3m})$ is *not* a subset of X. We first take X to be a rooted tree.

Theorem 2.11 (Tree metric spaces)**.** *For all $m \in \mathbf{N}$ and $n = 3m$, there exist (tree) metric spaces (X, d_X) and (Y, d_Y) such that (1) $|Y| = n$, and (2) For any $K \subset X$ with $|K| \leq n$, we have $d_{GH}(X, K) > d_{GH}(X, Y)$.*

Proof. Let $m = 1$. Let $X = \{x_1, \ldots, x_6\}$ be a rooted tree with three branches. Let r denote the root. For $i = \{1, 2, 3\}$, the i-th branch contains x_{2i-1}, x_{2i} with $d_X(r, x_{2i-1}) = 1$ and $d_X(x_{2i-1}, x_{2i}) = 1$. Then we have

$$d_X(x_1, x_2) = d_X(x_3, x_4) = d_X(x_5, x_6) = 1,$$
$$d_X(x_1, x_3) = d_X(x_1, x_5) = d_X(x_3, x_5) = 2,$$
$$d_X(x_2, x_4) = d_X(x_2, x_6) = d_X(x_4, x_6) = 4,$$
$$d_X(x_1, x_4) = d_X(x_1, x_6) = d_X(x_3, x_2) = d_X(x_3, x_6) = d_X(x_5, x_2) = d_X(x_5, x_4) = 3.$$

Let $Y = \{y_1, y_2, y_3\}$ be such that y_i lies on the i-th branch of X. Moreover we have for $i = \{1, 2, 3\}$, $d_Y(r, y_i) = \frac{3}{2}$.

We now show that for every $K \subset X$ with $|K| \leq 3$, $d_{GH}(X, K) > d_{GH}(X, Y)$.

We first calculate $d_{GH}(X, Y)$. Let R be a correspondence between X and Y. Then R induces a surjective map, say ϕ from X to Y. Suppose we have $\phi(x_1) = \phi(x_2) = y_1, \phi(x_3) = \phi(x_4) = y_2$ and $\phi(x_5) = \phi(x_6) = y_3$. Then $\mathrm{dis}(R(\phi)) = 1$. This implies that $d_{GH}(X, Y) \leq \frac{1}{2}$. Moreover we have $d_{GH}(X, Y) \geq \frac{1}{2}|\mathrm{diam}(X) - \mathrm{diam}(Y)| \geq \frac{1}{2}$. Thus we get $d_{GH}(X, Y) = \frac{1}{2}$. Now in order to prove the theorem, it suffices to show that for every $K \subset X$ with $|K| \leq 3$, $d_{GH}(X, K) > \frac{1}{2}$. We have the following cases:

1. $K = \{x_1, x_2, x_3\}$.

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq 3$; if $\phi(x_6) = x_2$ we get $\text{dis}(R(\phi)) \geq 4$ and if $\phi(x_6) = x_3$ we get $\text{dis}(R(\phi)) \geq 3$. This implies that $d_{GH}(X, K) \geq \frac{3}{2}$.

2. $K = \{x_1, x_2, x_4\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq 3$; if $\phi(x_6) = x_2$ we get $\text{dis}(R(\phi)) \geq 4$ and if $\phi(x_6) = x_4$ we get $\text{dis}(R(\phi)) \geq 4$. This implies that $d_{GH}(X, K) \geq \frac{3}{2}$.

3. $K = \{x_1, x_3, x_5\}$

   We use that $d_{GH}(X, K) \geq \frac{1}{2}|\text{diam}(X) - \text{diam}(K)|$. This gives $d_{GH}(X, K) \geq 1$.

4. $K = \{x_2, x_4, x_6\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_1) = x_2, \phi(x_3) = x_4$ and $\phi(x_5) = x_6$. This implies that $\text{dis}(R(\phi)) = 2$. Thus $d_{GH}(X, K) = 1$.

5. $K = \{x_1, x_3, x_6\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_2) = x_1, \phi(x_4) = x_3$ and $\phi(x_5) = x_6$. This implies that $\text{dis}(R(\phi)) = 2$. Thus $d_{GH}(X, K) = 1$.

6. $K = \{x_2, x_4, x_5\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_1) = x_2, \phi(x_3) = x_4$ and $\phi(x_6) = x_5$. This implies that $\text{dis}(R(\phi)) = 2$. Thus $d_{GH}(X, K) = 1$.

7. $K = \{x_1, x_2\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq 3$ and if $\phi(x_6) = x_2$, we get $\text{dis}(R(\phi)) \geq 4$. Thus $d_{GH}(X, K) \geq \frac{3}{2}$.

8. $K = \{x_1, x_3\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq 3$ and if $\phi(x_6) = x_3$, we again get $\text{dis}(R(\phi)) \geq 3$. Thus $d_{GH}(X, K) \geq \frac{3}{2}$.

9. $K = \{x_1, x_4\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq 3$ and if $\phi(x_6) = x_4$, we get $\text{dis}(R(\phi)) \geq 4$. Thus $d_{GH}(X, K) \geq \frac{3}{2}$.

10. $K = \{x_2, x_4\}$

    Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_2$, we get $\text{dis}(R(\phi)) \geq 4$ and if $\phi(x_6) = x_4$, we again get $\text{dis}(R(\phi)) \geq 4$. Thus $d_{GH}(X, K) \geq 2$.

11. $K \subset X$ with $|K| = 1$

    In this case $d_{GH}(X, K) = \frac{1}{2}\text{diam}(X) = 2$.

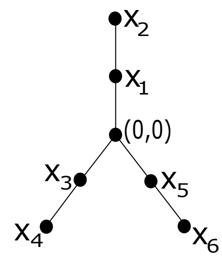

Figure 2: The space corresponding to $m = 1$ in Theorem 2.12.

The analysis for the remaining subsets K of X is symmetric to the cases above. This proves the theorem for $n = 3$. In general for any $m \in \mathbf{N}$ and $n = 3m$, we create m copies of X that are connected at the root r and prove the theorem similarly. $\square$

Now, we prove that we have a similar behavior even for subsets of Euclidean space.

Theorem 2.12 (Euclidean metric spaces)**.** *For all $m \in \mathbf{N}$ and $n = 3m$, there exist $X, Y \subset \mathbf{R}^2$ such that (1) $|Y| = n$, and (2) For any $K \subset X$ with $|K| \leq n$, we have $d_{GH}(X, K) > d_{GH}(X, Y)$.*

Proof. Let $m = 1$. We construct the set X on lines similar to those in the previous theorem. Let $x_1 = (0, 1), x_2 = (0, 2), x_3 = (-\frac{\sqrt{3}}{2}, -\frac{1}{2}), x_4 = (-\sqrt{3}, -1), x_5 = (\frac{\sqrt{3}}{2}, -\frac{1}{2}), x_6 = (\sqrt{3}, -1)$. Then, we have

$$d(x_1, x_2) = d(x_3, x_4) = d(x_5, x_6) = 1,$$
$$d(x_1, x_3) = d(x_1, x_5) = d(x_3, x_5) = \sqrt{3},$$
$$d(x_2, x_4) = d(x_2, x_6) = d(x_4, x_6) = 2\sqrt{3},$$
$$d(x_1, x_4) = d(x_1, x_6) = d(x_3, x_2) = d(x_3, x_6) = d(x_5, x_2) = d(x_5, x_4) = \sqrt{7}.$$

Let $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. Then $\operatorname{diam}(X) = 2\sqrt{3}$. Let $y_1 = \left(0, \frac{3}{2}\right), y_2 = \left(\frac{-3\sqrt{3}}{4}, -\frac{3}{4}\right)$ and $y_3 = \left(\frac{3\sqrt{3}}{4}, -\frac{3}{4}\right)$. Let $Y = \{y_1, y_2, y_3\}$. We now show that for every $K \subset X$ with $|K| = 3$, $d_{GH}(X, K) > d_{GH}(X, Y)$.

We first calculate $d_{GH}(X, Y)$. Let R be a correspondence between X and Y. Then R induces a surjective map, say ϕ from X to Y. Suppose we have $\phi(x_1) = \phi(x_2) = y_1, \phi(x_3) = \phi(x_4) = y_2$ and $\phi(x_5) = \phi(x_6) = y_3$. Then $\operatorname{dis}(R(\phi)) = 1$. This implies that $d_{GH}(X, Y) \leq \frac{1}{2}$. Now in order to prove the theorem, it suffices to show that for every $K \subset X$ with $|K| \leq 3$, $d_{GH}(X, K) > \frac{1}{2}$. We have the following cases:

1. $K = \{x_1, x_2, x_3\}$.

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \mathrm{id}$. Then if $\phi(x_6) = x_1$, we get $\operatorname{dis}(R(\phi)) \geq \sqrt{7}$; if $\phi(x_6) = x_2$ we get $\operatorname{dis}(R(\phi)) \geq 2\sqrt{3}$ and if $\phi(x_6) = x_3$ we get $\operatorname{dis}(R(\phi)) \geq \sqrt{7}$. This implies that $d_{GH}(X, K) \geq \frac{\sqrt{7}}{2}$.

2. $K = \{x_1, x_2, x_4\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \mathrm{id}$. Then if $\phi(x_6) = x_1$, we get $\operatorname{dis}(R(\phi)) \geq \sqrt{7}$; if $\phi(x_6) = x_2$ we get $\operatorname{dis}(R(\phi)) \geq 2\sqrt{3}$ and if $\phi(x_6) = x_4$ we get $\operatorname{dis}(R(\phi)) \geq 2\sqrt{3}$. This implies that $d_{GH}(X, K) \geq \frac{\sqrt{7}}{2}$.

3. $K = \{x_1, x_3, x_5\}$

   We use that $d_{GH}(X, K) \geq \frac{1}{2}|\text{diam}(X) - \text{diam}(K)|$. This gives $d_{GH}(X, K) \geq \frac{\sqrt{3}}{2}$.

4. $K = \{x_2, x_4, x_6\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_1) = x_2, \phi(x_3) = x_4$ and $\phi(x_5) = x_6$. This implies that $\text{dis}(R(\phi)) = \sqrt{3}$. Thus $d_{GH}(X, K) = \frac{\sqrt{3}}{2}$.

5. $K = \{x_1, x_3, x_6\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_2) = x_1, \phi(x_4) = x_3$ and $\phi(x_5) = x_6$. This implies that $\text{dis}(R(\phi)) = \sqrt{3}$. Thus $d_{GH}(X, K) = \frac{\sqrt{3}}{2}$.

6. $K = \{x_2, x_4, x_5\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then it is trivial to see that $\text{dis}(R(\phi))$ is minimized if $\phi(x_1) = x_2, \phi(x_3) = x_4$ and $\phi(x_6) = x_5$. This implies that $\text{dis}(R(\phi)) = \sqrt{3}$. Thus $d_{GH}(X, K) = \frac{\sqrt{3}}{2}$.

7. $K = \{x_1, x_2\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq \sqrt{7}$ and if $\phi(x_6) = x_2$, we get $\text{dis}(R(\phi)) \geq 2\sqrt{3}$. Thus $d_{GH}(X, K) \geq \frac{\sqrt{7}}{2}$.

8. $K = \{x_1, x_3\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq \sqrt{7}$ and if $\phi(x_6) = x_3$, we again get $\text{dis}(R(\phi)) \geq \sqrt{7}$. Thus $d_{GH}(X, K) \geq \frac{\sqrt{7}}{2}$.

9. $K = \{x_1, x_4\}$

   Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_1$, we get $\text{dis}(R(\phi)) \geq \sqrt{7}$ and if $\phi(x_6) = x_4$, we get $\text{dis}(R(\phi)) \geq 2\sqrt{3}$. Thus $d_{GH}(X, K) \geq \frac{\sqrt{7}}{2}$.

10. $K = \{x_2, x_4\}$

    Let $\phi : X \to K$ be a surjective map such that $\phi|_K = \text{id}$. Then if $\phi(x_6) = x_2$, we get $\text{dis}(R(\phi)) \geq 2\sqrt{3}$ and if $\phi(x_6) = x_4$, we again get $\text{dis}(R(\phi)) \geq 2\sqrt{3}$. Thus $d_{GH}(X, K) \geq \sqrt{3}$.

11. $K \subset X$ with $|K| = 1$

    In this case $d_{GH}(X, K) = \frac{1}{2}\text{diam}(X) = \sqrt{3}$.

The analysis for the remaining subsets K of X is symmetric to the cases above. This proves the theorem for $n = 3$. In general for any $m \in \mathbf{N}$ and $n = 3m$, we create m disjoint copies of X and prove the theorem similarly. $\square$

3 The case of metric measure spaces

In this section, we prove that for every $X \in \mathcal{M}_{k,w}$ and $k \in \mathbf{N}$, $\mathrm{Sketch}_{k,p}(X)$ is dual to $\mathrm{Shatter}_{k,p,\infty}(X)$ for $k = 1$ and any finite p, and for $p = \infty$ and any finite k. However, we prove in 3.2 that, in general, these objectives are not dual to each other. Thus in 3.3, we define Sturm's version of p-Gromov Wasserstein distance and prove in section 3.4 that $\mathrm{Sketch}^S_{k,p}(X)$ is dual to $\mathrm{Shatter}_{k,p,p}(X)$ for all $1 \le p \le \infty$.

3.1 The weak Gromov-Wasserstein distance

In this section, we define the p-Gromov Wasserstein distance following [39].

We say that two mm-spaces X and Y are isomorphic whenever there exists a map $\phi : X \to Y$ such that (1) $d_X(x, x') = d_Y(\phi(x), \phi(x'))$ for all $x, x' \in X$, and (2) $\mu_X(\phi^{-1}(B)) = \mu_Y(B)$ for all B measurable.

Definition 3.1 (**Measure coupling**). *Let $X, Y \in \mathcal{M}_w$. A probability measure μ on the product space $X \times Y$ is called a coupling of μ_X and μ_Y iff we have the following:*

- *For all measurable sets $A \subseteq X$, $\mu(A, Y) = \mu_X(A)$ and,*
- *For all measurable sets $B \subseteq Y$, $\mu(X, B) = \mu_Y(B)$.*

For $X, Y \in \mathcal{M}_w$, let $\mathcal{U}(X, Y)$ denote the set of all couplings of μ_X and μ_Y. For $\mu \in \mathcal{U}(X, Y)$, let $R(\mu) = \mathrm{supp}[\mu]$.

Definition 3.2 (**p-distortion**). *Given $X, Y \in \mathcal{M}_w$ and $\mu \in \mathcal{U}(X, Y)$, define the p-distortion of μ as follows: For $1 \le p < \infty$, we have*

$$\mathrm{dis}_p(\mu) := \left(\int_{X\times Y} \int_{X\times Y} |d_X(x, x') - d_Y(y, y')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p}.$$

For $p = \infty$, we have $\mathrm{dis}_\infty(\mu) := \sup_{x,x' \in X,\ y,y' \in Y,\ (x,y),(x',y') \in R(\mu)} |d_X(x, x') - d_Y(y, y')|$.

Definition 3.3 (**p-Gromov Wasserstein Distance**). *Given $X, Y \in \mathcal{M}_w$ and $1 \le p \le \infty$, define the p-Gromov Wasserstein distance between X and Y as $d_{GWp}(X, Y) := \frac{1}{2} \inf_{\mu \in \mathcal{U}(X,Y)} \mathrm{dis}_p(\mu)$.*

It is known [39] that the p-Gromov Wasserstein distance defines a proper metric on the collection of isomorphism classes of mm-spaces.

3.2 Relationship between $\mathrm{Sketch}_{k,p}$ and $\mathrm{Shatter}_{k,p,\infty}$

In this section, we show the relation between $\mathrm{Sketch}_{k,p}$ and $\mathrm{Shatter}_{k,p,\infty}$ for various values of k and p. It turns out that these are strictly dual to each other for certain values of p and k:

Theorem 3.4 (The case of $k = 1$ and p finite). *For all $X \in \mathcal{M}_w$ and $1 \le p < \infty$, we have* $\mathrm{Sketch}_{1,p}(X) = \frac{1}{2} \cdot \mathrm{Shatter}_{1,p,\infty}(X)$.

Proof. We have $\mathrm{Sketch}_{1,p}(X) = \inf_{(M_1,d_1,\mu_1)} d_{GW,p}(X, M_1)$. There is a unique measure μ_1 that can be defined on a one point metric space and hence a unique coupling between μ_X and μ_1. Thus we get that $\mathrm{Sketch}_{1,p}(X) = d_{GW,p}(X, M_1) = \frac{1}{2} \cdot \mathrm{diam}_p(X)$. We have $\mathrm{Shatter}_{1,p,\infty}(X) = \mathrm{diam}_p(X)$ and this proves the theorem. □

Theorem 3.5 (The case of k finite and $p = \infty$). *For all $X \in \mathcal{M}_w$ and $k \in \mathbf{N}$, we have* $\text{Sketch}_{k,\infty}(X) = \frac{1}{2}\text{Shatter}_{k,\infty,\infty}(X)$.

Proof. We observe that $d_{GW\infty}(X, M_k) \geq d_{GH}(X, M_k)$. Therefore, we get that

$$\text{Sketch}_{k,\infty}(X) = \inf_{(M_k,d_k,\mu_k)} d_{GW\infty}(X, M_k) \geq \inf_{(M_k,d_k)} d_{GH}(X, M_k) = \frac{1}{2} \cdot \text{Shatter}_k(X).$$

Since $\text{Shatter}_k(X) = \text{Shatter}_{k,\infty,\infty}(X)$, we get that $\text{Sketch}_{k,\infty}(X) \geq \frac{1}{2} \cdot \text{Shatter}_{k,\infty,\infty}(X)$. We now prove the opposite inequality. Let $\text{Shatter}_{k,\infty,\infty}(X) \leq \eta$. Then there exists a partition $\{B_1, \cdots B_k\}$ of X such that for all $i \in [k]$, $\text{diam}(B_i) \leq \eta$. Let (M_k, d_k, μ_k) be the k point metric space with $d_k(i, j) = d_H^X(B_i, B_j)$ and $\mu_k(i) = \mu_X(B_i)$. Let γ be the probability measure on $X \times M_k$ defined by: $\gamma(A \times \{i\}) = \mu_X(A \cap B_i)$ for all measurable sets $A \subseteq X$ and $i \in M_k$. Clearly, $\gamma(X \times \{i\}) = \mu_X(B_i) = \mu_k(i)\ \forall\ i \in M_k$. Now, for all $A \subset X$ measurable, $\gamma(A \times M_k) = \sum_{i\in[k]} \mu_X(A \cap B_i) = \mu_X(A)$. So, γ is a coupling between μ_X and μ_k.

Now, we have

$$d_{GW\infty}(X, M_k) \leq \frac{1}{2} \sup_{\substack{x,x' \in X \\ i,j \in P_k \\ (x,i),(x',j) \in supp(\mu)}} |d_X(x, x') - d_k(i, j)| \leq \frac{1}{2} \sup_{\substack{x \in B_i \\ x' \in B_j}} |d_X(x, x') - d_k(i, j)|.$$

We have the following two cases:

- **Case** $1 : i = j$

  Here we have $\sup_{x,x'\in B_i} d_X(x, x') = \max_i \text{diam}(B_i) \leq \eta$.

- **Case** $2 : i \neq j$

  Here we have

  $$\sup_{x\in B_i,y\in B_j} |d_X(x, y) - d_k(i, j)| = \sup_{x\in B_i,y\in B_j} |d_X(x, y) - d_H^X(B_i, B_j)|.$$

  Now, $d_H^X(B_i, B_j) \leq \eta + d_X(B_i, B_j)$, where $d_X(B_i, B_j) = \inf_{x\in B_i,y\in B_j} d_X(x, y)$. From the proof of Lemma 2.4, we know that for any $x \in B_i$ and $y \in B_j$,

  $$d_X(B_i, B_j) \leq d_X(x, y) \leq \eta + d_H^X(B_i, B_j)$$

  and $d_X(B_i, B_j) \leq d_H^X(B_i, B_j) \leq \eta + d_X(B_i, B_j)$. This gives $|d_X(x, y) - d_H^X(B_i, B_j)| \leq \eta$. Thus, $d_{GW\infty}(X, M_k) \leq \frac{\eta}{2}$. So $\text{Sketch}_{k,\infty}(X) \leq \frac{\eta}{2}$. Thus, we get that $\text{Sketch}_{k,\infty}(X) \leq \frac{1}{2}\text{Shatter}_{k,\infty,\infty}(X)$.

□

Proof of Theorem 1.2. Let $\text{Shatter}_{k,p,\infty}(X) \leq \eta$. This means that there exists a partition $B_1, \ldots, B_k$ of X such that for $1 \leq i \leq k$, $\text{diam}_p(B_i) \leq \eta$. Consider a set of k points M_k. We define a metric d_k on M_k as follows:

$$d_k(i, j) = d_{W_p}(\mu_{B_i}, \mu_{B_j})$$

where metric on B_i is $d_X|_{B_i}$ and $\mu_{B_i} = \frac{\mu_X|_{B_i}}{\mu_X(B_i)}$. The measure μ_k on M_k is defined as follows: $\mu_k(i) = \mu_X(B_i)$ for each $i \in [k]$. Let γ be the probability measure on $X \times M_k$ defined by: $\gamma(A \times \{i\}) = \mu_X(A \cap B_i)$ for all measurable sets $A \subseteq X$ and $i \in M_k$. Clearly, $\gamma(X \times \{i\}) = \mu_X(B_i) = \mu_k(i)\ \forall\ i \in M_k$. Now, for all $A \subset X$ measurable, $\gamma(A \times M_k) = \sum_{i\in[k]} \mu_X(A \cap B_i) = \mu_X(A)$. So, γ is a valid coupling between μ_X and μ_k. Now,

$$\begin{aligned}
d_{GWp}(X, M_k) &\leq \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_X(x,y) - d_k(i,j)|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&\leq \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_X(x,y) - d_{W_p}(\mu_{B_i},\mu_{B_j})|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&\leq \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_{W_p}(\delta_x,\delta_y) - d_{W_p}(\mu_{B_i},\mu_{B_j})|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&\leq \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_{W_p}(\delta_x,\mu_{B_i}) + d_{W_p}(\delta_y,\mu_{B_j})|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&\leq \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_{W_p}(\delta_x,\mu_{B_i})|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&+ \frac{1}{2}\left(\iint\limits_{(X\times M_k)\times(X\times M_k)} |d_{W_p}(\delta_y,\mu_{B_j})|^p d\gamma(x,i)d\gamma(y,j)\right)^{1/p} \\
&\leq \frac{1}{2}\left(\sum_i \int\limits_{B_i} |d_{W_p}(\delta_x,\mu_{B_i})|^p d\mu_X(x)\right)^{1/p} + \frac{1}{2}\left(\sum_j \int\limits_{B_j} |d_{W_p}(\delta_y,\mu_{B_j})|^p d\mu_X(y)\right)^{1/p} \\
&= \left(\sum_i \int\limits_{B_i} |d_{W_p}(\delta_x,\mu_{B_i})|^p d\mu_X(x)\right)^{1/p}
\end{aligned}$$

Here, the fourth inequality is due to triangle inequality and the fifth inequality is due to Minkowski's inequality. Now, $d_{W_p}(\delta_x, \mu_{B_i}) = \left(\int_{y\in B_i} d(x,y)^p \frac{d\mu_X(y)}{\mu_X(B_i)}\right)^{1/p}$. So, we get

$$\left(\sum_i \int\limits_{B_i} |d_{W_p}(\delta_x,\mu_{B_i})|^p d\mu_X(x)\right)^{1/p} = \left(\sum_i \int_{x\in B_i}\int_{y\in B_i} d(x,y)^p d\mu_X(x) \frac{d\mu_X(y)}{\mu_X(B_i)}\right)^{1/p}.$$

Now, $\mathrm{diam}_p(B_i) = \left(\int_{x\in B_i}\int_{y\in B_i} d(x,y)^p \frac{d\mu_X(x)}{\mu_X(B_i)}\frac{d\mu_X(y)}{\mu_X(B_i)}\right)^{1/p}$. Therefore, we get

$$\left(\sum_i \int_{x\in B_i}\int_{y\in B_i} d(x,y)^p d\mu_X(x)\frac{d\mu_X(y)}{\mu_X(B_i)}\right)^{1/p} = \left(\sum_i \mathrm{diam}_p^p(B_i)\cdot\mu_X(B_i)\right)^{1/p}.$$

We know that for every i, $\mathrm{diam}_p(B_i) \le \eta$ and $\sum_i \mu_X(B_i) = 1$. So we get

$$\left(\sum_i \mathrm{diam}_p^p(B_i)\cdot\mu_X(B_i)\right)^{1/p} \le \eta.$$

We have shown that whenever $\mathrm{Shatter}_{k,p,\infty}(X) \le \eta$, we get $\mathrm{Sketch}_{k,p}(X) \le \eta$. Thus, $\mathrm{Sketch}_{k,p}(X) \le \mathrm{Shatter}_{k,p,\infty}(X)$. □

In Theorem 1.2, we showed that $\mathrm{Sketch}_{k,p}(X) \le \mathrm{Shatter}_{k,p,\infty}(X)$ in general. In general, however, these objectives are not dual to each other as we show next.

$\mathrm{Shatter}_{k,p,\infty}$ **is not dual to** $\mathrm{Sketch}_{k,p}$. Let $\Delta_m = ([m], d_m, \mu_m)$ denote the metric measure space with m points such that (1) $d_m(i,j) = 1$ for $i \neq j$ and $d_m(i,j) = 0$ for $i = j$, and (2) for every $i \in [m]$, we have $\mu_m(i) = \frac{1}{m}$.

Theorem 3.6. *For every constant $C_1 > 0$ and every $r \in \mathbf{N}$ such that $r \ge 2$, there exists $k \in \mathbf{N}$ such that for $m = rk$, we have $C_1 \cdot \mathrm{Shatter}_{k,p,\infty}(\Delta_m) > \mathrm{Sketch}_{k,p}(\Delta_m)$, for every $1 \le p < \infty$.*

Proof. Fix $p \in [1,\infty)$. We first show that for any $r, k \in \mathbf{N}$ and $m = rk$, we have $\mathrm{Shatter}_{k,p,\infty}(\Delta_m) = \left(1-\frac{k}{m}\right)^{\frac{1}{p}}$. We know that for any $B \subseteq \Delta_m$ with $|B| = l$, we have $\mathrm{diam}_p(B) = \left(1-\frac{1}{l}\right)^{\frac{1}{p}}$. Therefore for any partition $\{B_1,\dots,B_k\}$ of Δ_m with $|B_i| = l_i$, we have $\max_i \mathrm{diam}_p(B_i) = \max_i\left(1-\frac{1}{l_i}\right)^{\frac{1}{p}} = \left(1-\frac{1}{\max_i l_i}\right)^{\frac{1}{p}}$. Thus,

$$\mathrm{Shatter}_{k,p,\infty}(\Delta_m) = \min_{\substack{(l_1,\dots,l_k)\\ \sum_{i=1}^k l_i = m}} \max_i \left(1-\frac{1}{l_i}\right)^{\frac{1}{p}} = \left(1-\frac{1}{\min_{\substack{(l_1,\dots,l_k)\\ \sum_{i=1}^k l_i = m}} \max_i l_i}\right)^{\frac{1}{p}}.$$

Since $m = rk$, we get that

$$\min_{\substack{(l_1,\dots,l_k)\\ \sum_{i=1}^k l_i = m}} \max_i l_i = \frac{m}{k} = r.$$

Thus, we get $\mathrm{Shatter}_{k,p,\infty}(\Delta_m) = \left(1-\frac{k}{m}\right)^{\frac{1}{p}}$.

From the definition of $\mathrm{Sketch}_{k,p}(X)$, we know that $\mathrm{Sketch}_{k,p}(\Delta_m) \le d_{GWp}(\Delta_m,\Delta_k)$. We use claim 5.1 from [39] to get that $d_{GWp}(\Delta_k,\Delta_m) \le \frac{1}{2}\left(\frac{1}{k}+\frac{1}{m}\right)^{\frac{1}{p}}$. This gives that $\mathrm{Sketch}_{k,p}(\Delta_m) \le \frac{1}{2}\left(\frac{1}{k}+\frac{1}{m}\right)^{\frac{1}{p}}$.

Now we fix a constant $C_1 > 0$ and $r \in \mathbf{N}$ such that $r \geq 2$. Let $k \in \mathbf{N}$ be such that $k > \frac{1+r^{-1}}{2^p \cdot C_1^p \cdot (1-r^{-1})}$. Then, we have that

$$C_1 \cdot (1-r^{-1})^{1/p} > \frac{1}{2 \cdot k^{1/p}} \cdot (1+r^{-1})^{1/p}.$$

This implies that $C_1 \cdot \mathrm{Shatter}_{k,p,\infty}(\Delta_m) > \mathrm{Sketch}_{k,p}(\Delta_m)$. □

From the definition of duality (equation (1)) and the above theorem, we conclude that $\mathrm{Shatter}_{k,p,\infty}$ is not dual to $\mathrm{Sketch}_{k,p}$.

3.3 Using Sturm's version of the Gromov-Wasserstein distance

In this section, we define another sketching objective for metric measure spaces, $\mathrm{Sketch}^S_{k,p}(X)$ using Sturm's definition of p-Gromov Wasserstein distance [43]. We show that $\mathrm{Sketch}^S_{k,p}(X)$ is dual to the clustering objective $\mathrm{Shatter}_{k,p,p}$, for all $k \in \mathbf{N}$ and $1 \leq p \leq \infty$.

Definition 3.7 (**Metric Coupling**). *A metric coupling between $X, Y \in \mathcal{M}_w$ is a metric d on the disjoint union $X \sqcup Y$ such that $d|_X = d_X$ and $d|_Y = d_Y$. For $X, Y \in \mathcal{M}_w$, let $\mathcal{D}(X,Y)$ denote the set of all metric couplings between X and Y.*

Definition 3.8 (**p-Wasserstein distance**, [46])**.** *Let $(X, d_X, \mu_X) \in \mathcal{M}_w$ and $A, B \subset X$ be compact subsets of X. Let μ_A, μ_B be Borel probability measures with $\mathrm{supp}[\mu_A] = A$ and $\mathrm{supp}[\mu_B] = B$. The p-Wasserstein distance between (A, μ_A) and (B, μ_B) is defined as*

$$d^X_{W_p}(\mu_A, \mu_B) := \inf_{\mu \in \mathcal{U}(\mu_A, \mu_B)} \left(\int_{A \times B} d^p_X(a,b) d\mu(a,b) \right)^{1/p},$$

for $1 \leq p < \infty$, and $d^X_{W_\infty}(\mu_A, \mu_B) := \inf_{\mu \in \mathcal{U}(\mu_A,\mu_B)} \sup_{(a,b) \in R(\mu)} d_X(a,b)$.

Definition 3.9 (**Sturm's version of p-Gromov Wasserstein distance**). *[43] Given $X, Y \in \mathcal{M}_w$ and $1 \leq p < \infty$, define*

$$\zeta_p(X,Y) := \inf_{\substack{d \in \mathcal{D}(X,Y) \\ \mu \in \mathcal{U}(X,Y)}} \left(\int_{X \times Y} d^p(x,y)\, d\mu(x,y) \right)^{1/p}.$$

For $p = \infty$, define

$$\zeta_\infty(X,Y) := \inf_{\substack{d \in \mathcal{D}(X,Y) \\ \mu \in \mathcal{U}(X,Y)}} \sup_{(x,y) \in R[\mu]} d(x,y).$$

It was shown in [39] that for all $X, Y \in \mathcal{M}_w$ and $1 \leq p \leq \infty$, $\zeta_p(X,Y) \geq d_{GW_p}(X,Y)$ and $\zeta_\infty(X,Y) = d_{GW_\infty}(X,Y)$.

Definition 3.10 (**Voronoi map for mm-spaces**). *Let $(X, d_X, \mu_X) \in \mathcal{M}_w$, $k \in \mathbf{N}$ and $\epsilon > 0$. We define a map $\mathcal{V}^w_{X,k,\epsilon} : \mathcal{M}_{k,w} \to \mathrm{Part}_k(X)$ as follows: for any $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$, let $d_\epsilon \in \mathcal{D}(d_X, d_k)$ and $\mu_\epsilon \in \mathcal{U}(\mu_X, \mu_k)$ be such that $\zeta_p(X, M_k) + \epsilon \geq \left(\int_{X \times M_k} d^p_\epsilon(x,i) d\mu_\epsilon(x \times i) \right)^{1/p}$. Let $P = \{B'_i\}_{i=1}^k$*

be the Voronoi partition of $X \sqcup M_k$ *with respect to* M_k *under the metric* d_ϵ. *Let* $B_i = B'_i \setminus \{i\}$. *We set*

$$\mathcal{V}^w_{X,k,\epsilon}((M_k, d_k, \mu_k)) = \{B_i\}_{i=1}^k.$$

Note that when X *is finite, we do not need an* ϵ, *since we can obtain* $d \in \mathcal{D}(d_X, d_k)$ *and* $\mu \in \mathcal{U}(\mu_X, \mu_k)$ *such that* $\zeta_p(X, M_k) = \left(\int_{X \times M_k} d^p(x,i) d\mu(x \times i)\right)^{1/p}$.

Definition 3.11 (Wasserstein map for mm-spaces). *Let* $(X, d_X, \mu_X) \in \mathcal{M}_w$ *and* $k \in \mathbf{N}$. *We define a map* $\mathcal{H}^w_{X,k} : \mathrm{Part}_k(X) \to \mathcal{M}_{k,w}$ *as follows: for any* $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$, *we define* (M_k, d_k, μ_k) *as*

$$d_k(i,j) = d^X_{W_p}(\mu_{B_i}, \mu_{B_j}) \ \forall\ i, j \in [k],\ i \neq j \quad \text{and} \quad \mu_k(i) = \mu_X(B_i) \ \forall\ i \in [k].$$

Here $d^X_{W_p}(\mu_{B_i}, \mu_{B_j})$ *is the* p*-Wasserstein distance (Definition 3.8) between* B_i *and* B_j *in* X *and* $\mu_{B_i} = \frac{\mu_X|_{B_i}}{\mu_X(B_i)}$ *for all* $i \in [k]$, *i.e.* μ_{B_i} *arises as the renormalization of the measure* μ_X *when restricted to* B_i. *We now set*

$$\mathcal{H}^w_{X,k}(P) = (M_k, d_k, \mu_k).$$

3.4 Relationship between $\mathrm{Sketch}^S_{k,p}$ and $\mathrm{Shatter}_{k,p,p}$

We have the following lemmas regarding the Voronoi map and the Wasserstein map for metric measure spaces.

Lemma 3.12. *For every* $X \in \mathcal{M}_w$, $k \in \mathbf{N}$, $M_k \in \mathcal{M}_{k,w}$, $\epsilon > 0$ *and* $1 \leq p < \infty$, *we have*

$$\Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k)) \leq 2 \cdot \zeta_p(X, M_k) + 2\epsilon.$$

Claim 3.13. *Given* $(X, d_X, \mu_X) \in \mathcal{M}_w$, $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ *and a coupling* d *between* d_X *and* d_k, *let* $\{B'_1, \ldots, B'_k\}$ *denote the Voronoi partition of* $X \sqcup M_k$ *with respect to* M_k *under the metric* d. *Let* $B_i = B'_i \setminus \{i\}$. *Define* $(M_k, d_k, \mu_k^{\mathbf{V}})$ *where for every* $i \in [k]$, $\mu_k^{\mathbf{V}}(i) = \mu_X(B_i)$. *Then, there exists a coupling* μ' *between* μ_X *and* $\mu_k^{\mathbf{V}}$ *such that* $\mathrm{supp}[\mu'] = \cup_{i \in [k]}(B_i \times \{i\})$ *i.e.* $(x,i) \in \mathrm{supp}[\mu']$ *if and only if* $x \in B_i$, *and for every coupling* μ *between* μ_X *and* μ_k,

$$\left(\int_{X \times M_k} d^p(x,i) d\mu'(x \times i)\right)^{1/p} \leq \left(\int_{X \times M_k} d^p(x,i) d\mu(x \times i)\right)^{1/p}.$$

Proof of Claim 3.13. Let $Z = X \sqcup M_k$. Let d be a metric on Z such that d is a coupling between d_X and d_k. Let μ be any coupling between μ_X and μ_k.

Since $\{B'_1, B'_2, \ldots, B'_k\}$ is a Voronoi partition of Z with respect to M_k, we have that if $x \in B'_i$, then $d(x,i) \leq d(x,j)$ for all $j \neq i$.

Let μ' be the probability measure on $X \times M_k$ defined by: $\mu'(A \times \{i\}) = \mu_X(A \cap B_i)$.

We first claim that $\mu' \in \mathcal{U}(\mu_X, \mu_k^{\mathbf{V}})$. Clearly, $\mu'(X \times \{i\}) = \mu_X(B_i) = \mu_k^{\mathbf{V}}(i)\ \forall\ i \in M_k$. Now, for all $A \subseteq X$ measurable, $\mu'(A \times M_k) = \sum_{i \in [k]} \mu_X(A \cap B_i) = \mu_X(A)$. So, μ' is a coupling between μ_X and $\mu_k^{\mathbf{V}}$. We observe that $\mathrm{supp}[\mu'] = \cup_{i \in [k]}(B_i \times \{i\})$.

Now, to prove the claim write:

$$\begin{aligned}\int_{X\times P_k} d^p(x,i)\,\mu(dx\times\{i\}) &= \sum_i \int_X d^p(x,i)\,\mu(dx\times\{i\})\\ &= \sum_i\sum_j \int_{B_j} d^p(x,i)\,\mu(dx\times\{i\})\\ &\geq \sum_i\sum_j \int_{B_j} d^p(x,j)\mu(dx\times\{i\})\\ &= \sum_j \int_{B_j} d^p(x,j)\mu_X(dx)\\ &= \int_{X\times P_k} d^p(x,j)\mu'(dx\times\{j\}).\end{aligned}$$

□

Proof of Lemma 3.12. Given $(X,d_X,\mu_X)\in\mathcal{M}_w$, $(M_k,d_k,\mu_k)\in\mathcal{M}_{k,w}$ and $\epsilon>0$, let $d_\epsilon\in\mathcal{D}(d_X,d_k)$ and $\mu_\epsilon\in\mathcal{M}(\mu_X,\mu_k)$ be such that

$$\zeta_p(X,M_k)+\epsilon\geq\left(\int\limits_{X\times M_k} d_\epsilon^p(x,i)d\mu_\epsilon(x\times i)\right)^{1/p}.$$

By claim 3.13, there exists a partition $P=\{B_i\}_{i=1}^k$ of X, an $(M_k,d_k,\mu_k^{\mathbf{V}})\in\mathcal{M}_{k,w}$ and, a coupling $\mu'_\epsilon\in\mathcal{U}(\mu_X,\mu_k^{\mathbf{V}})$ such that $\mathrm{supp}[\mu'_\epsilon]=\cup_{i\in[k]}(B_i\times\{i\})$, and it satisfies

$$\left(\int\limits_{X\times M_k} d_\epsilon^p(x,i)d\mu_\epsilon(x\times i)\right)^{1/p}\geq\left(\int\limits_{X\times M_k} d_\epsilon^p(x,i)d\mu'_\epsilon(x\times i)\right)^{1/p}.$$

We consider the partition $P=\{B_i\}_{i=1}^k$ of X. Then, we have

$$\left(\int_{X\times M_k} d_\epsilon^p(x,i)\,d\mu'_\epsilon(x\times i)\right)^{1/p}=\left(\sum_i\int_{B_i} d_\epsilon^p(x,i)\,d\mu_X(x)\right)^{1/p}\leq\zeta_p(X,M_k)+\epsilon.$$

From the above equation, we get

$$\begin{aligned}2\cdot\zeta_p(X,M_k)+2\epsilon &\geq \left(\sum_i\int_{B_i} d_\epsilon^p(x,i)d\mu_X(x)\right)^{1/p}+\left(\sum_i\int_{B_i} d_\epsilon^p(x',i)d\mu_X(x')\right)^{1/p}\\ &= \left(\sum_i\iint\limits_{B_i\times B_i}\frac{d_\epsilon^p(x,i)d\mu_X(x)d\mu_X(x')}{\mu_X(B_i)}\right)^{1/p}+\left(\sum_i\iint\limits_{B_i\times B_i}\frac{d_\epsilon^p(x',i)d\mu_X(x')d\mu_X(x)}{\mu_X(B_i)}\right)^{1/p}\\ &\geq \left(\sum_i\iint\limits_{B_i\times B_i}\frac{(d_\epsilon(x,i)+d_\epsilon(x',i))^p d\mu_X(x)d\mu_X(x')}{\mu_X(B_i)}\right)^{1/p}\end{aligned}$$

The last inequality is due to Minkowski's inequality. From triangle inequality, we get

$$2 \cdot \zeta_p(X, M_k) + 2\epsilon \geq \left(\sum_i \frac{1}{\mu_X(B_i)} \iint\limits_{B_i \times B_i} d_\epsilon^p(x, x') d\mu_X(x) d\mu_X(x') \right)^{1/p}$$
$$= \left(\sum_i \operatorname{diam}_p^p(B_i)\, \mu_X(B_i) \right)^{1/p}.$$

Thus, we have

$$2 \cdot \zeta_p(X, M_k) + 2\epsilon \geq \left(\sum_i \operatorname{diam}_p^p(B_i)\, \mu_X(B_i) \right)^{1/p}.$$

Since $\mathcal{V}^w_{X,k,\epsilon}(M_k) = \{B_i\}_{i=1}^k$ and $\Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k)) = \left(\sum_i \operatorname{diam}_p^p(B_i)\, \mu_X(B_i)\right)^{1/p}$, we get that $\Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k)) \leq 2 \cdot \zeta_p(X, M_k) + 2\epsilon$. □

Lemma 3.14. *For every $X \in \mathcal{M}_w$, $k \in \mathbf{N}$, $P \in \operatorname{Part}_k(X)$ and $1 \leq p < \infty$, we have*

$$\zeta_p(X, \mathcal{H}^w_{X,k}(P)) \leq \Phi_{p,p}(X, P).$$

Proof. Given $(X, d_X, \mu_X) \in \mathcal{M}_w$ and $P = \{B_i\}_{i=1}^k \in \operatorname{Part}_k(X)$, let

$$\eta = \Phi_{p,p}(X, P) = \left(\sum_i \operatorname{diam}_p^p(B_i)\, \mu_X(B_i) \right)^{1/p}.$$

Let (M_k, d_k, μ_k) be a k-point metric measure space such that

$$d_k(i, j) = d^X_{W_p}\left(\frac{\mu_X|_{B_i}}{\mu_X(B_i)}, \frac{\mu_X|_{B_j}}{\mu_X(B_j)} \right) \ \forall\ i, j \in [k],\ i \neq j \quad \text{and} \quad \mu_k(i) = \mu_X(B_i) \ \forall\ i \in [k].$$

Define d on $X \sqcup M_k$ as follows:

$$d(x, i) = d_{W_p}\left(\delta_x, \frac{\mu_X|_{B_i}}{\mu_X(B_i)} \right) \ \forall\ x \in X,\ i \in [k].$$

It follows from the definition that d satisfies triangle inequality. Let μ be a probability measure on $X \times M_k$ defined by: $\mu(A \times \{i\}) = \mu_X(A \cap B_i)$. Clearly, $\mu(X \times \{i\}) = \mu_X(B_i) = \mu_k(i)\ \forall\ i \in M_k$. Now, for all $A \subseteq X$ measurable, $\mu(A \times M_k) = \sum_{i \in [k]} \mu_X(A \cap B_i) = \mu_X(A)$. So, μ is a coupling between μ_X and μ_k.

Now, we have

$$\int_{X \times P_k} d^p(x, i) d\mu(x, i) = \sum_i \int_{B_i} d^p(x, i) \mu_X(dx) = \sum_i \int_{B_i} d^p_{W_p}\left(\delta_x, \frac{\mu_X|_{B_i}}{\mu_X(B_i)} \right) \mu_X(dx)$$
$$= \sum_i \int_{B_i} \int_{B_i} d^p(x, y) \frac{\mu_X(dy)}{\mu_X(B_i)} \mu_X(dx)$$
$$= \sum_i \mu_X(B_i) \iint\limits_{B_i \times B_i} d^p(x, y) \frac{\mu_X(dx) \mu_X(dy)}{\mu_X^2(B_i)}$$
$$= \sum_i \mu_X(B_i) \operatorname{diam}_p^p(B_i) \leq \eta^p.$$

This gives $\zeta_p(X, M_k) \leq \eta$. Since $\mathcal{H}^w_{X,k}(P) = (M_k, d_k, \mu_k)$, we get that $\zeta_p(X, \mathcal{H}^w_{X,k}(P)) \leq \Phi_{p,p}(X, P)$. $\square$

Using the above lemmas, we now prove the main result of this section.

Proof of Theorem 1.3. Fix $(X, d_X, \mu_X) \in \mathcal{M}_w$, $k \in \mathbf{N}$ and $1 \leq p < \infty$. Suppose $\mathrm{Sketch}^S_{k,p}(X) < \eta$. Then, there exists $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ such that $\zeta_p(X, M_k) < \eta$. Then, from Lemma 3.12, we have that, for every $\epsilon > 0$,

$$\Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k)) \leq 2 \cdot \zeta_p(X, M_k) + 2\epsilon < 2\eta + 2\epsilon.$$

Since $\mathcal{V}^w_{X,k,\epsilon}(M_k) \in \mathrm{Part}_k(X)$ and $\mathrm{Shatter}_{k,p,p}(X) \leq \Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k))$, we get that $\mathrm{Shatter}_{k,p,p}(X) < 2\eta + 2\epsilon$. This inequality holds true for every $\epsilon > 0$. Therefore, we conclude that $\mathrm{Shatter}_{k,p,p}(X) \leq 2\eta$. We have shown that whenever $\mathrm{Sketch}^S_{k,p}(X) < \eta$, we get $\mathrm{Shatter}_{k,p,p}(X) \leq 2\eta$. Hence, we have $\mathrm{Shatter}_{k,p,p}(X) \leq 2 \cdot \mathrm{Sketch}^S_{k,p}(X)$.

We now prove the second inequality. Let $\mathrm{Shatter}_{k,p,p}(X) < \eta$. Then, there exists $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$ such that $\Phi_{p,p}(X, P) < \eta$. Then, from Lemma 3.14, we have that

$$\zeta_p(X, \mathcal{H}^w_{X,k}(P)) \leq \Phi_{p,p}(X, P) < \eta.$$

Since $\mathcal{H}^w_{X,k}(P) \in \mathcal{M}_{k,w}$ and $\mathrm{Sketch}^S_{k,p}(X) \leq \zeta_p(X, \mathcal{H}^w_{X,k}(P))$, we get that $\mathrm{Sketch}^S_{k,p}(X) < \eta$. Thus, we have shown that whenever $\mathrm{Shatter}_{k,p,p}(X) < \eta$, we get $\mathrm{Sketch}^S_{k,p}(X) < \eta$. We conclude that $\mathrm{Sketch}^S_{k,p}(X) \leq \mathrm{Shatter}_{k,p,p}(X)$. $\square$

A direct corollary of Theorem 1.3 is the following.

Corollary 3.15. *We have that*

- *If* $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ *is such that* $\mathrm{Sketch}^S_{k,p}(X) = \zeta_p(X, M_k)$, *then for every* $\epsilon > 0$,

$$\mathrm{Shatter}_{k,p,p}(X) \leq \Phi_{p,p}(X, \mathcal{V}^w_{X,k,\epsilon}(M_k)) \leq 2 \cdot \mathrm{Shatter}_{k,p,p}(X) + 2\epsilon.$$

- *If* $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$ *is such that* $\mathrm{Shatter}_{k,p,p}(X) = \Phi_{p,p}(X, P)$, *then*

$$\mathrm{Sketch}^S_{k,p}(X) \leq \zeta_p(X, \mathcal{H}^w_{X,k}(P)) \leq 2 \cdot \mathrm{Sketch}^S_{k,p}(X).$$

Remark 3.1. *Note that Corollary 3.15 along with Lemmas 3.12 and 3.14 tells us that for any* $X \in \mathcal{M}_w$, $k \in \mathbf{N}$ *and* $1 \leq p < \infty$, *given* $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ *such that* $\mathrm{Sketch}^S_{k,p}(X) = \zeta_p(X, M_k)$, *and* $d \in \mathcal{D}(d_X, d_k)$, $\mu \in \mathcal{U}(\mu_X, \mu_k)$ *such that* $\zeta_p(X, M_k) = \left(\int_{X \times M_k} d^p(x, i) d\mu(x \times i)\right)^{1/p}$, *we can obtain a clustering of* X *that is a* 2*-approximation for* $\mathrm{Shatter}_{k,p,p}(X)$. *On the other hand, given* $P = \{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)$ *such that* $\mathrm{Shatter}_{k,p,p}(X) = \Phi_{p,p}(X, P)$, *a* 2*-approximation for* $\mathrm{Sketch}^S_{k,p}(X)$ *is obtained by calculating* p*-Wasserstein distance between the blocks of partition* P.

For $p = q = \infty$, we have the following strict duality result.

Theorem 3.16. *Let* $X \in \mathcal{M}_w$ *and* $k \in \mathbf{N}$. *For* $p = q = \infty$, *we have* $\frac{1}{2} \cdot \mathrm{Shatter}_{k,\infty,\infty}(X) = \mathrm{Sketch}^S_{k,\infty}(X)$.

Proof. For $p = \infty$, we use Theorem 5.1 from [39] that says for any $X, Y \in \mathcal{M}_w$, we have $d_{GW_\infty}(X, Y) = \zeta_\infty(X, Y)$. This gives that

$$\text{Sketch}^S_{k,\infty}(X) = \inf_{(P_k, d_k, \mu_k)} \zeta_\infty(X, P_k) = \inf_{(P_k, d_k, \mu_k)} d_{GW_\infty}(X, P_k) = \text{Sketch}_{k,\infty}(X).$$

From Theorem 3.5, we have that $\text{Sketch}_{k,\infty}(X) = \frac{1}{2} \cdot \text{Shatter}_{k,\infty,\infty}(X)$. This implies that $\text{Sketch}^S_{k,\infty}(X) = \frac{1}{2} \cdot \text{Shatter}_{k,\infty,\infty}(X)$. □

3.5 Computation of $\text{Sketch}^S_{k,p}$ and $\text{Shatter}_{k,p,p}$ for Δ_m

For the space $\Delta_m \in \mathcal{M}_w$, we can compute the objectives $\text{Sketch}^S_{k,p}$ and $\text{Shatter}_{k,p,p}$, which, in general, are difficult to compute. We show these computations in the following example.

Example 3.1. *For all $m \geq 2$, we have*

- $\displaystyle \frac{1}{2} \leq \frac{\text{Sketch}^S_{k,p}(\Delta_m)}{\text{Shatter}_{k,p,p}(\Delta_m)} \leq \frac{1}{2(1 - k \cdot m^{-1})^{1/p}}$ *for any $k < m$ and $1 \leq p \leq \infty$.*
- *Furthermore,* $\displaystyle \frac{\text{Sketch}^S_{1,1}(\Delta_m)}{\text{Shatter}_{1,1,1}(\Delta_m)} = \frac{1}{2(1 - m^{-1})}$*. This implies that the ratio* $\displaystyle \frac{\text{Sketch}^S_{1,1}(\Delta_m)}{\text{Shatter}_{1,1,1}(\Delta_m)}$ *takes infinitely many values in the interval* $\left[\frac{1}{2}, 1\right]$.

The above statements are proved in the following three lemmas.

Lemma 3.17. *For all $m, k \in \mathbf{N}$ with $k < m$ and every $p \geq 1$,* $\text{Shatter}_{k,p,p}(\Delta_m) = \left(1 - \frac{k}{m}\right)^{1/p}$.

Proof. Let $\{B_1, \cdots, B_k\}$ denote a partition of Δ_m into k blocks. For every $i \in [m]$, let $|B_i| = k_i$. Then $\sum_i k_i = m$. For any $p \geq 1$, $\text{diam}^p_p(B_i) = \sum_{x,y \in B_i} d^p_m(x, y) \cdot \frac{1}{m^2} \cdot \left(\frac{k_i}{m}\right)^{-2}$. Now, there are k_i^2 pairs of points x, y in B_i, out of which $k_i^2 - k_i$ pairs satisfy $d_m(x, y) = 1$. This gives $\text{diam}^p_p(B_i) = \frac{k_i^2 - k_i}{m^2} \cdot \frac{m^2}{k_i^2} = \frac{k_i - 1}{k_i}$. Now,

$$\sum_i \mu_m(B_i) \text{diam}^p_p(B_i) = \sum_i \frac{k_i}{m} \left(\frac{k_i - 1}{k_i}\right) = \frac{1}{m} \sum_i k_i - 1 = \frac{m - k}{m}.$$

The above equation holds true for any partition $\{B_1, \cdots, B_k\}$ of Δ_m. Thus we conclude that for any $m, k \in \mathbf{N}$ and $k < m$, $\text{Shatter}_{k,p,p}(\Delta_m) = \left(1 - \frac{k}{m}\right)^{1/p}$. □

Lemma 3.18. *For all $m, k \in \mathbf{N}$ with $k < m$ and for every $p \geq 1$,* $\frac{1}{2}\left(1 - \frac{k}{m}\right)^{1/p} \leq \text{Sketch}^S_{k,p}(\Delta_m) \leq \frac{1}{2}$.

Proof. We have $\text{Sketch}^S_{k,p}(\Delta_m) = \inf_{(M_k, d_k, \mu_k)} \zeta_p(\Delta_m, M_k)$. Let $P_k = \Delta_k$. Then, $\text{Sketch}^S_{k,p}(\Delta_m) \leq \zeta_p(\Delta_m, \Delta_k)$. Now, $\zeta_p(\Delta_m, \Delta_k) \leq \inf_{\mu,d} \left(\int d^p(x, y) d\mu(x \times y)\right)^{1/p}$, where $x \in \Delta_m,\ y \in \Delta_k$, d varies over all metric couplings between d_m and d_k and μ varies over all measure couplings between μ_m and μ_k. Let $d(x, y) = \frac{1}{2}$ for all $x \in \Delta_m,\ y \in \Delta_k$. It can be verified that d is a valid coupling. Let the coupling μ be the product measure. Then

$$\zeta_p(\Delta_m, \Delta_k) \leq \left(\sum_{x \in \Delta_m} \sum_{y \in \Delta_k} \frac{1}{2^p} \frac{1}{mk}\right)^{1/p} = \frac{1}{2}.$$

This establishes that $\text{Sketch}^S_{k,p}(\Delta_m) \le \frac{1}{2}$. The lower bound follows from Theorem **??** and Lemma 3.17. Therefore, we get

$$\frac{1}{2} \le \frac{\text{Sketch}^S_{k,p}(\Delta_m)}{\text{Shatter}^S_{k,p,p}(\Delta_m)} \le \frac{1}{2(1-k\cdot m^{-1})^{1/p}}.$$

□

Lemma 3.19. *For all $m \ge 2$, $\text{Sketch}^S_{1,1}(\Delta_m) > \frac{1}{2}\cdot \text{Shatter}_{1,1,1}(\Delta_m)$. Furthermore, we have:*

$$\frac{\text{Sketch}^S_{1,1}(\Delta_m)}{\text{Shatter}_{1,1,1}(\Delta_m)} = \frac{1}{2(1-m^{-1})}.$$

Proof. We have that $\text{Shatter}_{1,1,1}(\Delta_m) = \text{diam}_1(\Delta_m) = \frac{m^2-m}{m^2} = 1 - \frac{1}{m}$. We have $\text{Sketch}^S_{1,1}(\Delta_m) = \inf_{(M_1,d_1,\mu_1)} \zeta_1(\Delta_m, M_1) = \zeta_1(\Delta_m, M_1)$. This is because there is a unique one point mm-space. Moreover, there is a unique measure coupling between μ_{Δ_m} and μ_1 and that is the product measure. Let d be a metric coupling between d_{Δ_m} and d_1. Then, we have

$$\text{Sketch}^S_{1,1}(\Delta_m) = \inf_d \sum_{x\in\Delta_m} d(x,\{1\})\mu_{\Delta_m}(x) = \inf_d \sum_{x\in\Delta_m} \frac{d(x,\{1\})}{m}.$$

The distance d satisfies the inequality $d(x,x') \le d(x,\{1\}) + d(x',\{1\})$. There are $\binom{m}{2}$ pairs x, x' for which this inequality holds. We add all such inequalities to get

$$\binom{m}{2} \le (m-1)\sum_{x\in\Delta_m} d(x,\{1\}).$$

Thus, we get $\text{Sketch}^S_{1,1}(\Delta_m) \ge \frac{1}{2}$. Moreover, we know that $\text{Sketch}^S_{1,1}(\Delta_m) \le \frac{1}{2}$ from Lemma 3.18. Therefore, we have $\text{Sketch}^S_{1,1}(\Delta_m) = \frac{1}{2}$. However, $\frac{1}{2}\text{Shatter}_{1,1,1}(\Delta_m) = \frac{1}{2}\left(1-\frac{1}{m}\right) < \frac{1}{2}$ for $m \ge 2$. Thus, we conclude that $\text{Sketch}^S_{1,1}(\Delta_m) > \frac{1}{2}\cdot\text{Shatter}_{1,1,1}(\Delta_m)$. □

3.6 Approximation results for $\text{Sketch}^S_{k,p}$ for finite p

We use the duality between $\text{Sketch}^S_{k,p}$ and $\text{Shatter}_{k,p,p}$ established in the last section to obtain approximation results for $\text{Sketch}^S_{k,p}$ via the approximations for $\text{Shatter}_{k,p,p}$. We show the following result.

Theorem 3.20 (Approximation results for $\text{Sketch}^S_{k,p}$)**.** *For any $0 < \epsilon < 1$ and $t \in \mathbf{Z}$ such that $k \ge t > 1$, there is an $f(p)$-approximation for $\text{Sketch}^S_{k,p}(X)$, where $f(p)$ is as follows:*

1. *For $p = 1$, $f(p) = 12 + \frac{8}{t} + \epsilon$.*
2. *For $p = 2$, $f(p) = 20 + \frac{16}{t} + \epsilon$.*
3. *For all reals $p > 2$, $f(p) = \left(12 + \frac{8}{t}\right)p + \epsilon$.*

The running time of the approximation algorithm is $|X|^{O(t)}\cdot\epsilon^{-1}$.

We need the following preliminaries in order to prove the above theorem.

Definition 3.21. *For $p \in [1,\infty)$, given $X \in \mathcal{M}_w$ and a subset $B \subseteq X$, we define the p-radius of B as*

$$\mathrm{rad}_p(B) := \inf_{a\in B} \left(\int d_X^p(a,x) \frac{d\mu_X(x)}{\mu_X(B)} \right)^{1/p}.$$

For $p=\infty$, we define

$$\mathrm{rad}_\infty(B) = \inf_{a\in B} \sup_{x\in B} d_X(a,x).$$

Theorem 3.22. *For all $p \in [1,\infty]$, given a metric measure space (X, d_X, μ_X) and subset $B \subseteq X$,*

$$2\,\mathrm{rad}_p(B) \geq \mathrm{diam}_p(B) \geq \mathrm{rad}_p(B).$$

Proof. Let us first consider the case of finite p. For $B \subseteq X$, we have $\mathrm{diam}_p(B) = \left(\iint_{B\times B} d_X^p(x,x') \frac{d\mu_X(x) d\mu_X(x')}{\mu_X^2(B)} \right)^{1/p}$. For any $a \in X$, we have

$$\begin{aligned}
2\left(\int_B d_X^p(a,x) \frac{d\mu_X(x)}{\mu_X(B)} \right)^{1/p} &= \left(\int_B d_X^p(a,x) \frac{d\mu_X(x)}{\mu_X(B)} \right)^{1/p} + \left(\int_B d_X^p(a,x') \frac{d\mu_X(x')}{\mu_X(B)} \right)^{1/p} \\
&= \left(\iint\limits_{B\times B} d_X^p(a,x) \frac{d\mu_X(x) d\mu_X(x')}{\mu_X^2(B)} \right)^{1/p} + \left(\iint\limits_{B\times B} d_X^p(a,x') \frac{d\mu_X(x') d\mu_X(x)}{\mu_X^2(B)} \right)^{1/p} \\
&\geq \left(\iint\limits_{B\times B} d_X^p(x,x') \frac{d\mu_X(x) d\mu_X(x')}{\mu_X^2(B)} \right)^{1/p} \\
&= \mathrm{diam}_p(B).
\end{aligned}$$

This gives that $2\,\mathrm{rad}_p(B) \geq \mathrm{diam}_p(B)$. For the other inequality, we observe that

$$\begin{aligned}
\mathrm{diam}_p^p(B) &= \iint\limits_{B\times B} d_X^p(x,x') \frac{d\mu_X(x) d\mu_X(x')}{\mu_X^2(B)} \\
&= \int_B \int_B d_X^p(x,x') \frac{d\mu_X(x) d\mu_X(x')}{\mu_X^2(B)} \\
&= \int_B \frac{d\mu_X(x')}{\mu_X(B)} \int_B d_X^p(x,x') \frac{d\mu_X(x)}{\mu_X(B)} \\
&\geq \int_B \mathrm{rad}_p^p(B) \frac{d\mu_X(x')}{\mu_X(B)} \\
&= \mathrm{rad}_p^p(B).
\end{aligned}$$

This gives that $\mathrm{diam}_p(B) \geq \mathrm{rad}_p(B)$.

It remains to consider the case $p=\infty$. We recall that for any $X \in \mathcal{M}_w$ and $B \subseteq X$, we have

$$\mathrm{diam}_\infty(X) = \sup_{x,x'\in X} d_X(x,x') \quad \text{and} \quad \mathrm{diam}_\infty(B) = \sup_{x,x'\in B} d_X(x,x').$$

For any $a \in B$, we have

$$\mathrm{diam}_\infty(B) = \sup_{x,x' \in B} d_X(x,x') \leq \sup_{x,x' \in B} \big(d_X(x,a) + d_X(a,x')\big) = \sup_{x \in B} d_X(x,a) + \sup_{x' \in B} d_X(a,x').$$

Since the above inequality holds for any $a \in B$, we have that

$$\mathrm{diam}_\infty(B) \leq \inf_{a \in B} \left(\sup_{x \in B} d_X(x,a) + \sup_{x' \in B} d_X(a,x') \right) = 2\mathrm{rad}_\infty(B).$$

The other inequality holds because

$$\mathrm{rad}_\infty(B) = \inf_{a \in B} \sup_{x \in B} d_X(a,x) \leq \sup_{a \in B} \sup_{x \in B} d_X(a,x) = \mathrm{diam}_\infty(B).$$

□

We now define an analogue of $\mathrm{Shatter}_{k,p,q}(X)$ that we denote by $\mathrm{Shatter}^{\mathrm{rad}}_{k,p,q}(X)$ as follows:

Definition 3.23 ($\mathrm{Shatter}^{\mathrm{rad}}_{k,p,q}$)**.** *For $X \in \mathcal{M}_w$, $k \in \mathbf{N}$ and $p, q \in \mathbf{R}$ with $1 \leq p < \infty$ and $1 \leq q < \infty$, define*

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,q}(X) = \min_{\{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)} \left(\sum_{i=1}^k \mathrm{rad}_p^q(B_i)\, \mu_X(B_i) \right)^{1/q}.$$

For $1 \leq p \leq \infty$ and $q = \infty$, define

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,\infty}(X) = \min_{\{B_i\}_{i=1}^k \in \mathrm{Part}_k(X)} \max_i \mathrm{rad}_p(B_i).$$

A direct corollary of Theorem 3.22 is the following:

Corollary 3.24. *For every $X \in \mathcal{M}_w$, for all $p \in [1, \infty]$ and for all $k \in \mathbf{N}$, we have*

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) \leq \mathrm{Shatter}_{k,p,p}(X) \leq 2 \cdot \mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X).$$

Definition 3.25 (**k-median problem**)**.** *[41] Given a metric space (X, d), a function $w : X \to \mathbf{R}_{\geq 0}$ and $k \in \mathbf{Z}_+$, the k-median problem asks to find an $F \subseteq X$ such that $|F| = k$ and F minimizes the objective function*

$$\Phi(F) := \sum_{a \in X} d(a, F)\, w(a).$$

Here for all $a \in X$, $d(a, F) = \min_{b \in F} d(a, b)$.

Let $|X| = n$. For a non-empty $C \subseteq X$, define

$$||C||_p := \left(\sum_{x \in X} d^p(x, C) \mu_X(x) \right)^{1/p}.$$

For $1 \leq p < \infty$, define

$$\mathrm{opt}_p(X) := \min_{C \subseteq X, |C| = k} \left(\sum_{x \in X} d^p(x, C) \mu_X(x) \right)^{1/p}.$$

$$C_p(X) := \mathrm{argmin}_{C\subseteq X, |C|=k} \left(\sum_{x\in X} d^p(x, C)\mu_X(x) \right)^{1/p}.$$

For $p = \infty$, define

$$\mathrm{opt}_\infty(X) := \min_{C\subseteq X, |C|=k} \max_{x\in X} d(x, C).$$

$$C_\infty(X) := \mathrm{argmin}_{C\subseteq X, |C|=k} \max_{x\in X} d(x, C).$$

Note that there can be more than one minimizing set. In that case, we pick one arbitrarily.

Lemma 3.26. *For $1 \le p \le \infty$, any finite $X \in \mathcal{M}_w$ and any $k \in \mathbf{N}$, we have*

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) = \mathrm{opt}_p(X).$$

Proof. We first assume that $1 \le p < \infty$. For a finite metric space X, we have $\mathrm{rad}_p(X) = \inf_{a\in X} \left(\sum_{x\in X} d^p(a, x)\mu_X(x)\right)^{1/p}$. Similarly for $B \subseteq X$, we have $\mathrm{rad}_p(B) = \inf_{a\in B} \left(\sum_{x\in B} d^p(a, x)\frac{\mu_X(x)}{\mu_X(B)}\right)^{1/p}$. Let $\{B_1 \ldots, B_k\}$ be a partition of X such that

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) = \left(\sum_{i=1}^{k} \mathrm{rad}_p^p(B_i)\mu_X(B_i) \right)^{1/p}.$$

From the above definitions, we get that

$$\sum_{i=1}^{k} \mathrm{rad}_p^p(B_i)\mu_X(B_i) = \sum_{i=1}^{k} \sum_{x\in B_i} d^p(a_i, x)\frac{\mu_X(x)}{\mu_X(B_i)}\mu_X(B_i) = \sum_{i=1}^{k} \sum_{x\in B_i} d^p(a_i, x)\mu_X(x).$$

where for every $i \in [k]$, $a_i \in B_i$ is such that $\mathrm{rad}_p(B_i) = \left(\sum_{x\in B_i} d^p(a_i, x)\frac{\mu_X(x)}{\mu_X(B)}\right)^{1/p}$. This gives that

$$\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) := \left(\sum_{i=1}^{k} \sum_{x\in B_i} d^p(a_i, x)\mu_X(x) \right)^{1/p}.$$

For every $i \in [k]$, let $c_i = a_i$. Let $C = \{c_1, \ldots, c_k\}$. Then for any $i \in [k]$ and any $x \in B_i$, we have $d_X(x, C) \le d_X(x, a_i)$. This implies that

$$\sum_{x\in X} d^p(x, C)\mu_X(x) = \sum_{i=1}^{k} \sum_{x\in B_i} d^p(x, C)\mu_X(x) \le \sum_{i=1}^{k} \sum_{x\in B_i} d^p(a_i, x)\mu_X(x).$$

This gives that $\mathrm{opt}_p(X) \le \mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X)$.

We now prove that $\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) \ge \mathrm{opt}_p(X)$. Let $C_p(X) = \{c_1, \ldots, c_k\}$ and $\{B_1, \ldots, B_k\}$ be a Voronoi partition of X with respect to $C_p(X)$ such that $c_i \in B_i$. For every $i \in [k]$, let $a_i \in B_i$ be such that $\mathrm{rad}_p(B_i) = \left(\sum_{x\in B_i} d^p(a_i, x)\frac{\mu_X(x)}{\mu_X(B)}\right)^{1/p}$. Then we have that

$$\sum_{x\in X} d^p(x, C_p(X))\mu_X(x) = \sum_{i=1}^{k} \sum_{x\in B_i} d^p(x, c_i)\mu_X(x) \ge \sum_{i=1}^{k} \sum_{x\in B_i} d^p(x, a_i)\mu_X(x) \ge \left(\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X)\right)^p.$$

Thus we have established that $\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X) = \mathrm{opt}_p(X)$. Therefore a c-approximation algorithm for $\mathrm{opt}_p(X)$ gives a c-approximation for $\mathrm{Shatter}^{\mathrm{rad}}_{k,p,p}(X)$. The same result holds for $p = \infty$. $\square$

3.7 Proof of Theorem 3.20

3.7.1 The case $p = 1$

For $p = 1$, Charikar and Guha [11] give a polynomial time 4-approximation algorithm for $\text{opt}_1(X)$. For the case when X has uniform measure, the best known approximation factor for $\text{opt}_1(X)$ is $(3 + \epsilon)$ for any $\epsilon > 0$. This approximation factor is due to Arya et al. [9] and is obtained from a local search algorithm. We describe this local search method in the next paragraph.

Let K be a set of centers that gives an α-approximation for $\text{opt}_1(X)$. Since K gives an α-approximation, we have $||K||_1 \leq \alpha ||C_1(X)||_1$. We start with this set K and perform a local search as follows: we choose an arbitrary $\epsilon \in (0, 1)$ and take $\tau = \frac{\epsilon}{10k}$, $k = |K|$. We fix these values of ϵ and τ. We now choose an arbitrary $p \in X \setminus K$ and an arbitrary $r \in K$. We consider the new set of centers obtained by swapping r by p. In particular we define $K_{new} = K \cup \{p\} \setminus \{r\}$. Now if $||K_{new}||_1 \leq (1 - \tau)||K||_1$, then we set $K = K_{new}$; otherwise K remains the same. We then consider another swap and repeat the same procedure. There are $O(nk)$ possible swaps since $|X \setminus K| = n - k$ and $|K| = k$. We consider all such swaps. We stop when we obtain a set of centers that can no longer be improved by any swap. Let L denote this locally optimal set of centers.

We now analyze the running time of this algorithm. There are $O(nk)$ possible swaps for a fixed set of centers K. For every swap computation of $||K_{new}||_1$ takes $O(nk)$ time, since we have to calculate the distance of every $p \in X \setminus K_{new}$ to its closest point in K_{new}. Since $\frac{1}{1-\tau} \geq 1 + \tau$, we get that the running time of the local search algorithm is

$$O\left((nk)^2 \log_{1/(1-\tau)} \frac{||K||_1}{||C_1(X)||_1}\right) = O\left((nk)^2 \log_{1+\tau} \alpha\right) = O\left((nk)^2 \frac{\log \alpha}{\tau}\right).$$

In the above algorithm, we replace one center from a locally optimal solution K with an arbitrary center from $X \setminus K$. Another method of performing local search is to simultaneously replace an arbitrary set of t centers from K with an arbitrary set of t centers from $X \setminus K$. This method is called the *t-swap method*. Since the algorithm involves checking all subsets of both X and K of size t, its running time is $n^{O(t)}\epsilon^{-1}$. Note that the running time is polynomial for constant values of t and for polynomially small ϵ.

The t-swap local search algorithm has been analyzed by Arya et al. in [11] and also by Gupta and Tangwongsan in [21]. Both papers give the same approximation factor of $\left(3 + \frac{2}{t} + \epsilon\right)$ but for the case when measure on X is uniform. Here, we consider the simpler analysis of Gupta and Tangwongsan and observe that their analysis works for any arbitrary measure on X. When measure on X is non-uniform, we modify their objective function so as to include the measures.

Thus we get that for any $0 < \epsilon < 1$ and $k > t > 1$, there is a $\left(3 + \frac{2}{t} + \epsilon\right)$-approximation algorithm for $\text{opt}_1(X)$ with running time $|X|^{O(t)}\epsilon^{-1}$. From Lemma 3.26, we get that this gives a $\left(3 + \frac{2}{t} + \epsilon\right)$-approximation for $\text{Shatter}^{\text{rad}}_{k,1,1}(X)$. Using Corollary 3.24, we get a $\left(6 + \frac{4}{t} + \epsilon\right)$-approximation of $\text{Shatter}_{k,1,1}(X)$. Let P be a partition of X that achieves this $\left(6 + \frac{4}{t} + \epsilon\right)$-approximation of $\text{Shatter}_{k,1,1}(X)$. We apply the Wasserstein map (Definition 3.11) to P to obtain a k-point metric measure space. From Remark 3.1, we get that the k-point metric space obtained is a $\left(12 + \frac{8}{t} + \epsilon\right)$-approximation for $\text{Sketch}^S_{k,1}(X)$, We note that computing the k-point metric measure space takes polynomial time, since the computation involves calculating 1-Wasserstein distance between blocks of partition P, and this can be done in polynomial time.

3.7.2 The case $2 \leq p < \infty$

Definition 3.27 (**λ-approximate metric**)**.** *Given $\lambda \geq 1$, a non-negative and symmetric distance function d defined on a set M is called a λ-approximate metric if for any sequence of points $\langle x_0, \ldots, x_m \rangle$ in M, we have $d(x_0, x_m) \leq \lambda \cdot \sum_{0 \leq i \leq m} d(x_i, x_{i+1})$.*

Definition 3.28 (**Weakly λ-approximate metric**)**.** *Given $\lambda \geq 1$, a non-negative and symmetric distance function d defined on a set M is called a weakly λ-approximate metric if it satisfies the following inequality: for any $x, y, z \in M$, we have $d(x,z) \leq \lambda(d(x,y) + d(y,z))$. The above inequality is a weaker form of the triangle inequality.*

For example, if we consider the distance function d^p on our metric space (X, d), where $1 < p < \infty$, then we can show that for any $x, y, z \in X$,

$$d^p(x,y) \leq (d(x,z) + d(z,y))^p \leq 2^{p-1}(d^p(x,z) + d^p(z,y)).$$

Thus d^p is a weakly 2^{p-1}- approximate metric.

We refer to the work of Mettu and Plaxton [41] in this paragraph. They give a constant factor approximation algorithm for the k-median problem on spaces endowed with a weakly λ-approximate metric. Let the input metric space be U and n be the cardinality of U. Let l be the ratio of the diameter of U to the shortest distance between any pair of distinct points in U. We note that the length of any sequence σ_i as described in section 3.1 of [41] is at most $O(\log l)$. From Lemma 4.1 of [41], we have that if U is a weakly λ-approximate metric then for points $x_0, x_1, \ldots, x_m$ in U with $m \geq 1$, $d(x_0, x_m) \leq \lambda^{\lceil \log_2 m \rceil} \sum_{0 \leq i \leq m} d(x_i, x_{i+1})$. Since all sequences used in proving Theorem 3.1 of [41] have length $O(\log l)$, we get an approximation factor of $\lambda^{O(\log_2 \log l)}$ for the k-median problem on a weakly λ-approximate metric. Since d^p is a weakly 2^{p-1}-approximate metric, we get a $(\log l)^{O(p-1)}$-approximation for $\mathrm{opt}_p(X)$.

Let K denote the set of centers obtained from this approximation algorithm. We now perform local search on K using the t-swap method for $t \geq 1$, as done in the previous section for $p = 1$. This method has been analyzed for the metric d^p by Gupta and Tangwongsan in [21]. The problem studied by Gupta and Tangwongsan in [21] is the l^p-facility location problem. This problem can be stated as follows: Given a set P of m points, find a set $F \subseteq P$ of k points such that the quantity $\Phi_p(F) = \left(\sum_{j \in P} d^p(j, F)\right)^{1/p}$ is minimized. For this problem, they give the following result:

Theorem 3.29. [[21], Theorem 3.5] *For any value of $t \in \mathbf{Z}_+$, the natural t-swap local-search algorithm for the l_p-facility location problem yields the following guarantees:*

1. *For $p = 2$, it is a $\left(5 + \frac{4}{t}\right)$ approximation.*
2. *For all reals $p \geq 2$, it is a $\left(3 + \frac{2}{t}\right) p$ approximation.*

It is straightforward to see that the objective function $\Phi_p(F)$ is same as $\mathrm{opt}_p(X)$ when X has uniform measure. However the analysis of Theorems 3.2 and 3.5 of [21] also works when X has non-uniform measure. Therefore we can perform the same analysis on the locally optimal solution obtained after performing local search on K and obtain the following result:

Lemma 3.30. *For any $t \in \mathbf{Z}_+$ and any $0 < \epsilon < 1$, there is an $f(p)$-approximation for $\mathrm{opt}_p(X)$ where $f(p)$ is as follows:*

1. For $p = 2$, $f(p) = 5 + \frac{4}{t} + \epsilon$.

2. For all reals $p > 2$, $f(p) = \left(3 + \frac{2}{t}\right) p + \epsilon$.

The running time of the approximation algorithm is $|X|^{O(t)}\epsilon^{-1}$.

From Lemma 3.26 and Lemma 3.30, we get an $f(p)$-approximation for $\text{Shatter}^{\text{rad}}_{k,p,p}(X)$. Using Corollary 3.24, we get a $2f(p)$-approximation for $\text{Shatter}_{k,p,p}(X)$. Let P be a partition of X that achieves this $2f(p)$-approximation of $\text{Shatter}_{k,p,p}(X)$. We apply the Wasserstein map (Definition 3.11) to P to obtain a k-point metric measure space. From Remark 3.1, we get that the k-point metric space obtained is a $4f(p)$-approximation of $\text{Sketch}^S_{k,p}(X)$, We note that computing the k-point metric measure space takes polynomial time, since the computation involves calculating p-Wasserstein distance between blocks of partition P, and this can be done in polynomial time.

4 Relating the weak and strong sketching objectives for mm-spaces

Without imposing some control on the class of mm-spaces we consider, there is no hope to obtain a comparability result between $\text{Sketch}_{k,p}$ and $\text{Sketch}^S_{k,p}$. This is demonstrated by the following example.

Example 4.1 (Blow-up of $\frac{\text{Sketch}^S_{k,1}(\Delta_m)}{\text{Sketch}_{k,1}(\Delta_m)}$)**.** *We know from the proof of Theorem 3.1 that for all $k, m \in \mathbf{N}$ with $k \le m$ and $p \ge 1$, we have*

$$\text{Sketch}^S_{k,p}(\Delta_m) \ge \frac{1}{2}\left(1 - \frac{k}{m}\right)^{1/p}.$$

We know from Claim 5.1 of [39] that $d_{GW_p}(\Delta_m, \Delta_k) \le \frac{1}{2}\left(\frac{1}{k} + \frac{1}{m}\right)^{1/p}$. This implies that $\text{Sketch}_{k,p}(\Delta_m) \le \frac{1}{2}\left(\frac{1}{k} + \frac{1}{m}\right)^{1/p}$. Therefore, we get that

$$\frac{\text{Sketch}^S_{k,1}(\Delta_m)}{\text{Sketch}_{k,1}(\Delta_m)} \ge \frac{1 - \frac{k}{m}}{\frac{1}{k} + \frac{1}{m}}.$$

For $m = 2^{n+1}$ and $k = 2^n$, we get that

$$\frac{\text{Sketch}^S_{k,1}(\Delta_m)}{\text{Sketch}_{k,1}(\Delta_m)} \ge \frac{m}{6}. \tag{2}$$

We observe that as $m \to \infty$, the right hand side of the above inequality goes to ∞. In what follows, we show that for a rich family of metric measure spaces, we can upper bound $\text{Sketch}^S_{k,p}(X)$ by a suitable function of $\text{Sketch}_{k,p}(X)$.

Definition 4.1 (Doubling mm-spaces). [Chapter 1, [26]] *A metric measure space (X, d_X, μ_X) is called doubling if there exists a constant $C \ge 1$ such that for all $x \in X$ and $r > 0$, we have*

$$\mu_X(B_X(x, 2r)) \le C \cdot \mu_X(B_X(x, r)).$$

The constant C is called the doubling constant of X.

Note that for any $m \in \mathbf{N}$ such that $m \geq 2$, the space $\Delta_m \in \mathcal{M}_w$ has doubling constant m, i.e. it is not bounded independently of m.

We invoke some results from Section 5 of [39]. Let $(X, d_X, \mu_X) \in \mathcal{M}_w$. Given $\delta > 0$, define $f_\delta(\epsilon) = \mu_X(x \in X \mid \mu_X(B_X(x,\epsilon)) \leq \delta)$. We now define $v_\delta(X)$ as follows: $v_\delta(X) = \inf\{\epsilon > 0 \mid f_\delta(\epsilon) \leq \epsilon\}$. Note that $v_\delta(X)$ is an increasing function of δ i.e. if $\delta_1 \leq \delta_2$ then $v_{\delta_1}(X) \leq v_{\delta_2}(X)$.

We have the following result:

Theorem 4.2 ([39])**.** *Let $X, Y \in \mathcal{M}_w$, $p \in [1,\infty)$ and $\delta \in (0, 1/2)$. Then*

$$\zeta_p(X,Y) \leq (4 \cdot \min(v_\delta(X), v_\delta(Y)) + \delta)^{1/p} \cdot M,$$

whenever $d_{GWp}(X,Y) < \delta^5$, where $M = 2 \cdot \max(\mathrm{diam}(X), \mathrm{diam}(Y)) + 45$.

Let $\mathcal{F} \subset \mathcal{M}_w$ be a family for which there exists a surjective function $\rho_\mathcal{F} : [0,\infty) \to [0, \delta_\mathcal{F}]$ with $\delta_\mathcal{F} > 0$ such that for all $\epsilon \geq 0, x \in X$ and $X \in \mathcal{F}$, we have $\mu_X(B_X(x,\epsilon)) \geq \rho_\mathcal{F}(\epsilon)$. Then for all $\delta \in (0, \delta_\mathcal{F})$, we have $\sup_{X \in \mathcal{F}} v_\delta(X) \leq \inf\{\epsilon > 0 \mid \rho_\mathcal{F}(\epsilon) > \delta\}$. [1]

If $\mathcal{F} \subset \mathcal{M}_w$ is the set of all metric measure spaces with doubling dimension C, then we can take $\rho_\mathcal{F}(\epsilon) = \left(\frac{\epsilon}{2D}\right)^N$, where $D = \mathrm{diam}_\infty(X)$ and $N = \log_2 C$. We observe that if $\rho_\mathcal{F}$ is an increasing function then given $\delta > 0$, the quantity $\inf\{\epsilon > 0 \mid \rho_\mathcal{F}(\epsilon) > \delta\}$ is equal to $\rho_\mathcal{F}^{-1}(\delta)$. Since $\rho_\mathcal{F}(\epsilon) = \left(\frac{\epsilon}{2D}\right)^N$ is an increasing function, we get that $\inf\{\epsilon > 0 \mid \rho_\mathcal{F}(\epsilon) > \delta\} = \rho_\mathcal{F}^{-1}(\delta) = 2D\delta^{1/N}$.

Therefore we get that for all doubling metric measure spaces X with doubling constant $C > 0$, we have $v_\delta(X) \leq 2 \cdot \mathrm{diam}(X) \cdot \delta^{1/\log_2 C}$.

Proof of Theorem 1.4. Let $\mathrm{Sketch}_{k,p}(X) < \delta'^5$, where $\delta' \in \left(0, \frac{1}{2}\right)$. Then, there exists $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ such that $d_{GW_p}(X, M_k) < \delta'^5$. We may assume that $\mathrm{diam}(M_k) \leq \mathrm{diam}(X)$. This is because of the following claim.

Claim 4.3. *Given $X \in \mathcal{M}_w$, let $(M_k, d_k, \mu_k) \in \mathcal{M}_{k,w}$ be such that $\mathrm{diam}(M_k) > \mathrm{diam}(X)$. Then, there exists $(M'_k, d'_k, \mu'_k) \in \mathcal{M}_{k,w}$ such that $d_{GW_p}(X, M'_k) \leq d_{GW_p}(X, M_k)$ and $\mathrm{diam}(M_k) \leq \mathrm{diam}(X)$.*

Proof of Claim. We consider the k-point space (M'_k, d'_k, μ'_k) defined as follows: for every $i \in [k]$, $\mu'_k(i) = \mu_k(i)$ and

$$d'_k(i,j) = \min(d_k(i,j), \mathrm{diam}(X)) \quad \forall\ i,j \in [k].$$

Define $S \subseteq [k] \times [k]$ as the set such that for all $(i,j) \in S$, $d_k(i,j) \neq d'_k(i,j)$. This means that for all $(i,j) \in S$, $d_k(i,j) > \mathrm{diam}(X)$ and therefore $d'_k(i,j) = \mathrm{diam}(X)$. Let $\overline{S} = [k] \times [k] \setminus S$. Then for any measure coupling μ between μ_X and μ_k, we have

$$\begin{aligned}
&\iint\limits_{(X \times M'_k) \times (X \times M'_k)} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i) d\mu(x',j) \\
&\qquad = \iint\limits_{(X \times S) \times (X \times S)} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i) d\mu(x',j) \\
&\qquad + \iint\limits_{(X \times \overline{S}) \times (X \times \overline{S})} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i) d\mu(x',j).
\end{aligned}$$

[1]This is because for a fixed $\delta \in (0, \delta_\mathcal{F})$, if $\epsilon \geq 0$ is such that $\rho_\mathcal{F}(\epsilon) > \delta$ then for all $x \in X$, $X \in \mathcal{F}$ we have $\mu_X(B_X(x,\epsilon)) > \delta$. This gives $f_\delta(\epsilon) = 0$ and $v_\delta(X) \leq \epsilon$. Since $\epsilon > 0$ was arbitrary, we get the above inequality.

Note that

$$\iint\limits_{(X\times\overline{S})\times(X\times\overline{S})} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i)d\mu(x',j)$$
$$= \iint\limits_{(X\times\overline{S})\times(X\times\overline{S})} |d_X(x,x') - d_k(i,j)|^p d\mu(x,i)d\mu(x',j).$$

For the other integral we have

$$\iint\limits_{(X\times S)\times(X\times S)} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i)d\mu(x',j)$$
$$= \iint\limits_{(X\times S)\times(X\times S)} |d_X(x,x') - \mathrm{diam}(X)|^p d\mu(x,i)d\mu(x',j)$$
$$= \iint\limits_{(X\times S)\times(X\times S)} \big(\mathrm{diam}(X) - d_X(x,x')\big)^p\, d\mu(x,i)d\mu(x',j)$$
$$\leq \iint\limits_{(X\times S)\times(X\times S)} \big(d_k(i,j) - d_X(x,x')\big)^p\, d\mu(x,i)d\mu(x',j)$$
$$= \iint\limits_{(X\times S)\times(X\times S)} |d_k(i,j) - d_X(x,x')|^p d\mu(x,i)d\mu(x',j).$$

Therefore we get that

$$\iint\limits_{(X\times M'_k)\times(X\times M'_k)} |d_X(x,x') - d'_k(i,j)|^p d\mu(x,i)d\mu(x',j)$$
$$\leq \iint\limits_{(X\times S)\times(X\times S)} |d_k(i,j) - d(x,x')|^p d\mu(x,i)d\mu(x',j)$$
$$+ \iint\limits_{(X\times\overline{S})\times(X\times\overline{S})} |d_X(x,x') - d_k(i,j)|^p d\mu(x,i)d\mu(x',j)$$
$$= \iint\limits_{(X\times M_k)\times(X\times M_k)} |d_X(x,x') - d_k(i,j)|^p d\mu(x,i)d\mu(x',j).$$

This gives that $d_{GW_p}(X, M'_k) \leq d_{GW_p}(X, M_k)$ and by definition of (M'_k, d'_k, μ'_k), we have that $\mathrm{diam}(M'_k) \leq \mathrm{diam}(X)$. □

Now, we have that $d_{GW_p}(X, M_k) < \delta'^5$ and $\mathrm{diam}(M_k) \leq \mathrm{diam}(X)$. We now use Theorem 4.2 to get

$$\zeta_p(X, M_k) \leq (4 \cdot \min(v_{\delta'}(X), v_{\delta'}(M_k)) + \delta')^{1/p} \cdot M$$

where $M = 2 \cdot \max(\mathrm{diam}(X), \mathrm{diam}(M_k)) + 45$. Since X is a doubling metric measure space with doubling constant $C > 0$, we have that $v_{\delta'}(X) \leq 2 \cdot \mathrm{diam}(X)\delta'^{1/\log_2 C}$. This gives

$$\min(v_{\delta'}(X), v_{\delta'}(M_k)) \leq v_{\delta'}(X) \leq 2 \cdot \mathrm{diam}(X)\delta'^{1/\log_2 C}.$$

Since $\mathrm{diam}(M_k) \leq \mathrm{diam}(X)$, we get $M = 2 \cdot \mathrm{diam}(X) + 45$. Therefore we have the following result:

$$\mathrm{Sketch}^S_{k,p}(X) \leq \zeta_p(X, M_k) \leq \left(8 \cdot \mathrm{diam}(X)\, \delta'^{1/\log_2 C} + \delta'\right)^{1/p} \cdot M$$

where $M = 2 \cdot \mathrm{diam}(X) + 45$. Let $\mathrm{Sketch}_{k,p}(X) = \delta$. Then, for every $\epsilon > 0$, the above inequality holds true for $\delta' = \delta^{1/5}(1+\epsilon)$. Moreover, we know from [39] that for every $X \in \mathcal{M}_w$ and $M_k \in \mathcal{M}_{k,w}$, $\zeta_p(X, M_k) \geq d_{GW_p}(X, M_k)$. Therefore, we get that

$$\delta \leq \mathrm{Sketch}^S_{k,p}(X) < \left(8 \cdot \mathrm{diam}(X) \cdot \delta^{1/(5\log_2 C)} + \delta^{1/5}\right)^{1/p} \cdot M$$

whenever $\mathrm{Sketch}_{k,p}(X) = \delta < 2^{-5}$, where $M = 2 \cdot \mathrm{diam}(X) + 45$. □

5 Impossibility results for sketching of metric spaces

In order to prove that there exist clustering functions that do not admit a dual sketching function, we need to impose some conditions on our clustering and sketching cost functions. Otherwise, for any clustering cost function we can take our sketching cost function to be equal to the clustering cost function. However, such a freedom in the choice of sketching cost function is, of course, unreasonable.

We need the following definition: given $X \in \mathcal{M}$, the *k-covering radius* of X is the smallest $r > 0$ such that there exist k balls of radius r centered at points of X that cover the whole space X. We use $\mathrm{Cov}_k(X)$ to denote the k-covering radius of X.

We say that a clustering cost function Φ is *admissible* if the following properties are satisfied for all $X \in \mathcal{M}$ and $k \in \mathbf{N}$:

1. $\mathrm{Shatter}^\Phi_k(\lambda X) = \lambda\, \mathrm{Shatter}^\Phi_k(X)$ for all $\lambda \geq 0$.
2. $\mathrm{Shatter}^\Phi_k(X) \leq \mathrm{Shatter}^\Phi_{k-1}(X)$.

In the rest of the section, we refer to an admissible clustering cost function as admissible clustering.

We say that a sketching cost function Ψ is *admissible* if the following properties are satisfied for all $X \in \mathcal{M}$ and $k \in \mathbf{N}$:

1. $\mathrm{Sketch}^\Psi_k(\lambda X) = \lambda\, \mathrm{Sketch}^\Psi_k(X)$ for all $\lambda \geq 0$.
2. $\mathrm{Sketch}^\Psi_k(X) \leq \mathrm{Sketch}^\Psi_{k-1}(X)$.
3. $\mathrm{Sketch}^\Psi_k(X) \geq \alpha_k \cdot \mathrm{Cov}_k(X)$, where $\alpha_k > 0$ is a constant.

In the rest of the section, we refer to an admissible sketching cost function as admissible sketching.

We now give an example of an admissible clustering for which a dual admissible sketching does not exist. Given $(X, d_X) \in \mathcal{M}$, we consider u_X, the maximal sub-dominant ultrametric on X. This is defined precisely as follows. First, given $x, x' \in X$, let $S_{x,x'}$ denote the collection of all sequences of points in X starting at x and ending at x'. Then, define

$$u_X(x, x') := \min\left\{\, \max_i d_X(x_i, x_{i+1}) \mid (x_0, x_1, \ldots, x_n) \in S_{x,x'} \right\}.$$

Let $U(X) = (X, u_X)$. For every $k \leq |X|$, we consider the clustering cost function:

$$\Phi_U(X, P) = \Phi(U(X), P).$$

It is starightforward to see that Φ_U is admissible. Let $D(X)$ denote the dendrogram describing the single linkage hierarchical clustering (SLHC) of (X, d_X) [31]. We define $t_k(X) = \inf\{t \geq 0 \mid D(X)$ has at most k bars$\}$. We observe that

$$\text{Shatter}_k^{\Phi_U}(X) = \inf_{P \in \text{Part}_k(X)} \Phi_U(X, P) = t_k(X),$$

and this corresponds to finding a partition of X into k blocks that maximizes the minimum inter-cluster distance.

Theorem 5.1 (Φ_U does not admit a dual admissible sketching)**.** *For every $k \in \mathbf{N}$, there exists a sequence of spaces $\{Y_{n,k}\}_{n\in\mathbf{N}} \in \mathcal{M}$ such that for any admissible sketching Ψ, we have $\text{Sketch}_k^{\Psi}(Y_{n,k}) \geq \Theta(1)$ but as $n \to \infty$, $\text{Shatter}_k^{\Phi_U}(Y_{n,k}) \to 0$.*

Proof. For $k = 1$, consider the space $Y_{n,1} = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\} \subset \mathbf{R}$. For an arbitrary $k \in \mathbf{N}$, we define $Y_{n,k} = \cup_{i=0}^{k-1}(2k + Y_{n,1})$, where $\alpha + Y_{n,1} = \left\{\alpha, \alpha + \frac{1}{n}, \alpha + \frac{2}{n}, \ldots, \alpha + \frac{n-1}{n}, \alpha + 1\right\}$. In particular for every $k \in \mathbf{N}$, we get a sequence of spaces $\{Y_{n,k}\}_{n\in\mathbf{N}} \in \mathbf{R}$. We observe that for any $n, k \in \mathbf{N}$, $\text{Shatter}_k^{\Phi_U}(Y_{n,k}) = \frac{1}{n}$. In addition, for all $n, k \in \mathbf{N}$ we have $\text{Cov}_k(Y_{n,k}) \geq \frac{1}{2}$. Therefore for any admissible clustering Ψ, we have $\text{Sketch}_k^{\Psi}(Y_{n,k}) \geq \frac{\alpha_k}{2} > 0$. However as $n \to \infty$, $\text{Shatter}_k^{\Phi_U}(Y_{n,k}) = \frac{1}{n} \to 0$. Thus we get that for every $k \in \mathbf{N}$, $\{Y_{n,k}\}_{n\in\mathbf{N}} \in \mathcal{M}$ is such that as $n \to \infty$, $\text{Shatter}_k^{\Phi_U}(Y_{n,k}) \to 0$ but any admissible sketching Ψ satisfies $\text{Sketch}_k^{\Psi}(Y_{n,k}) \geq \Theta(1)$. This implies that the admissible clustering Φ_U does not admit a dual admissible sketching. □

We now assume that $X \subset \mathbf{R}^d$. We consider the Hamming distance r_X on X that is defined as follows: for $x = (x_1, \ldots, x_d)$ and $x' = (x'_1, \ldots, x'_d)$, $r_X(x, x') = |\{i \in [d] : x_i \neq x'_i\}|$. It follows from the definition that r_X gives a metric on X. For any $X \subset \mathbf{R}^d$, let $R(X) = (X, r_X)$. For every $k \in \mathbf{N}$ and partition P of X into k blocks, we consider the clustering cost function

$$\Phi_R(X, P) = \Phi(R(X), P).$$

It is straightforward to see that $\Phi_R(X, P)$ is admissible.

Theorem 5.2. *The admissible clustering Φ_R does not admit a dual admissible sketching.*

Proof. Suppose Ψ is an admissible compression that is dual to Φ_R. Let $X \in \mathcal{M}$ be such that $X \subset \mathbf{R}$ and $|X| \geq 2$. Let $k = 1$. Since $|X| \geq 2$ and $\text{Part}_1(X) = X$, we have $\text{Shatter}_1^{\Phi_R}(X) = 1$. We observe that $\text{Cov}_1(X) \geq \frac{\text{diam}(X)}{2}$. This gives that $\text{Sketch}_1^{\Psi}(X) \geq \alpha_1 \cdot \text{diam}(X)$, where $c > 0$ is a constant. Since Ψ is dual to Φ_R, there exist constants $C_2 \geq C_1 > 0$ such that for all $X \in \mathcal{M}$,

$$C_1 \leq \text{Sketch}_1^{\Psi}(X) \leq C_2.$$

This implies that $\alpha_1 \cdot \text{diam}(X) \leq C_2$. We now consider $X' \in \mathcal{M}$ such that $X' \subset \mathbf{R}$ and $\text{diam}(X') > \frac{C_2}{\alpha_1}$. Such an X' exists because $\frac{C_2}{\alpha_1} > 0$ is a finite constant. Now for X', we again have $\text{Shatter}_1^{\Phi_R}(X') = 1$ but the inequality $\alpha_1 \cdot \text{diam}(X') \leq C_2$ does not hold. In particular, the above inequality does not hold with same constants $C_2 \geq C_1 > 0$ for all $X \in \mathcal{M}$ because we can choose an X with diameter large enough. Thus we get a contradiction to our assumption that Ψ is dual to Φ_R. We can similarly construct counter-examples for higher values of k. □

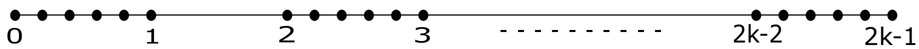

Figure 3: The set $Y_{n,k}$. Points within an interval of length 1 are at distance $\frac{1}{n}$ from each other.

Consider another clustering cost function $\overline{\Phi}$ defined as follows: for a partition $P = \{B_i\}_{i=1}^k$ of X, we define

$$\overline{\Phi}(X, P) = \max_i(\mathrm{diam}(X) - \mathrm{diam}(B_i)).$$

Theorem 5.3. *The admissible clustering $\overline{\Phi}$ does not admit a dual admissible sketching.*

Proof. Let $X \in \mathcal{M}$ be arbitrary and let $k = 1$. Then we have $\mathrm{Part}_1(X) = \{X\}$ and this gives that $\mathrm{Shatter}_1^{\overline{\Phi}}(X) = \overline{\Phi}(X, \{X\}) = 0$. However for any admissible clustering Ψ, we have $\mathrm{Sketch}_1^{\Psi}(X) \geq \alpha_1 \cdot \mathrm{diam}(X) > 0$ since $\mathrm{Cov}_1(X) \geq \frac{\mathrm{diam}(X)}{2}$. Here $\alpha_1 > 0$ is a constant. This shows that the inequality $\mathrm{Sketch}_1^{\Psi}(X) \leq C_2 \cdot \mathrm{Shatter}_1^{\overline{\Phi}}(X)$ does not hold for any constant $C_2 > 0$. Since $X \in \mathcal{M}$ was arbitrary, we conclude that the admissible clustering $\overline{\Phi}$ does not admit a dual admissible sketching. □

Given $1 < p < \infty$, we consider a metric transform $M_p : \mathcal{M} \to \mathcal{M}$ given by $M_p((X, d_X)) = (X, m_p)$. For any $x, x' \in X$, let $S_{x,x'}$ denote the collection of all sequences of points in X starting at x and ending at x'. We define

$$m_p(x, x') := \min_{(x_0,\ldots,x_m)\in S_{x,x'}} \left(\sum_{i=0}^{m-1} d_X^p(x_i, x_{i+1}) \right)^{1/p}.$$

For any $k \in \mathbf{N}$ and $P \in \mathrm{Part}_k(X)$, define $\Phi_p(X, P) = \Phi(M_p(X), P)$. It is straightforward to see that $\Phi_p(X, P)$ is admissible.

Theorem 5.4. *For any $1 < p < \infty$, the admissible clustering Φ_p does not admit a dual admissible sketching.*

Proof. We use a similar construction as in the proof of Theorem 5.1. Let $p \in (1, \infty)$ be fixed. For $k = 1$, consider the space $Y_{n,1} = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\} \subset \mathbf{R}$. For an arbitrary $k \in \mathbf{N}$, we define $Y_{n,k} = \cup_{i=0}^{k-1}\{2k + Y_{n,1}\}$, where for $\alpha \in \mathbf{R}$, $\alpha + Y_{n,1} = \left\{\alpha, \alpha + \frac{1}{n}, \alpha + \frac{2}{n}, \ldots, \alpha + \frac{n-1}{n}, \alpha + 1\right\}$. In particular for every $k \in \mathbf{N}$, we get a sequence of spaces $\{Y_{n,k}\}_{n\in\mathbf{N}} \in \mathbf{R}$. We observe that for all $n, k \in \mathbf{N}$, $\mathrm{diam}(Y_{n,k}) = 2k - 1$. For any $n \in \mathbf{N}$, we have

$$\mathrm{Shatter}_1^{\Phi_p}(Y_{n,1}) = \mathrm{diam}(M_p(Y_{n,1})) = \left(\frac{n}{n^p}\right)^{1/p} = \frac{1}{n^{1-1/p}}.$$

Thus we get that as $n \to \infty$, $\text{Shatter}_1^{\Phi_p}(Y_{n,1}) \to 0$. However $\text{Cov}_1(Y_{n,1}) \geq \frac{\text{diam}(Y_{n,1})}{2} = \frac{1}{2}$. This gives that for any admissible sketching Ψ, $\text{Sketch}_1^{\Psi}(Y_{n,1}) \geq \frac{\alpha_1}{2} > 0$ for all $n \in N$ and some constant $\alpha_1 > 0$. Thus the inequality $\text{Sketch}_1^{\Psi}(Y_{n,1}) \leq C_2 \cdot \text{Shatter}_1^{\Phi_p}(Y_{n,1})$ does not hold for any constant $C_2 > 0$.

For all $k > 1$, we consider $Y_{n,k}$. We again have $\text{Cov}_k(Y_{n,k}) \geq \frac{1}{2}$. This gives that for any admissible sketching Ψ, $\text{Sketch}_k^{\Psi}(Y_{n,k}) \geq \frac{\alpha_k}{2} > 0$ for all $n \in N$ and some constant $\alpha_k > 0$. We have $\text{Shatter}_k^{\Phi_p}(Y_{n,k}) = \frac{1}{n^{1-1/p}}$ since $Y_{n,k}$ consists of k blocks of $Y_{n,1}$ arranged in a path with the distance between adjacent blocks being 1. Thus we can achieve an equipartition of $Y_{n,k}$ into k blocks. We again have that $\text{Shatter}_k^{\Phi_p}(Y_{n,k}) \to 0$ as $n \to \infty$. Thus the inequality $\text{Sketch}_k^{\Psi}(Y_{n,k}) \leq C_2 \cdot \text{Shatter}_k^{\Phi_p}(Y_{n,k})$ does not hold for any constant $C_2 > 0$. We conclude that the admissible sketching Φ_p does not admit a dual admissible compression. Since $p \in (1, \infty)$ was arbitrary, the statement of the theorem holds. □

References

[1] Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat. Np-hardness of euclidean sum-of-squares clustering. *Machine learning*, 75(2):245–248, 2009.

[2] Alexandr Andoni, Jiecao Chen, Robert Krauthgamer, Bo Qin, David P Woodruff, and Qin Zhang. On sketching quadratic forms. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 311–319. ACM, 2016.

[3] Alexandr Andoni, Khanh Do Ba, Piotr Indyk, and David Woodruff. Efficient sketches for earth-mover distance, with applications. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 324–330. IEEE, 2009.

[4] Alexandr Andoni, Assaf Goldberger, Andrew McGregor, and Ely Porat. Homomorphic fingerprints under misalignments: sketching edit and shift distances. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 931–940. ACM, 2013.

[5] Alexandr Andoni, Robert Krauthgamer, and Ilya Razenshteyn. Sketching and embedding are equivalent for norms. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 479–488. ACM, 2015.

[6] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Impossibility of sketching of the 3d transportation metric with quadratic cost. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 55. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[7] David Arthur and Sergei Vassilvitskii. How slow is the k-means method? In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 144–153. ACM, 2006.

[8] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[9] Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on computing*, 33(3):544–562, 2004.

[10] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. URL: https://books.google.com/books?id=afnlx8sHmQIC.

[11] Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for facility location problems. *SIAM Journal on Computing*, 34(4):803–824, 2005.

[12] Samir Chowdhury and Facundo Mémoli. Constructing geodesics on the space of compact metric spaces. *arXiv preprint arXiv:1603.02385*, 2016.

[13] Sanjoy Dasgupta. *The hardness of k-means clustering*. Department of Computer Science and Engineering, University of California, San Diego, 2008.

[14] Inderjit S Dhillon and Dharmendra S Modha. Concept decompositions for large sparse text data using clustering. *Machine learning*, 42(1-2):143–175, 2001.

[15] Petros Drineas, Alan Frieze, Ravi Kannan, Santosh Vempala, and V Vinay. Clustering large graphs via the singular value decomposition. *Machine learning*, 56(1-3):9–33, 2004.

[16] Khaled M. Elbassioni and Kazuhisa Makino, editors. *Algorithms and Computation - 26th International Symposium, ISAAC 2015, Nagoya, Japan, December 9-11, 2015, Proceedings*, volume 9472 of *Lecture Notes in Computer Science*. Springer, 2015.

[17] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[18] Edward W Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769, 1965.

[19] Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38(Supplement C):293 – 306, 1985. URL: http://www.sciencedirect.com/science/article/pii/0304397585902245, doi:https://doi.org/10.1016/0304-3975(85)90224-5.

[20] Sudipto Guha, Piotr Indyk, and Andrew McGregor. Sketching information divergences. *Machine Learning*, 72(1-2):5–19, 2008. URL: https://doi.org/10.1007/s10994-008-5054-x, doi:10.1007/s10994-008-5054-x.

[21] Anupam Gupta and Kanat Tangwongsan. Simpler analyses of local search algorithms for facility location. *CoRR*, abs/0809.2554, 2008. URL: http://arxiv.org/abs/0809.2554.

[22] Sariel Har-Peled. *Geometric approximation algorithms*, volume 173. American mathematical society Boston, 2011.

[23] Sariel Har-Peled and Akash Kushal. Smaller coresets for k-median and k-means clustering. *Discrete & Computational Geometry*, 37(1):3–19, 2007. URL: http://dx.doi.org/10.1007/s00454-006-1271-x, doi:10.1007/s00454-006-1271-x.

[24] Sariel Har-Peled and Soham Mazumdar. On coresets for k-means and k-median clustering. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 291–300. ACM, 2004.

[25] Sariel Har-Peled and Bardia Sadri. How fast is the k-means method? *Algorithmica*, 41(3):185–202, 2005.

[26] Juha Heinonen. *Lectures on Analysis on Metric Spaces*. Springer-Verlag New York. doi:10.1007/978-1-4613-0131-8.

[27] Dorit S Hochbaum and David B Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM (JACM)*, 33(3):533–550, 1986.

[28] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM (JACM)*, 53(3):307–323, 2006.

[29] Anil K Jain and Richard C Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[30] Kamal Jain and Vijay V Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of the ACM (JACM)*, 48(2):274–296, 2001.

[31] Nicholas Jardine and Robin Sibson. Mathematical taxonomy. *London etc.: John Wiley*, 1971.

[32] Daniel M Kane, Jelani Nelson, and David P Woodruff. On the exact space complexity of sketching and streaming small norms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1161–1178. SIAM, 2010.

[33] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. A local search approximation algorithm for k-means clustering. In *Proceedings of the eighteenth annual symposium on Computational geometry*, pages 10–18. ACM, 2002.

[34] Madhukar R Korupolu, C Greg Plaxton, and Rajmohan Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of algorithms*, 37(1):146–188, 2000.

[35] Stuart P Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.

[36] Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM (JACM)*, 41(5):960–981, 1994.

[37] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. The planar k-means problem is np-hard. In *WALCOM: Algorithms and Computation*, pages 274–285. Springer, 2009.

[38] Jirı Matoušek. On approximate geometric k-clustering. *Discrete & Computational Geometry*, 24(1):61–84, 2000.

[39] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, Aug 2011. URL: https://doi.org/10.1007/s10208-011-9093-5, doi:10.1007/s10208-011-9093-5.

[40] Facundo Mémoli. Some properties of Gromov–Hausdorff distances. *Discrete & Computational Geometry*, 48(2):416–440, 2012.

[41] Ramgopal R. Mettu and C. Greg Plaxton. The online median problem. *SIAM J. Comput.*, 32(3):816–832, March 2003. URL: http://dx.doi.org/10.1137/S0097539701383443, doi:10.1137/S0097539701383443.

[42] Felix Schmiedl. Computational aspects of the gromov-hausdorff distance and its application in non-rigid shape matching. *Discrete & Computational Geometry*, 57(4):854–880, 2017. URL: https://doi.org/10.1007/s00454-017-9889-4, doi:10.1007/s00454-017-9889-4.

[43] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta Mathematica*, 196(1):65–131, Jul 2006. URL: https://doi.org/10.1007/s11511-006-0002-8, doi:10.1007/s11511-006-0002-8.

[44] Mikkel Thorup. Quick k-median, k-center, and facility location for sparse graphs. In *Automata, Languages and Programming*, pages 249–260. Springer, 2001.

[45] Andrea Vattani. K-means requires exponentially many iterations even in the plane. *Discrete & Computational Geometry*, 45(4):596–616, 2011.

[46] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. URL: https://books.google.com/books?id=GqRXYFxe0l0C.