2018 IEEE 59th Annual Symposium on Foundations of Computer Science

ε-Coresets for Clustering (with Outliers) in Doubling Metrics

(Full version: [1])

Lingxiao Huang*, Shaofeng H.-C. Jiang†, Jian Li‡ and Xuan Wu‡
*École polytechnique fédérale de Lausanne, Lausanne, Switzerland
Email: lingxiao.huang@epfl.ch
†Weizmann Institute of Science, Rehovot, Israel
Email: shaofeng.jiang@weizmann.ac.il
‡ Tsinghua University, Beijing, China
Email: {lijian83@mail.tsinghua.edu.cn, wu3412790@gmail.com}

***Abstract*— We study the problem of constructing ε-coresets for the (k, z)-clustering problem in a doubling metric $M(X, d)$. An ε-coreset is a weighted subset $S \subseteq X$ with weight function $w : S \rightarrow \mathbb{R}_{\geq 0}$, such that for any k-subset $C \in [X]^k$, it holds that $\sum_{x \in S} w(x) \cdot d^z(x, C) \in (1 \pm \varepsilon) \cdot \sum_{x \in X} d^z(x, C)$.**

We present an efficient algorithm that constructs an ε-coreset for the (k, z)-clustering problem in $M(X, d)$, where the size of the coreset only depends on the parameters k, z, ϵ and the doubling dimension $\mathsf{ddim}(M)$. To the best of our knowledge, this is the first efficient ϵ-coreset construction of size *independent* of $|X|$ for general clustering problems in doubling metrics.

To this end, we establish the first relation between the doubling dimension of $M(X, d)$ and the shattering dimension (or VC-dimension) of the range space induced by the distance d. Such a relation is not known before, since one can easily construct instances in which neither one can be bounded by (some function of) the other. Surprisingly, we show that if we allow a small $(1 \pm \epsilon)$-distortion of the distance function d (the distorted distance is called the smoothed distance function), the shattering dimension can be upper bounded by $O(\epsilon^{-O(\mathsf{ddim}(M))})$. For the purpose of coreset construction, the above bound does not suffice as it only works for unweighted spaces. Therefore, we introduce the notion of τ-error *probabilistic shattering dimension*, and prove a (drastically better) upper bound of $O(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log \frac{1}{\tau})$ for the probabilistic shattering dimension for weighted doubling metrics. As it turns out, an upper bound for the probabilistic shattering dimension is enough for constructing a small coreset. We believe the new relation between doubling and shattering dimensions is of independent interest and may find other applications.

Furthermore, we study *robust coresets* for (k, z)-clustering with outliers in a doubling metric. We show an improved connection between α-approximation and robust coresets. This also leads to improvement upon the previous best known bound of the size of robust coreset for Euclidean space [Feldman and Langberg, STOC 11]. The new bound entails a few new results in clustering and property testing.

As another application, we show constant-sized (ε, k, z)-centroid sets in doubling metrics can be constructed by extending our coreset construction. Prior to our result, constant-sized centroid sets for general clustering problems were only known for Euclidean spaces. We can apply our centroid set to accelerate the local search algorithm (studied in [Friggstad et al., FOCS 2016]) for the (k, z)-clustering problem in doubling metrics.

***Keywords*-coreset, clustering, doubling dimension, centroid set, outlier**

I. Introduction

We study the (k, z)-clustering problem in a metric space $M(X, d)$. In the (k, z)-clustering problem, the objective is to find a k-subset $C \in [X]^k$ (which we call the set of centers), such that the objective function $\mathcal{K}_z(X, C) := \sum_{x \in X} d^z(x, C)$ is minimized, where $d(x, C) := \min_{y \in C} d(x, y)$. The (k, z)-clustering problem is a general and fundamental problem in many areas including approximation algorithms, unsupervised learning and computational geometry [2], [3], [4], [5]. In particular, $(k, 1)$-clustering is the well known k-median problem, $(k, 2)$-clustering the k-means problem, and (k, ∞)-clustering the k-center problem.

Coresets. A powerful technique for solving the (k, z)-clustering problem is to construct coresets [6], [7], [8], [9]. A coreset is a weighted subset of the point set, such that for any set of k centers, the objective function computed from the coreset is approximately the same as that computed from all points in X. Hence, a coreset can be used as proxy for the full data set: one can apply the same algorithm on the coreset, and the result on the coreset approximates that on the full data set.

Definition I.1. *An ε-coreset for the (k, z)-clustering problem in metric space $M(X, d)$ is a weighted subset S of X with weight $w : S \rightarrow \mathbb{R}_{\geq 0}$* [1]*, such that for any k-subset $C \in [X]^k$,*

$$\sum_{x \in S} w(x) \cdot d^z(x, C) \in (1 \pm \varepsilon) \cdot \mathcal{K}_z(X, C).$$

Typically, we require that the size of the coreset depends on $1/\varepsilon$, k and z (independent of $|X|$). Apparently, a small coreset is much cheaper to store and can be used to estimate the objective function more efficiently. In fact, constructing

[1] Some previous work needs negative weights, but we only need nonnegative weights.

DOI 10.1109/FOCS.2018.00082

coresets can be useful in designing more efficient approximation algorithms for many clustering problems, with various constraints and outliers [8], [10], [9], [11], [12], [13].

Doubling Metrics. In this paper, we mainly consider metric spaces with bounded *doubling dimension* [14], [15]. The doubling dimension of a metric space M, denoted as $\mathsf{ddim}(M)$, is the smallest integer t such that any ball can be covered by at most 2^t balls of half the radius. A doubling metric is a metric space of bounded doubling dimension. The doubling dimension measures the intrinsic dimensionality of a general metric space, and it generalizes the dimension of normed vector spaces, where t-dimensional ℓ_p space has doubling dimension $O(t)$ [14].

Many problems have been studied in doubling metrics, such as spanners [16], [17], [18], [19], [20], [21], [22], [23], [24], metric embedding [15], [25], [26], nearest neighbor search [27], [28], [29], and approximation algorithms [30], [31], [32], [33], [34], [12]. Apart from the above work, some machine learning problems have also been studied in the context of doubling metrics [35], [36]. However, to the best of our knowledge, no previous work has studied constructing coresets in doubling metrics.

A. Our Results

Our main result is an efficient construction of ε-coresets for the (k, z)-clustering problem in doubling metrics. The size of our coreset does not depend on the number of input points. Moreover, both the running time and the size of the coreset depend *polynomially* on the doubling dimension and k. The result is stated in the following theorem.

Theorem I.1. *(informal version of Theorem VI.1) Consider a metric space $M(X, d)$ with n points. Let real numbers $0 < \varepsilon, \tau < 1/100$, $z > 0$, and integer $k \geq 1$. There exists an algorithm running in $\mathrm{poly}(n)$ time (assuming oracle access to the distance function), that constructs an ε-coreset of size $\tilde{O}(2^{O(z \log z)} \cdot k^3 \cdot \mathsf{ddim}(M)/\varepsilon^2)$ for the (k, z)-clustering problem with probability at least $1 - \tau$.*

A first natural attempt is to embed the doubling space to the Euclidean space and use the existing Euclidean construction. As shown in [15, Theorem 4.5], for a doubling metric $M(X, d)$, it is possible to embed $d^{\frac{1}{2}}$ to an $O(\mathsf{ddim}(M) \cdot \log \mathsf{ddim}(M))$-dimensional ℓ_2 space with $O(\mathsf{ddim}(M))$-distortion. Then an ε-coreset for $(k, 2z)$-clustering problem in ℓ_2 would imply an $O(\mathsf{ddim}(M)^z)$-coreset for the (k, z)-clustering problem in M. However, it is generally not possible to embed $(X, d^{\frac{1}{2}})$ into ℓ_2 with $(1 + \varepsilon)$-distortion for an arbitrarily small constant $\varepsilon > 0$ and doubling metric $M(X, d)$ (where an example can be found in the full version). Hence, in order to construct an ε-coreset in a doubling metric, we need new ideas.

A by now standard technique for constructing small coresets for clustering problems is importance sampling, developed in a series of work [37], [8], [38]. In particular, by the framework in [8], [38], one can obtain an ε-coreset by taking $\tilde{O}(2^{O(z \log z)} \cdot k^3 \cdot \mathsf{dim}/\varepsilon^2)$ samples. Here dim is the (shattering) dimension of the range space induced by the distance function. (i.e., the range space consists of all balls of different radii [2]). Hence, if one can show that dim is bounded by some function of $\mathsf{ddim}(M)$, the construction of an ε-coreset would be finished.

Doubling Dimension and Shattering Dimension. We discuss the relation between the doubling dimension $\mathsf{ddim}(M)$ and the (shattering) dimension dim of the range space. While the dimension dim measures the combinatorial complexity of the metric space, doubling dimension $\mathsf{ddim}(M)$ is the intrinsic geometric dimension of the metric space. They both generalize the ordinary Euclidean dimension, but from different perspectives. In particular, for $\mathbb{R}^d$, both the (shattering) dimension and the doubling dimension are $O(d)$. Although both dimensions are subjects of extensive research, to the best of our knowledge, there is no nontrivial relation known between the two. This may not be a surprise, as we can easily construct a doubling metric, which has unbounded shattering dimension on the corresponding induced range space (see Theorem IV.2). The other direction cannot be bounded neither [3]. Hence, studying their relation may appear to be hopeless. However, we observe that in the bad instance in Theorem IV.2, if we allow a $(1 \pm \varepsilon)$-distortion to the distance function d, then the instance actually has a small shattering dimension.

Inspired by this observation, we introduce the *smoothed distance function*. A ε-smoothed distance function $\delta : X \times X \to \mathbb{R}_{\geq 0}$ satisfies $\delta(x, y) \in (1 \pm O(\varepsilon)) \cdot d(x, y)$ for all $x, y \in X$. Basically, it is a small perturbation of the original distance function d. We show, somewhat surprisingly, that if we use a certain smoothed distance function δ, defined by a hierarchical net of the doubling metric, the shattering dimension of the range space (induced by δ, instead of d) can be upper bounded by some function of the doubling dimension $O(\mathsf{ddim}(M))$, as in the following theorem.

Theorem I.2. *(informal, unweighted case) Suppose $M(X, d)$ is a metric space. Let $0 < \varepsilon \leq \frac{1}{8}$ be a constant. There is some ε-smoothed distance function such that $\mathsf{dim}(\mathcal{F}) \leq O(1/\varepsilon)^{\mathsf{ddim}(M)}$, where $\mathcal{F} := \{\delta(x, \cdot) \mid x \in X\}$ is the set of ε-smoothed distance functions.*

While the above theorem is encouraging, there are still some drawbacks. First, the dimension bound is exponential in $\mathsf{ddim}(M)$ (in contrast to the linear dependency in Euclidean case). It is a natural question whether one can obtain a better bound in general. More importantly, the above bound is not sufficient for the purpose of constructing small

[2] In fact, we will deal with the range space in a certain function space. See Section IV for the precise definition.

[3] Consider a star with n leaves. It is immediate that the metric induced by the star has doubling dimension $\Omega(\log n)$. However, the shattering dimension of the range space induced by the star metric is $O(1)$.

coresets, for which we need a dimension bound for weighted spaces. Unfortunately, it seems difficult to extend the proof of Theorem IV.2 to the weighted case.

Weighted Space and Probabilistic Shattering Dimension. Recall that our goal is to construct ε-coresets for the (k, z)-clustering problem. Following the framework [8], we consider the set of *weighted* distance functions $g_x : [X]^k \to \mathbb{R}_{\geq 0}$ for each point $x \in X$, defined as $g_x(C) := w(x) \cdot \delta^z(x, C)$ for $C \in [X]^k$, where $w : X \to \mathbb{R}_{\geq 0}$ is a weight function and δ is an $O(\varepsilon/z)$-smoothed distance function. We consider the function set $\mathcal{G} := \{g_x \mid x \in X\}$, and show that there is a subset $\mathcal{S} \subseteq \mathcal{G}$ such that $\mathcal{S}$ is an α-approximation for the range space of $\mathcal{G}$ for a certain constant α. Then we can apply [8, Theorem 4.1] to show that one can efficiently find an $\varepsilon(\alpha)$-coreset of size $|\mathcal{S}|$.

In order to prove an α-approximation result, it suffices to bound the shattering dimension of $\mathcal{G}$. We recall that $\mathcal{G} = \{g_x \mid x \in X\}$ with $g_x(C) = w(x) \cdot \delta^z(x, C)$ for $C \in [X]^k$. Let $\mathcal{F} := \{f_x \mid x \in X\}$ be a collection of functions $f_x(y) := w(x) \cdot \delta^z(x, y)$ for $y \in X$. Note that the difference between $\mathcal{F}$ and $\mathcal{G}$ is that, the ground set of $\mathcal{F}$ consists of singletons and that of $\mathcal{G}$ contains k-subsets. By a standard argument, one can show that roughly the shattering dimension of $\mathcal{G}$ is at most k times of the shattering dimension of $\mathcal{F}$. Hence, the key is to bound the shattering dimension of $\mathcal{F}$ (the set of weighted smoothed distance function with ground set X). It turns out the proof for the weighted case is much more involved than the unweighted case. Instead of using a deterministic δ defined with respect to d, we introduce a random smoothed distance function, defined on top of a randomized hierarchical decomposition introduced by [25] in the doubling metric. A key property of the randomized hierarchical decomposition is that a set with small diameter is cut by a large cluster in the decomposition with very small probability. Intuitively, the property enhances the smooth property, so that we can still hang the balls centered at any $x \in X$ to a net point of higher layer in the weighted space.

Consider an arbitrary fixed $H \subseteq X$ and $\mathcal{F}_H := \{f_x \in \mathcal{F} \mid x \in H\}$. Due to the randomness in the randomized hierarchical decomposition, we can only show $|\mathsf{ranges}(\mathcal{F}_H)|$ is bounded with *constant* (close to 1) probability. Hence, we introduce the notion of *probabilistic shattering dimension* for the range space induced by a family of random functions (formally in Definition III.4): for any subset of (random) functions $\mathcal{F}_H$, if the probability that $|\mathsf{ranges}(\mathcal{F}_H)| \leq O(|\mathcal{F}_H|^t)$ with probability $1 - \tau$ (note that $|\mathsf{ranges}(\mathcal{F}_H)|$ is a random variable), we say that the probabilistic shattering dimension $\mathsf{pdim}_\tau(\mathcal{F})$ of the range space is t. Our main technical result is the following theorem.

Theorem I.3. *(informal, weighted case) Suppose $M(X, d)$ is a metric space together with a gap-2 weight function (Definition V.1) $w : X \to \mathbb{R}_{\geq 0}$. Let $0 < \varepsilon \leq 1/100z$ and $0 < \tau < 1$ be constants. There exists a random ε-smoothed distance function δ and a collection $\mathcal{F} := \{f_x = w(x) \cdot \delta^z(x, \cdot) \mid x \in X\}$, such that the following holds: for any fixed $H \subseteq X$ and $\mathcal{F}_H := \{f_x : x \in H\}$, we have*

$$\Pr_\delta \left[|\mathsf{ranges}(\mathcal{F}_H)| \leq O\left(\varepsilon^{O(-\mathsf{ddim}(M))} \cdot \log \frac{1}{\tau} \cdot \mathrm{poly}(|H|) \right) \right] \geq 1 - \tau.$$

In other words, $\mathsf{pdim}_\tau(\mathcal{F}) = O(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log 1/\tau)$.

The above theorem drastically improves the dimension from $O(1/\varepsilon)^{\mathsf{ddim}(M)}$ (in Theorem IV.2) to $\tilde{O}\left(\mathsf{ddim}(M) \cdot \log(1/\varepsilon)\right)$ (albeit with a weaker probabilistic guarantee). Note that one cannot afford to apply a union bound over all different H's to show that $\dim(\mathcal{F})$ is bounded. Hence, the bound on the probabilistic shattering dimension does not directly lead to an α-approximation by the standard PAC learning theory. However, we prove in Lemma III.1 a probabilistic analogue of the α-approximation lemma, which only requires a bounded probabilistic shattering dimension.

Robust Coreset. We also consider robust coresets which are coresets for (k, z)-clustering problems with outliers. The notion of robust coreset was first introduced in [8]. In the following, we give the definition of robust coreset for the (k, z)-clustering problem.

Definition I.2 (robust coresets)**.** *Let $M(X, d)$ be a metric space. Let $0 < \gamma \leq 1$, $0 \leq \varepsilon, \alpha \leq \frac{1}{4}$, $k \geq 1$ and $z > 0$. For any $H \subseteq X$ and $C \in [X]^k$, let*

$$\mathcal{K}_z^{-\gamma}(H, C) := \min_{H' \subseteq H : |H'| = \lceil (1-\gamma)|H| \rceil} \sum_{x \in H'} d^z(x, C)$$

denote the sum of the smallest $\lceil (1-\gamma)|H| \rceil$ values $d^z(x, C)$ over $x \in H$ (i.e., we exclude the largest $\gamma|H|$ values as outliers). An (α, ε)-robust coreset for the (k, z)-clustering problem with outliers is a subset $S \subseteq X$ such that for any k-subset $C \in [X]^k$ and any $\alpha < \gamma < 1 - \alpha$,

$$(1-\varepsilon) \cdot \frac{\mathcal{K}_z^{-(\gamma+\alpha)}(X, C)}{|X|} \leq \frac{\mathcal{K}_z^{-\gamma}(S, C)}{|S|} \leq (1+\varepsilon) \cdot \frac{\mathcal{K}_z^{-(\gamma-\alpha)}(X, C)}{|X|}.$$

Our result for robust coreset for (k, z)-clustering is presented in the following theorem, which generalizes and improves the prior result in [8] for Euclidean space.

Theorem I.4 (informal, robust coreset)**.** *Let $M(X, d)$ be a doubling metric (a d-dimensional Euclidean space resp.). Let S be a uniform sample of size $\tilde{O}(k \cdot \mathrm{ddim}(M)/\alpha^2)$ ($\tilde{O}(kd/\alpha^2)$ resp.) from X. Then with constant probability, S is an (α, ε)-robust coreset ($(\alpha, 0)$-robust coreset resp.) for the (k, z)-clustering problem with outliers.*

The definition of robust coreset in [8] is slightly different from ours. [4] One can directly check that in Euclidean space, an $(\gamma\varepsilon/4, 0)$-robust coreset in Definition I.2 is an (γ, ε)-coreset in [8, Definition 8.1]. Thus the above theorem improves the size of (γ, ε)-coreset in [8, Corollary 8.4] from $\tilde{O}(kd\gamma^{-2}\varepsilon^{-4})$ to $\tilde{O}(kd\gamma^{-2}\varepsilon^{-2})$.

Furthermore, we demonstrate an application of robust coresets in property testing (omitted, see the full version). Our testing for (k, z)-clustering problem is similar to the testing for k-center problem proposed in Alon et al. [39]. We design a simple testing algorithm for (k, z)-clustering.

Constructing robust coresets is also a useful subroutine in several other problems, such as robust median and bi-criteria approximation for projective clustering (see [8]). Hence, our improvement may lead to certain improvements of these problems as well. Since this is not the focus of the this paper, we do not go into the details.

Centroid Set. We also consider a notion closely related to coreset, called centroid set. Roughly speaking, a centroid set can be viewed as a coreset that contains an $(1+\varepsilon)$-approximate solution (which is a k-subset) to the clustering objective (see Definition VI.1). Applying our coreset result, we show the existence of succinct centroid sets in doubling metrics, which is presented in the following theorem. To the best of our knowledge, this is the first result on centroid sets beyond Euclidean spaces.

Theorem I.5 (informal, centroid set)**.** *Let $M(X, d)$ be a metric space of n discrete points. Let S be an $\frac{\varepsilon}{2}$-coreset for the (k, z)-clustering problem on X. There is an algorithm running in $\mathrm{poly}(n)$ time, that finds a centroid set of size at most $(\frac{z}{\varepsilon})^{O(\mathsf{ddim}(M))} \cdot |S|^2$.*

Applying the above theorem, we can also accelerate the local search algorithm [12] for (k, z)-clustering in doubling metrics, from $n^{O(\rho)}$ to $(2^{O(z\log z)} \cdot \frac{k}{\varepsilon})^{O(\rho)}$ running time per iteration, where $n = |X|$ is the number of points and $\rho := \rho(\epsilon, \mathsf{ddim}(M), z)$ is a large constant (depending only on $\epsilon, \mathsf{ddim}(M), z$).

B. Overview of Our Techniques

The Feldman-Langberg Framework [8]. Our coreset construction makes use of the the framework of [8], which we briefly discuss below. Let $[X]^k$ be the ground set (the set of k-tuples) and δ be an $O(\varepsilon/z)$-smoothed distance function. For the (k, z)-clustering problem, assign a *weighted* distance function $g_x : [X]^k \to \mathbb{R}_{\geq 0}$ to each point $x \in X$, such that $g_x(C) := w(x) \cdot \delta^z(x, C)$ for $C \in [X]^k$, where $w : X \to \mathbb{R}_{\geq 0}$ is a weight function. Consider the function set $\mathcal{G} := \{g_x \mid x \in X\}$. The range spaces of $\mathcal{G}$ is defined as $(\mathcal{G}, \mathsf{ranges}(\mathcal{G}))$, where $\mathsf{ranges}(\mathcal{G}) := \{\mathsf{range}(\mathcal{G}, C, r) \mid C \in [X]^k, r \geq 0\}$ and $\mathsf{range}(\mathcal{G}, C, r) := \{g_x \in \mathcal{G} \mid g_x(C) \leq r\}$. To interpret the definition, one can think of $\{g_x \in \mathcal{G} \mid g_x(C) \leq r\}$ as a ball of functions in $\mathcal{G}$ that is centered at C with radius r, and the distance from C to $g_x \in \mathcal{G}$ is measured as $g_x(C)$. In the unweighted case, $\mathsf{range}(\mathcal{G}, x, r)$ indeed corresponds to a ball in the metric space. So $|\mathsf{ranges}(\mathcal{G})|$ counts the number of distinct balls (of functions in $\mathcal{G}$) that may be formed by any center in the ground set and radii.

Recall that a subset $\mathcal{S} \subseteq \mathcal{G}$ is an α-approximation for the range space $(\mathcal{G}, \mathsf{ranges}(\mathcal{G}))$ if for any $\mathcal{R} \in \mathsf{ranges}(\mathcal{G})$, $\big||\mathcal{R}|/|\mathcal{G}| - |\mathcal{R} \cap \mathcal{S}|/|\mathcal{S}|\big| \leq \alpha$. In other words, $\mathcal{S}$ can be used as a good estimator for the density of $\mathcal{R}$ relative to $\mathcal{G}$. It is shown in [8, Theorem 4.1] that, if there is a subset $\mathcal{S} \subseteq \mathcal{G}$ such that $\mathcal{S}$ is an α-approximation for the range space of $\mathcal{G}$, then we can efficiently construct an $\varepsilon(\alpha)$-coreset $S \subseteq X$ of size $|\mathcal{S}|$ in $M(X, \delta)$. Since δ is a small perturbation of the original distance function d, S is also an $\varepsilon(\alpha)$-coreset in $M(X, d)$. Constructing an α-approximation of small size is extensively studied in the PAC learning theory. In particular, if a range space has bounded shattering (or VC) dimension, then a small sample (whose size depends on α and shattering dimension) from the set of functions would be an α-approximation with constant probability (see e.g., [40]). Hence, if $\mathcal{G}$ has bounded shattering dimension, we can apply the existing α-approximation construction. This is also the approach taken in [8].

α-Approximation. As one can imagine, in order to obtain an α-approximation, we would like to apply Theorem I.3 (to bound the shattering dimension of $(\mathcal{G}, \mathsf{ranges}(\mathcal{G}))$). More precisely, for any $H \subseteq X$, we want to bound $|\mathsf{ranges}(\mathcal{G}_H)|$ where $\mathcal{G}_H := \{g_x \in \mathcal{G} \mid x \in H\}$. However, the ground set of $\mathcal{G}_H$ is the set of k-subsets $[X]^k$ (with the distance function in $\mathcal{G}_H$ defined as $g_x(C) := w(x) \cdot \delta^z(x, C)$), but the ground set of $\mathcal{F}_H$ in Theorem I.3 is the point set X of the metric space $M(X, d)$. This is easy to handle: one can show that $|\mathsf{ranges}(\mathcal{G}_H)| \leq |\mathsf{ranges}(\mathcal{F}_H)|^k$. Hence, we only need to bound $|\mathsf{ranges}(\mathcal{F}_H)|$. Another problem is that the bound for $|\mathsf{ranges}(\mathcal{F}_H)|$ only holds with constant probability. As a result, we cannot directly use the standard α-approximation result. In Lemma III.1, we introduce a probabilistic analogue of the α-approximation lemma from the PAC learning theory, which only requires a bounded probabilistic shattering dimension.

Our proof borrows the classical double sampling idea from the construction of α-nets in the PAC learning theory (see for example [41]). An obvious challenge is that we cannot afford to guarantee $|\mathsf{ranges}(\mathcal{F}_H)|$ is small for many H simultaneously by the union bound, which is required in the original proof. Note that our guarantee has an additional randomness from δ, and it is important to take advantage the additional randomness. We crucially use the fact that $\mathcal{F}$ is actually indexed by X, that is, for each $x \in X$, a function $f_x \in \mathcal{F}$ is generated by applying a *random* map

[4] In [8, Definition 8.1], $S \subset X$ is called a (γ, ε)-coreset if for every $C \in [X]^k$, $\gamma_1 \geq \gamma$ and $\varepsilon_1 \geq \varepsilon$, $(1-\varepsilon_1) \cdot \frac{1}{|X|}\mathcal{K}_1^{-(1-\gamma_1+\varepsilon_1\gamma_1)}(X, C) \leq \frac{1}{|S|}\mathcal{K}_1^{-(1-\gamma_1)}(S, C) \leq (1+\varepsilon_1) \cdot \frac{1}{|X|}\mathcal{K}_1^{-(1-\gamma_1-\varepsilon_1\gamma_1)}(X, C)$.

from x to $w(x) \cdot \delta(x, \cdot)$. This enables us to separate the two randomness, in a way that we view a sample from $\mathcal{F}$, as firstly sampling from X then applying a *random* map on the sample. The randomness of sampling from X is used in the similar way as in the original proof, but the randomness of δ is used in another conditional probability argument to avoid the overlarge union bound. The details can be found in the full version.

Doubling Dimension and Shattering Dimension. Now, we highlight some technical aspects of Theorems I.2 and I.3, which are the key technical contributions of this paper. Our smoothed distance function δ is defined over the hierarchical net tree (see Section III-A for the definition) of the doubling metric $M(X, d)$, i.e., for any $x, y \in X$, we define $\delta(x, y)$ to be the distance between their ancestors of a proper height in the hierarchical net tree (see Definition IV-A). We also define $B^{\delta}(x, r) := \{y \in X \mid \delta(x, y) \leq r\}$ to be the ball of radius r centered at $x \in X$ with respect to δ. We can show the smoothed distance function δ satisfies several useful properties. One is the *smooth* property (Lemma IV.3): roughly speaking, for any radius r, $B^{\delta}(x, r) = B^{\delta}(u, r)$ for any point $x \in X$ and a nearby net point u with higher height (relative to r). This intuitively means we can "hang" the center x to the net point u. Since the number of net points with higher height is smaller, the smooth property greatly reduces the number of possible balls we need to consider. Another important property is the cross-free property (Lemma IV.4), which implies that, if we let $\mathcal{F}$ be the range space induced by δ, then for any fixed $\mathcal{D} \subseteq \mathcal{F}$, for any ball $\mathsf{range}(\mathcal{F}, x, r)$ in the range space, $\mathsf{range}(\mathcal{F}, x, r) \cap \mathcal{D}$ can be represented by a union of at most $\varepsilon^{-O(\mathsf{ddim}(M))}$ subsets, and all these subsets are from a support of size $O(|\mathcal{D}|)$. This implies that at most $|\mathcal{D}|^{\varepsilon^{-O(\mathsf{ddim}(M))}}$ possible subsets of $\mathcal{D}$ can be formed by intersecting with balls of a fixed radius in the metric, which is the main observation used in our proof for Theorem I.2.

Unfortunately, it seems difficult to extend the above idea to the weighted case, and our proof for Theorem I.3 is much more involved. We restrict our attention to the weight function such that the set of distinct weights $\{w_1, w_2, \ldots, w_l\}$ satisfies $w_1 \geq 2w_2 \geq 4w_3 \geq \ldots \geq 2^{l-1}w_l$ (this suffices for the coreset construction). Fix a set $H \subseteq X$ and let $H_i = \{x \in H \mid w(x) = w_i\}$. Essentially, we need to bound the number of different ranges $\bigcup_{i \in [l]} B^{\delta}(x, r/w_i) \cap H_i$ $(r > 0, x \in X)$. For this purpose, we divide $[0, +\infty)$ (the range of r) into at most $O(|H|^4)$ *critical intervals*. Inside each critical interval, we enforce some invariance properties. For a critical interval $[a, b)$ with $b \leq 2a$, we simply apply the packing property to bound the number of different ranges for $r \in [a, b)$. For a critical interval $[a, b)$ with $b \gg a$, we need to use the randomized hierarchical decomposition developed in [25] to enhance the smooth property.

Robust Coreset. We prove an improved connection between α-approximation and robust coreset, which improves the one in [8, Theorem 8.3]. Our proof is much simpler. Combining with the α-approximation result, we can construct an (α, ε)-robust coreset in Euclidean space or doubling metrics. The algorithm is extremely simple: to take a uniform sample of size $\tilde{O}(kd/\alpha^2)$ or $\tilde{O}\big(k \cdot \mathsf{ddim}(M)/\alpha^2\big)$.

II. Related Work

In the seminal paper [42], Agarwal et al. proposed the notion of coresets for the directional width problem (in which a coreset is called an ϵ-kernel) and several other geometric shape-fitting problems. Since then, coresets have become increasingly more relevant in the era of big data as they can reduce the size of a dataset with provable guarantee that the answer on the coreset is a close approximation of the one on the whole dataset. Many efficient algorithms for constructing small coresets for clustering problems in Euclidean spaces are known (see e.g., [43], [44], [7], [45], [37], [8], [9], [11]). In particular, Feldman and Langberg [8] (see their latest full version) showed a construction for ε-coresets of size $\tilde{O}(dk/\varepsilon^{2z})$ for general (k, z)-clustering problems with arbitrary k and z, in $\tilde{O}(nk)$ time. For the special case that $z = 2$ which is the k-means clustering, Braverman et al. [11] improved the size to $\tilde{O}(k \min\{k/\varepsilon, d\}/\varepsilon^2)$, which is *independent* of the dimensionality d. For another special case $z = \infty$, which is the k-center clustering, an ε-coreset of size $O(k/\varepsilon^d)$ can be constructed in $O(n + k/\varepsilon^d)$ time, for $\mathbb{R}^d$ [43], [44]. For general metrics, an ε-coreset for the (k, z)-clustering problem of size $O(k \log n/\varepsilon^{2z})$ can be constructed in time $\tilde{O}(nk)$ [8], and for k-means clustering, Braverman et al. [11] showed a construction of size $O(k \log k \log n/\varepsilon^2)$. We also refer interested readers to Phillips's survey [46] for more construction algorithms as well as the applications of coresets in many other areas.

Feldman and Langberg [8] first studied the notion of robust coreset to handle the clustering problems with outliers. In $\mathbb{R}^d$, they showed how to construct a (γ, ε)-coreset [5] of size $\tilde{O}(kd\varepsilon^{-4}\gamma^{-2})$ by uniform sampling. We improve the bound to $\tilde{O}(kd\varepsilon^{-2}\gamma^{-2})$. Later, Feldman et al. [10] developed another notion called *weighted coreset* to handle outliers. They used such coresets to design an $(1+\varepsilon)$-approximation algorithm for the k-median problem with outliers.

Constructing coresets for clustering problems in Euclidean spaces has been also investigated in the streaming and distributed settings in the literature e.g., [8], [9], [47], [11], [48]). However, it is unclear how to define the streaming or distributed model in a general doubling metric, since there is no coordinate representation for each point and we need all distances between the new coming point and the prior

[5] Note that their definition [8, Definition 8.1] is similar but slightly different to ours. However, considering the (k, z)-clustering problem with outliers, one can check that an $(\varepsilon\gamma/4, \varepsilon)$-robust coreset in our Definition I.2 is a (γ, ε)-coreset in [8, Definition 8.1]. In fact, our defintion is more general. It is unclear whether their result applies to our definition.

points. Hence, in this paper, we focus on the centralized setting.

Besides unsupervised clustering problem, some supervised learning problems are also studied in the context of doubling metrics, and the connections between doubling dimension and VC dimension (and closely related notions) have been investigated in a variety of settings. Bshouty, Li and Long [35] provided a generalization bound in terms of the maximum of the doubling dimension and the VC-dimension of the hypothesis class F. They also showed that the doubling dimension of metric (F, d), where the distance d is defined as $d(f, g) = \Pr_x[f(x) \neq g(x)]$ for any two classifiers f and g, cannot be bounded by the VC-dimension of F in general. Gottlieb et al. [36] studies the classification problem of points in a metric space, and obtained a generalization bound with respect to the doubling dimension. Abraham et al. [35] introduced the concept of highway dimension (which is closely related to doubling dimension) in the context of designing efficient shortest path algorithm, and they showed that VC dimension and learning theory are also useful in this context.

Alon et al. [39] first considered the property testing problem in the context of clustering. In particular, they studied the testing algorithm for k-center clustering. In this paper, we use robust coreset to develop a unified testing algorithm for (k, z)-clustering (for constant k and z). As pointed out in [39], the testing algorithms can be converted into a sublinear time approximation algorithms for clustering with outliers. One interesting benefit of such algorithms is that they can answer the query "which cluster does a data point belong to", without actually having to partition all the data points.

III. Preliminaries

Let $[m] := \{1, 2, \ldots, m\}$ for an integer $m \geq 1$. For a function f defined on some ground set $\mathcal{U}$ and $S \subseteq \mathcal{U}$, let $f(S) := \{f(x) \mid x \in S\}$. For a set S and integer $k \geq 1$, let $[S]^k := \{P \mid P \subseteq S, |P| = k\}$. Consider a metric space $M(X, d)$. Define $B^{\xi}(x, r) := \{y \in X \mid \xi(x, y) \leq r\}$ to be the ball of radius r centered at $x \in X$, with respect to some function $\xi : X \times X \to \mathbb{R}_{\geq 0}$. For $S \subseteq X$ define the diameter of S as $\mathsf{diam}(S) := \max_{x,y \in S} \{d(x, y)\}$. For $S, T \subseteq X$, define $d(S, T) := \min_{x \in S, y \in T} d(x, y)$.

A. Doubling Dimension and Hierarchical Nets

Definition III.1 (doubling dimension)**.** *A metric space has doubling dimension at most t, if any ball can be covered by at most 2^t balls of half the radius. The doubling dimension of a metric space M is denoted as $\mathsf{ddim}(M)$.*

Covering, Packing and Net. Consider a subset of points $S \subseteq X$. S is a ρ-covering, if for any $x \in X$, there exists $y \in S$ such that $d(x, y) \leq \rho$. S is a ρ-packing, if for all $x, y \in S$, it holds that $d(x, y) \geq \rho$. S is a ρ-net, if S is both a ρ-packing and a ρ-covering.

Fact III.1 (packing property. see. e.g., [15])**.** *Given a metric space $M(X, d)$, if $S \subseteq X$ is a ρ-packing then $|S| \leq (\frac{2 \cdot \mathsf{diam}(S)}{\rho})^{\mathsf{ddim}(M)}$.*

Hierarchical Nets and Net Trees. Now, we introduce some useful concepts that are well known in the doubling metric literature (see e.g. [30], [24]). Rescale the metric such that the minimum intra-point distance is 1. Suppose the diameter of the space is between $[2^{L-1}, 2^L)$. Construct nets $N_L \subseteq N_{L-1} \subseteq \ldots \subseteq N_1 \subseteq N_0 = N_{-1} = \ldots = N_{-\infty} = X$, where N_i is a 2^i-net of N_{i-1}. The set of nets $\{N_i \mid i \leq L\}$ is called a hierarchical net.

We identify a point $u \in N_i$ in the tree by $u^{(i)}$ for $i \leq L$. Note that the same point may belong to several N_i's, but they have different identities. A net tree is a rooted tree with node set $\{u^{(i)} \mid i \leq L, u \in N_i\}$, and the root is defined as the only node in N_L (observing that $|N_L| = 1$). For each $u \in N_i$, $u^{(i)}$ has a unique parent node $v^{(i+1)}$ such that $v \in N_{i+1}$, and we denote $\mathsf{par}(u^{(i)}) = v^{(i+1)}$.

For a net tree, define $\mathsf{des}(u^{(i)}) \subseteq X$ to be the set of points in the metric space corresponding to descendants of $u^{(i)} \in N_i$ in the net tree. For a leaf node $x \in X$, define $\mathsf{par}^{(i)}(x)$ to be the ancestor of x in N_i.

Definition III.2 (c-covering net trees)**.** *A net tree is c-covering ($c \geq 1$), if for each height i and each $u \in N_i$, it holds that $d(u, \mathsf{par}(u^{(i)})) \leq c \cdot 2^{i+1}$.*

The following fact is immediate from Definition III.2.

Fact III.2. *In a c-covering net tree, for each $x \in X$ it holds that $d(x, \mathsf{par}^{(i)}(x)) \leq c \cdot 2^{i+1}$.*

B. Range Space, Shattering Dimension and α-Approximation

We adopt the function representation used in [8, Definition 7.2], but specifically tailored to our own needs. In particular, since we focus on the clustering problems in a doubling metric $M(X, d)$, the ground set is $[X]^k$ (the set of k-subsets) throughout the paper. When $k = 1$, we use X to represent $[X]^1$ for simplicity.

Indexed Function Sets. As seen in Section I, we mainly focus on range spaces induced by a metric space. Hence we always consider *indexed* function sets. A set of functions $\mathcal{F}$ is called *indexed*, if there exists an index set V such that $\mathcal{F} = \{f_x \mid x \in V\}$. In most cases, we simply use $V = X$ as the index set. We will make necessary clarification when we use other index set. For an indexed function set $\mathcal{F}$, define $\mathcal{F}_H := \{f_x \mid x \in H\}$ for a subset $H \subseteq V$ of the index set. There are technical reasons to consider the indexed function set (rather than a general set of functions). See Remark III.1.

Range Space. Let $\mathcal{F}$ be an indexed function set. Define $\mathsf{range}(\mathcal{F}, C, r) := \{f_x \in \mathcal{F} \mid f_x(C) \leq r\}$ for $C \in [X]^k, r \geq 0$. Define $\mathsf{ranges}(\mathcal{F}) := \{\mathsf{range}(\mathcal{F}, C, r) \mid C \in [X]^k, r \geq 0\}$ to be the collection of all the range sets. The range space of $\mathcal{F}$ is defined as the pair $(\mathcal{F}, \mathsf{ranges}(\mathcal{F}))$.

Now, We define the dimension of a range space, following [8].

Definition III.3 ((shattering) dimension of a range space)**.** *Suppose $\mathcal{F}$ is an indexed function set with ground set $[X]^k$. The (shattering) dimension of the range space $(\mathcal{F}, \mathsf{ranges}(\mathcal{F}))$, or simply the (shattering) dimension of $\mathcal{F}$, denoted as $\mathsf{dim}(\mathcal{F})$, is the smallest integer t, such that for any $\mathcal{D} \subseteq \mathcal{F}$ with $|\mathcal{D}| \geq 2$, $|\mathsf{ranges}(\mathcal{D})| \leq |\mathcal{D}|^t$. We note that in $\mathsf{ranges}(\mathcal{D})$, the same ground set $[X]^k$ is implicit.*

However, as discussed in Section I, our guarantee of the dimension for the weighted doubling distance functions only holds in a probabilistic sense. We capture this formally in the following.

Definition III.4 (probabilistic (shattering) dimension of a range space)**.** *Suppose $\mathcal{F}$ is a random indexed function set with a deterministic index set denoted as V. The τ-error probabilistic (shattering) dimension of $(\mathcal{F}, \mathsf{ranges}(\mathcal{F}))$, or simply the τ-error probabilistic dimension of $\mathcal{F}$, denoted as $\mathsf{pdim}_\tau(\mathcal{F})$, is the smallest integer t such that for any fixed $H \subseteq V$ with $|H| \geq 2$, $|\mathsf{ranges}(\mathcal{F}_H)| \leq |H|^t$ with probability at least $1 - \tau$.*

We need a well studied notion in the PAC learning theory, called α-approximation.

Definition III.5 (α-approximation of a range space)**.** *Given a range space $(\mathcal{F}, \mathsf{ranges}(\mathcal{F}))$ (with ground set $[X]^k$), a set $\mathcal{S} \subseteq \mathcal{F}$ is an α-approximation of the range space, if for every $\mathsf{range}(\mathcal{F}, C, r) \in \mathsf{ranges}(\mathcal{F})$ ($C \in [X]^k, r \geq 0$)*

$$\left| \frac{|\mathsf{range}(\mathcal{F}, C, r)|}{|\mathcal{F}|} - \frac{|\mathcal{S} \cap \mathsf{range}(\mathcal{F}, C, r)|}{|\mathcal{S}|} \right| \leq \alpha.$$

In particular, it was shown that a small sized (depending on α and the VC dimension[6]) independent sample from the function set is an α-approximation with constant probability (see for example [40]). However, the traditional results are for range spaces with bounded VC dimension only, and our probabilistic dimension is very different in nature. We prove the following version of the sampling bound that only requires a bounded probabilistic dimension. The proof can be found in the full version.

Lemma III.1. *Suppose $\mathcal{F}$ is a random indexed function set with fixed index set V. In addition, suppose $T : \mathbb{N} \times \mathbb{R}_{\geq 0}$ satisfies for any $H \subseteq V$ and $0 < \gamma < 1$,*

$$\Pr[|\mathsf{ranges}(\mathcal{F}_H)| \leq T(|H|, \gamma)] \geq 1 - \gamma.$$

Let $\mathcal{S}$ be a collection of m uniformly independent samples from $\mathcal{F}$. Then with probability at least $1 - \tau$, $\mathcal{S}$ is an α-approximation of the range space $(\mathcal{F}, \mathsf{ranges}(\mathcal{F}))$, where the randomness is taken over $\mathcal{S}, \mathcal{F}$ and

$$\alpha := \sqrt{\frac{48\left(\log(T(2m, \frac{\tau}{4})) + \log\frac{8}{\tau}\right)}{m}}.$$

In this bound, we directly use the size $|\mathsf{ranges}(\mathcal{F}_H)|$ (rather than pdim), which can provide a slightly more precise bound. [7]

Remark III.1. *There are also technical reasons for considering indexed function sets. As discussed in Section I, regarding the α-approximation, we crucially use the fact that the function set is indexed (in particular the index set is fixed), and we do not manage to prove the α-approximation lemma (Lemma III.1) for more general function sets.*

IV. Warmup: Unweighted Doubling Metrics

Let $M(X, d)$ be a doubling metric. Consider the function set $\mathcal{F} := \{f_x(\cdot) \mid x \in X\}$ indexed by X with $f_x(y) := d(x, y)$ for $y \in X$. It is well known that a bounded dimensional Euclidean space is a special case of doubling metrics, and $\mathsf{dim}(\mathcal{F}) \leq O(t)$ if M is the t-dimensional Euclidean space. However, for a general doubling metric M, $\mathsf{dim}(\mathcal{F})$ may not be bounded, as stated in the following theorem.

Theorem IV.1. *For any integer $n \geq 1$, there is a metric space $M_n(X_n, d_n)$ with $2^n + n$ unweighted points such that $\mathsf{ddim}(M_n) \leq 2$ and $\mathsf{dim}(\mathcal{F}^{M_n}) \geq n/\log n$, where $\mathcal{F}^{M_n} := \{d_n(x, \cdot) \mid x \in X_n\}$.*

Proof: We start with the definition of $M_n(X_n, d_n)$.

Define $L_n := \{u_1, u_2, \ldots, u_n\}$, $R_n := \{v_0, v_2, \ldots, v_{2^n-1}\}$. Define the point set of M_n to be $X_n := L_n \cup R_n$. For $1 \leq i \leq j \leq n$, define $d_n(u_i, u_j) := |j - i|$. For $0 \leq i \leq j \leq 2^n - 1$, define $d_n(v_i, v_j) := |j - i|$. For $u_i \in L_n$ and $v_j \in R_n$, define $d_n(v_j, u_i) := 2^{n+1} + 1$ if the i-th digit in the binary representation of j is 1, and $d_n(v_j, u_i) := 2^{n+1}$ if the i-th digit in the binary representation of j is 0. This completes the definition of M_n. It is immediate that M_n is a metric space.

Doubling Dimension. Consider a ball with center $x \in X_n$ and radius r. We distinguish the following two cases.

1) If $r < 2^{n+1}$, then either $B^{d_n}(x, r) \subseteq L_n$ or $B^{d_n}(x, r) \subseteq R_n$. Since the distance between points in L_n is induced by a 1-dimensional line, each ball $B^{d_n}(x, r) \subseteq L_n$ can be covered by at most 3 balls

[6] Our definition of the dimension is the shattering dimension of a range space, which tightly relates to the VC-dimension (see for example [41]). In particular, if $\mathsf{dim}(\mathcal{F})$ is t, then the VC-dimension of $\mathcal{F}$ is bounded by $O(t \log t)$.

[7] In Corollary V.1, we can actually show $|\mathsf{ranges}(\mathcal{F}_H)| \leq \varepsilon^{-O(\mathsf{ddim}(M))} \cdot \log\frac{1}{\tau} \cdot \mathrm{poly}(|H|)$ with probability $1 - \tau$, for a set $\mathcal{F}$ of weighted doubling distance functions. Of course we can also say $\mathsf{pdim}_\tau(\mathcal{F}) \leq O(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log 1/\tau)$, but this would lead to a slightly looser bound.

of radius $\frac{r}{2}$. This argument also holds for each ball $B^{d_n}(x, r) \subseteq R_n$.

2) If $r \geq 2^{n+1}$, $B^{d_n}(x, r)$ is a union of a subset of L_n and a subset of R_n. Then there exists $u \in L_n \cap B^{d_n}(x, r)$ and $v \in R_n \cap B^{d_n}(x, r)$. Note that L_n is covered by $B^{d_n}(u, 2^n)$ and R_n is covered by $B^{d_n}(v, 2^n)$. Hence, each ball $B^{d_n}(x, r) \subseteq R_n$ can be covered by at most 2 balls $B^{d_n}(u, \frac{r}{2})$ and $B^{d_n}(v, \frac{r}{2})$.

Therefore, $\mathsf{ddim}(M_n) \leq 2$.

Dimension of the Range Space. Let $\mathcal{D}$ be the subset of functions $\{d_n(u_i, \cdot) \mid i \in [n]\} \subseteq \mathcal{F}^{M_n}$. Consider balls $B^{d_n}(v_j, 2^{n+1})$ for $v_j \in R_n$. By definition, $|\left\{L_n \cap B^{d_n}(v_j, 2^{n+1}) \mid v_j \in R_n\right\}| = 2^n$. Note that

$$\begin{aligned}&\{f_{u_i} \in \mathcal{D} \mid u_i \in L_n \cap B^{d_n}(v_j, 2^{n+1})\} \\ &= \{f_{u_i} \in \mathcal{D} \mid f_{u_i}(v_j) = d_n(u_i, v_j) \leq 2^{n+1}\} \in \mathsf{ranges}(\mathcal{D}).\end{aligned}$$

Hence, we have $|\mathsf{ranges}(\mathcal{D})| \geq 2^n \geq |\mathcal{D}|^{n/\log n}$. Therefore, $\mathsf{dim}(\mathcal{F}^{M_n})$ is at least $n/\log n$. ■

In light of Theorem IV.1, it is impossible to bound the dimension of $\mathcal{F}$ for doubling metric M. However, we observe that, from the hard instance in the proof of Theorem IV.1, if we allow a small distortion to the distance functions (i.e., to modify all distances $2^{n+1}+1$ to 2^{n+1}), the dimension of the range space becomes bounded. Inspired by this observation, we introduce the notion of *smoothed distance functions* for doubling metrics in the next subsection. Then we prove that the range space induced by the smoothed distance functions indeed has bounded dimension in a doubling metric (see Theorem IV.2), which is the main result of this section.

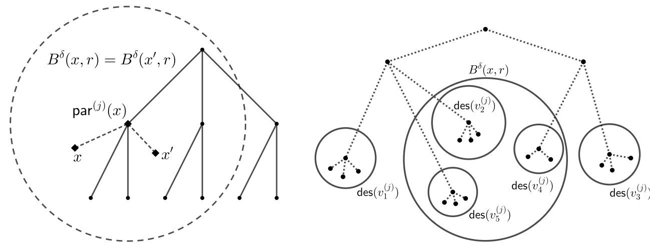
A. Smoothed Distance Functions

The smoothed distance function is defined with respect to a metric space $M(X, d)$ and a net tree T of the space (definition in Section III-A). The proofs in this section can be found in the full version.

Definition IV.1 (ε-smoothed distance function)**.** *Given a net tree T of a metric space $M(X, d)$, for $0 < \varepsilon < 1$, define $\delta_\varepsilon : X \times X \to \mathbb{R}_{\geq 0}$ as the ε-smoothed distance function induced by T as follows. For any $x, y \in X$, let $h_\varepsilon(x, y)$ be the largest integer j such that $d(\mathsf{par}^{(j)}(x), \mathsf{par}^{(j)}(y)) \geq \frac{2^j}{\varepsilon}$.* [8] *Define $j = h_\varepsilon(x, y)$ and $\delta_\varepsilon(x, y) := d(\mathsf{par}^{(j)}(x), \mathsf{par}^{(j)}(y))$.*

We assume that there is an underlying net tree T, and we drop the subscript in δ and h whenever the context is clear. Note that δ may *not* be a distance function since it may not satisfy the triangle inequality. But it satisfies the non-negativity and symmetry properties. Nonetheless, it is a close approximation of the original distance function $d(\cdot, \cdot)$, as in the following lemma.

[8]Such j must exist, because $j = -\infty$ always satisfies the condition.



(a) smooth property (b) cross-free property

Figure 1. Illustration for the smooth property and the cross-free property

Lemma IV.1 (small distortion)**.** *If T is c-covering, then for any $x, y \in X$ and any $\varepsilon > 0$,*

$$(1 - 4c \cdot \varepsilon) \cdot \delta(x, y) \leq d(x, y) \leq (1 + 4c \cdot \varepsilon) \cdot \delta(x, y).$$

Next, we show that the ε-smoothed distance function has several useful properties, which we will use extensively. The first is the descendant property, which says that if the smoothed distance of x and y is defined by two nodes u (an ancester of x) and v (an ancester of y) in layer j, then any descendant of $u^{(j)}$ has the same smoothed distance to any descendant of $v^{(j)}$.

Lemma IV.2 (descendant property)**.** *For any $x, y \in X$, assume that $j = h(x, y)$, $u = \mathsf{par}^{(j)}(x)$ and $v = \mathsf{par}^{(j)}(y)$. Then for any $x' \in \mathsf{des}(u^{(j)})$ and $y' \in \mathsf{des}(v^{(j)})$, we have $\delta(x, y) = \delta(x', y') = d(u, v)$.*

The second is the smooth property which says that at a certain distance scale r, if we move the center of the ball (of radius r) from x to x' (x' is a nearby point in a small subtree), the ball does not change, when the ball is defined w.r.t. δ.

Lemma IV.3 (smooth property, illustrated in Figure 1a)**.** *Suppose $\{N_i \mid i \leq L\}$ is a hierarchical net and T is a c-covering net tree with respect to $\{N_i\}_i$. Consider $0 < \varepsilon \leq \frac{1}{8c}$ and $r > 0$. Let $\lambda := \frac{\varepsilon \cdot (1 - 5c\varepsilon)}{20(1 + 4c\varepsilon)}$. Define j to be the integer satisfying that $2^{j-1} \leq \lambda \cdot r$. Then for any $x, x' \in X$, if $\mathsf{par}^{(j)}(x) = \mathsf{par}^{(j)}(x')$, we have $B^\delta(x, r) = B^\delta(x', r)$.*

The third is the cross-free property. Consider a ball $B^\delta(x, r)$. The next lemma says that any small subtree (with distance scale less than ϵr) is either completely contained in the ball, or does not intersect the ball at all. A consequence useful later is that each ball can be viewed as the union of some small subtrees.

Lemma IV.4 (cross-free property, illustrated in Figure 1b)**.** *Suppose $\{N_i \mid i \leq L\}$ is a hierarchical net and T is a c-covering net tree with respect to $\{N_i\}_i$. Consider $0 < \varepsilon \leq \frac{1}{8c}$ and $r > 0$. Let $\lambda := \frac{\varepsilon \cdot (1 - 5c\varepsilon)}{20(1 + 4c\varepsilon)}$. Suppose j is an integer such that $2^{j-1} \leq \lambda \cdot r$. Then for any $x \in X$ and $v \in N_j$, either $\mathsf{des}(v^{(j)}) \subseteq B^\delta(x, r)$ or $\mathsf{des}(v^{(j)}) \cap B^\delta(x, r) = \emptyset$.*

B. Bounded Dimension for Smoothed Doubling Distance Functions

In this section, we showcase the use of the smoothed distance function. In particular, we show in Theorem IV.2 that the range space induced by smoothed doubling distance functions has bounded dimension. The ε-smoothed distance function in this section is defined with respect to the *simple net tree*, which is the following natural net tree built on a hierarchical net.

Definition IV.2 (simple net trees)**.** *In a simple net tree, for each $u \in N_i$, $\mathsf{par}(u^{(i)})$ is defined to be the nearest point $v \in N_{i+1}$ to u (ties are broken arbitrarily).*

The following fact follows immediately from the definition of simple net trees.

Fact IV.1. *A simple net tree is 1-covering.*

Theorem IV.2. *Suppose $M(X, d)$ is a metric space and T is a simple net tree on X. Let $0 < \varepsilon \leq \frac{1}{8}$ be a constant. Let δ be the ε-smoothed distance function induced by T. Let $\mathcal{F} := \{\delta(x, \cdot) \mid x \in X\}$ be the function set induced by the ε-smoothed distance functions. Then $\mathsf{dim}(\mathcal{F}) \leq O(\frac{1}{\varepsilon})^{O(\mathsf{ddim}(M))}$.*

Proof: Consider any subset $H \subseteq X$ of size $|H| = m \geq 2$. It suffices to show

$$\left|\left\{H \cap B^{\delta}(x, r) \mid x \in X, r \geq 0\right\}\right| \leq m^{O(\varepsilon)^{-O(\mathsf{ddim}(M))}}.$$

Let $\lambda := \frac{\varepsilon\cdot(1-5c\varepsilon)}{20(1+4c\varepsilon)}$ as defined in Lemma IV.4. Let us first fix some $r \geq 0$ and $x \in X$. Define j to be the integer such that $2^{j-1} \leq \lambda \cdot r < 2^j$. By Lemma IV.4, $B^{\delta}(x, r)$ is the union of some $\mathsf{des}(v)$'s for v in a subset of N_j. Next, we show the number of such $\mathsf{des}(v)$'s is a constant (depending on $\mathsf{ddim}(M)$ and ε).

Let P be the set of $v \in N_j$ satisfying that $\mathsf{des}(v) \subseteq B^{\delta}(x, r)$, i.e., $B^{\delta}(x, r) = \bigcup_{v\in P} \mathsf{des}(v)$. Since $P \subseteq N_j$ is a 2^j-packing, the distance between any two points in P is at least 2^j. On the other hand, since $P \subseteq B^{\delta}(x, r)$, we have $\mathsf{diam}(P) \leq 2(1 + 4\varepsilon) \cdot r < 2^{j+2}/\lambda$. Then by packing property (Fact III.1), $|P| \leq O(\frac{1}{\lambda})^{\mathsf{ddim}(M)}$. Define $H^{(j)} := \left\{\mathsf{par}^{(j)}(x) \mid x \in H\right\}$. We have $H \cap B^{\delta}(x, r) = \bigcup_{v\in P}(H \cap \mathsf{des}(v)) = \bigcup_{v\in P\cap H^{(j)}}(H \cap \mathsf{des}(v))$. This implies every ball $B^{\delta}(x, r)$ is formed by first choosing at most $\Lambda := O(\frac{1}{\lambda})^{\mathsf{ddim}(M)}$ points $v \in H^{(j)}$, and then letting $B^{\delta}(x, r)$ be the union of these $\mathsf{des}(v)$'s.

Now we turn to general x and r. For $r \geq 0$, define

$$Q_r := \left\{\bigcup_{x\in S}\left(H \cap \mathsf{des}(\mathsf{par}^{(j)}(x))\right) \mid S \subseteq H, |S| \leq \Lambda\right\},$$

where $2^{j-1} \leq \lambda \cdot r < 2^j$. By the above argument, we know that $H \cap B^{\delta}(x, r) \in Q_r$ for any $x \in X$. Hence, $\left\{H \cap B^{\delta}(x, r) \mid x \in X, r \geq 0\right\} \subseteq \bigcup_{r\geq 0} Q_r$. Then to bound $\left|\left\{H \cap B^{\delta}(x, r) \mid x \in X, r \geq 0\right\}\right|$, it suffices to bound $|\bigcup_{r\geq 0} Q_r|$. Note that for any fixed $r \geq 0$, $|Q_r| \leq O(m^{\Lambda})$. We claim that there are at most $m + 1$ different collections Q_r for all $r \geq 0$. If the claim is true, we can bound $|\bigcup_{r\geq 0} Q_r|$ by $O((m + 1) \cdot m^{\Lambda}) = O(m^{\Lambda+2})$, and this would conclude the theorem.

It remains to prove the claim that there are at most $m + 1$ different collections Q_r for all $r \geq 0$. Observe that the cardinality of $H^{(j)}$ is non-increasing as j increases. Assume that $|H^{(i)}| = |H^{(j)}|$ for some $i \leq j$. Then for any $x, y \in H$, we have $\mathsf{par}^{(i)}(x) = \mathsf{par}^{(i)}(y)$ if and only if $\mathsf{par}^{(j)}(x) = \mathsf{par}^{(j)}(y)$.

Now fix some $x \in X$. Let $u := \mathsf{par}^{(i)}(x) \in N_j$ and $v := \mathsf{par}^{(j)}(x) \in N_j$. If $|H^{(i)}| = |H^{(j)}|$, we have for any $y \in H$, $y \in \mathsf{des}(\mathsf{par}^{(i)}(x)) \Leftrightarrow \mathsf{par}^{(i)}(y) = u \Leftrightarrow \mathsf{par}^{(j)}(y) = v \Leftrightarrow y \in \mathsf{des}(\mathsf{par}^{(j)}(x))$, which implies that $H \cap \mathsf{des}(\mathsf{par}^{(i)}(x)) = H \cap \mathsf{des}(\mathsf{par}^{(j)}(x))$. Hence, for any r', r, define i to be the integer such that $2^{i-1} \leq \lambda \cdot r' < 2^i$ and j to be the integer such that $2^{j-1} \leq \lambda \cdot r < 2^j$. If $|H^{(i)}| = |H^{(j)}|$, then $Q_r = Q_{r'}$. Since there are at most $m+1$ possible cardinalities for $|H^{(j)}|$, there are at most $m + 1$ different Q_r's. This proves the claim and thus concludes the theorem. ∎

V. Weighted Doubling Metrics

In the last section, we provide a bound of the shattering dimension by the doubling dimension. Note that the dimension bound in Theorem IV.2 is quite large in that it is exponential in $\mathsf{ddim}(M)$. However, considering the Euclidean case, the dependency is only linear. Moreover, Theorem IV.2 is not sufficient for the purpose of constructing coresets, for which we need a dimension bound for weighted spaces. In this section, we provide a new proof that can reduce the exponential dependency to a polynomial dependency, in a certain probabilistic sense. Moreover, the proof also works for weighted doubling metrics, where each point $x \in X$ is associated with a weight $w(x)$. In particular, we consider the following type of weight functions, which suffices for coreset construction.

Definition V.1. *We say $w : X \to \mathbb{R}_{\geq 0}$ is a gap-c weight function if for any $x, y \in X$, we have either $w(x) = w(y)$ or $\max\left\{\frac{w(x)}{w(y)}, \frac{w(y)}{w(x)}\right\} \geq c$.*

To achieve the polynomial dependence in $\mathsf{ddim}(M)$, we construct a *random* ε-smoothed distance function δ. Let $\mathcal{F} := \{w(x) \cdot \delta(x, \cdot) \mid x \in X\}$ be the function set induced by the random ε-smoothed distance function δ. We show that $\mathsf{pdim}_{\tau}(\mathcal{F}) \leq O\left(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log 1/\tau\right)$ in Theorem V.1[9]. The proofs can be found in the full version.

Theorem V.1. *Suppose $M(X, d)$ is a metric space together with a gap-2 weight function $w : X \to \mathbb{R}_{\geq 0}$. Let $0 < \varepsilon \leq \frac{1}{100}$ and $0 < \tau < 1$ be constant. There exists a random ε-smoothed distance function δ (defined*

[9]Actually, Theorem V.1 provides a better bound than the one with respect to pdim.

with respect to some random net tree), such that for $\mathcal{F} := \{w(x) \cdot \delta(x, \cdot) \mid x \in X\}$*, and any* $H \subseteq X$,

$$\Pr_{\delta}\left[|\mathsf{ranges}(\mathcal{F}_H)| \le O\left(\frac{1}{\varepsilon}\right)^{O(\mathsf{ddim}(M))} \cdot \log\frac{|H|}{\tau} \cdot |H|^6\right] \ge 1 - \tau.$$

In other words, $\mathsf{pdim}_\tau(\mathcal{F}) = O(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log 1/\tau)$.

Using a similar argument, we can generalize the above result to the case where the distance is taken to the power of z, as in the following corollary.

Corollary V.1. *Suppose* $M(X, d)$ *is a metric space with a gap-2 weight function* $w : X \to \mathbb{R}_{\ge 0}$*. Let* $z > 0$*,* $0 < \epsilon \le \frac{1}{100z}$ *and* $0 < \tau < 1$ *be constant. There exists a random* ϵ*-smoothed distance function* δ *(defined with respect to some random net tree), such that for* $\mathcal{F} := \{w(x) \cdot \delta^z(x, \cdot) \mid x \in X\}$ *(*$\mathcal{F}$ *is defined with respect to the* z*-th power of the random smoothed distance function* δ*), and any* $H \subseteq X$,

$$\Pr_{\delta}\left[|\mathsf{ranges}(\mathcal{F}_H)| \le O\left(\frac{1}{\epsilon}\right)^{O(\mathsf{ddim}(M))} \cdot \log\frac{|H|}{\tau} \cdot |H|^6\right] \ge 1 - \tau,$$

In addition, for $x, y \in X$*, it holds that* $(1 - O(\varepsilon \cdot z)) \cdot \delta^z(x, y) \le d^z(x, y) \le (1 + O(\varepsilon \cdot z)) \cdot \delta^z(x, y)$*. In other words,* $\mathsf{pdim}_\tau(\mathcal{F}) \le O\left(\mathsf{ddim}(M) \cdot \log(1/\varepsilon) + \log\log 1/\tau\right)$.

VI. Applications

In this section, we provide three applications of our main result. Due to the page limit, we skip all details and they can be found in the full version. The major application is an efficient ϵ-coreset construction algorithm for the (k, z)-clustering problem in doubling metrics (Section VI-A). The overall approach is to apply the Feldman-Langberg framework. As noted in Section I, one important building block is an α-approximation for the weighted range space induced by the metric space. This is done by combining the probabilistic dimension upper bound of the weighted range space (Corollary V.1), and the α-approximation lemma for the bounded probabilistic dimension (Lemma III.1). Although the construction of the smoothed distance function in Corollary V.1 is quite involved, we only use it in the analysis. The algorithm is almost as simple as in the Euclidean case. In particular, the core of the algorithm is a weighted sampling of the points in the *original* metric. In the analysis, we consider an *auxiliary* range space resulted from Corollary V.1, and we relate the sample on the original point set to a sample on the auxiliary range space. We show that the sample is a good approximation for the auxiliary space with high probability, and we translate it into a good coreset in the original space.

In Section VI-B, we discuss the robust coreset. The construction of the robust coreset is simply a *uniform* sample of points from the metric space. The key proof for the correctness is a new lemma that presents a simple (yet previously unknown) relationship between α-approximation and robust coreset. Having this lemma, we can then follow a similar argument as in the coreset construction (Section VI-A) to get a robust coreset in doubling metrics.

Another application is the construction of the centroid set and its application to accelerate the local search algorithms for the (k, z)-clustering problem in doubling metrics. The centroid set is essentially an extension of a coreset, such that a $(1 + \varepsilon)$-approximate solution to the clustering objective is *included* in the centroid set. The centroid set was first considered by [49] and was applied to a constant approximation for the geometric k-means clustering problem in Euclidean space. In a high level, our construction of the centroid set is similar with that in [49], but our construction does not rely on the specific properties in Euclidean spaces and the k-means objective. We obtain a small sized centroid set for the (k, z)-clustering problem with arbitrary k and z, and for any doubling metric. Recently, Friggstad et al. [12] showed that the local search algorithm actually gives a PTAS for the (k, z)-clustering problem in doubling metrics. For the special case of k-means in Euclidean spaces, they used the centroid set in [49] to improve the running time. However, a centroid set for doubling metrics was not known and hence the running time was not improved for more general doubling metrics. Using our new result, we obtain a similar speedup comparable to theirs in Euclidean spaces. The construction of the centroid set as well as its application is discussed in Section VI-C.

A. Coreset Construction in Doubling Metrics

Theorem VI.1 is the formal statement of Theorem I.1.

Theorem VI.1. *We are given a doubling metric* $M(X, d)$ *with* X *being a set of* n *discrete points. Let real numbers* $0 < \varepsilon, \tau < 1/100$*,* $z > 0$*, and integer* $k \ge 1$*. There exists an algorithm running in* $\mathrm{poly}(n)$ *time, that constructs a weighted subset* $S \subseteq X$ *of size*

$$\Gamma := O\left(\frac{2^{O(z\log z)}k^2\log(1/\tau)}{\varepsilon^2}\right) + (\mathsf{ddim}(M) \cdot \log(z/\varepsilon) + \log k + \log\log(1/\tau)) \cdot \left(\frac{2^{O(z\log z)}k^3}{\varepsilon^2}\right),$$

such that S *is an* ε*-coreset for the* (k, z)*-clustering problem with probability at least* $1 - \tau$.

B. Robust Coreset

We generalize and improve the prior result [8] for Euclidean space, and prove the existence of robust coresets with smaller size in doubling metrics.

Theorem VI.2. *Let $M(X,d)$ be a doubling metric space (a d-dimensional Euclidean space resp.). Suppose S is a uniform independent sample of Γ (Γ' resp.) points from X, where*

$$\Gamma := O\left(\frac{\log(1/\tau)}{\alpha^2}\right) + O\left(\frac{k}{\alpha^2}(\mathrm{ddim}(M)\cdot\log(z/\varepsilon)+\log k+\log\log(1/\tau))\right)$$

and

$$\Gamma' := O\left(\frac{1}{\alpha^2}(kd\log k+\log(1/\tau))\right).$$

Then with probability at least $1-\tau$, $\mathcal{S}$ is an (α,ε)-robust coreset ($(\alpha,0)$-robust coreset resp.) for the (k,z)-clustering problem with outliers.

C. Centroid Set and Fast Local Search Algorithm

Definition VI.1 (centroid set)**.** *Let $k\geq 1$ be an integer and $\varepsilon, z>0$. Let $M(X,d)$ be a metric space. Given a weighted point set $S\subseteq X$ with weight function $w: S\to\mathbb{R}_{\geq 0}$, an (ε,k,z)-centroid set is a subset H of points such that*

1) *$S\subseteq H\subseteq X$.*
2) *there exists a k-point set $C\subseteq H$ such that, $\sum_{x\in S} w(x)\cdot d^z(x,C)\leq(1+\varepsilon)\cdot\min_{C'\in[X]^k}\sum_{x\in S} w(x)\cdot d^z(x,C')$.*

In other words, H extends S in the sense that a $(1+\varepsilon)$-approximate solution to the weighted (k,z)-clustering instance S is contained in H. Then if S is an ε-coreset of X, we have a natural corollary that the centroid set H must contain a $(1+2\varepsilon)$-approximate solution for the (k,z)-clustering problem on X.

Theorem VI.3 (centroid set)**.** *Let $k\geq 1$ be an integer, $z>0$ and $0<\varepsilon<\frac{1}{z}$. Given a ground set X and a weighted point set $S\subseteq X$ with weight function $w: S\to\mathbb{R}_{\geq 0}$, there is an algorithm running in $\mathrm{poly}(|X|)$ time, that finds an $(O(z\cdot\varepsilon),k,z)$-centroid set of size at most $O(\frac{1}{\varepsilon})^{O(\mathsf{ddim}(M))}\cdot|S|^2$.*

Recently, Friggstad et al. [12] analyzed the local search algorithm for the (k,z)-clustering problem in doubling metrics. They improved the running time for the special of bounded dimensional Euclidean spaces using centroid sets. With the help of Theorem VI.3, it is possible to improve the running time for doubling metrics as well.

Corollary VI.1. *Let $M(X,d)$ be a (finite) metric space, and consider the (k,z)-clustering problem in M. The local search algorithm for the (k,z)-clustering problem that swaps $\rho := \mathsf{ddim}(M)^{O(\mathsf{ddim}(M))}\cdot(\frac{2^z}{\varepsilon})^{O(2^z)\cdot\mathsf{ddim}(M)\cdot\varepsilon^{-1}}$ centers (as defined in [12]) in each iteration, gives a $(1+\varepsilon)$-approximate solution after polynomial (in the input size) number of iterations. Furthermore, with a $\mathrm{poly}(|X|)$-time preprocessing procedure that succeeds with probability at least $1-\tau$, the local search algorithm runs in $(2^{O(z\log z)}\cdot\frac{k}{\varepsilon}\cdot\log\frac{1}{\tau})^{O(\rho)}$ time per iteration.*

ACKNOWLEDGMENT

We are grateful to Robert Krauthgamer for fruitful discussions and insightful comments, particularly for pointing to us the snowflake embedding literature.

REFERENCES

[1] L. Huang, S. H. Jiang, J. Li, and X. Wu, "ϵ-coresets for clustering (with outliers) in doubling metrics," *CoRR*, vol. abs/1804.02530, 2018.

[2] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Information Theory*, vol. 28, no. 2, pp. 129–136, 1982.

[3] P.-N. Tan, M. Steinbach, V. Kumar *et al.*, "Cluster analysis: basic concepts and algorithms," *Introduction to data mining*, vol. 8, pp. 487–568, 2006.

[4] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *SODA*, 2007, pp. 1027–1035.

[5] A. Coates and A. Y. Ng, "Learning feature representations with k-means," in *Neural Networks: Tricks of the Trade - Second Edition*, 2012, pp. 561–580.

[6] S. Har-Peled and S. Mazumdar, "On coresets for k-means and k-median clustering," in *STOC*, 2004, pp. 291–300.

[7] K. Chen, "On *k*-median clustering in high dimensions," in *SODA*. ACM Press, 2006, pp. 1177–1185.

[8] D. Feldman and M. Langberg, "A unified framework for approximating and clustering data," in *STOC*, 2011, pp. 569–578, https://arxiv.org/abs/1106.1379.

[9] D. Feldman, M. Schmidt, and C. Sohler, "Turning big data into tiny data: Constant-size coresets for *k*-means, PCA and projective clustering," in *SODA*, 2013, pp. 1434–1453.

[10] D. Feldman and L. J. Schulman, "Data reduction for weighted and outlier-resistant clustering," in *SODA*. SIAM, 2012, pp. 1343–1354.

[11] V. Braverman, D. Feldman, and H. Lang, "New frameworks for offline and streaming coreset constructions," *CoRR*, vol. abs/1612.00889, 2016.

[12] Z. Friggstad, M. Rezapour, and M. R. Salavatipour, "Local search yields a PTAS for k-means in doubling metrics," in *FOCS*. IEEE Computer Society, 2016, pp. 365–374.

[13] M. Lucic, M. Faulkner, A. Krause, and D. Feldman, "Training mixture models at scale via coresets," *arXiv preprint arXiv:1703.08110*, 2017.

[14] P. Assouad, "Plongements lipschitziens dans $\mathbf{R}^n$," *Bull. Soc. Math. France*, vol. 111, no. 4, pp. 429–448, 1983.

[15] A. Gupta, R. Krauthgamer, and J. R. Lee, "Bounded geometries, fractals, and low-distortion embeddings," in *FOCS*. IEEE Computer Society, 2003, pp. 534–543.

[16] J. Gao, L. J. Guibas, and A. T. Nguyen, "Deformable spanners and applications," in *Symposium on Computational Geometry*. ACM, 2004, pp. 190–199.

[17] R. Cole and L. Gottlieb, "Searching dynamic point sets in spaces with bounded doubling dimension," in *STOC*. ACM, 2006, pp. 574–583.

[18] L. Gottlieb and L. Roditty, "Improved algorithms for fully dynamic geometric spanners and geometric routing," in *SODA*. SIAM, 2008, pp. 591–600.

[19] ——, "An optimal dynamic spanner for doubling metric spaces," in *ESA*, ser. Lecture Notes in Computer Science, vol. 5193. Springer, 2008, pp. 478–489.

[20] T. H. Chan and A. Gupta, "Small hop-diameter sparse spanners for doubling metrics," *Discrete & Computational Geometry*, vol. 41, no. 1, pp. 28–44, 2009.

[21] T. H. Chan, M. Li, L. Ning, and S. Solomon, "New doubling spanners: Better and simpler," *SIAM J. Comput.*, vol. 44, no. 1, pp. 37–53, 2015.

[22] T. H. Chan, M. Li, and L. Ning, "Sparse fault-tolerant spanners for doubling metrics with bounded hop-diameter or degree," *Algorithmica*, vol. 71, no. 1, pp. 53–65, 2015.

[23] S. Solomon, "From hierarchical partitions to hierarchical covers: optimal fault-tolerant spanners for doubling metrics," in *STOC*. ACM, 2014, pp. 363–372.

[24] T. H. Chan, A. Gupta, B. M. Maggs, and S. Zhou, "On hierarchical routing in doubling metrics," *ACM Trans. Algorithms*, vol. 12, no. 4, pp. 55:1–55:22, 2016.

[25] I. Abraham, Y. Bartal, and O. Neiman, "Advances in metric embedding theory," in *STOC*. ACM, 2006, pp. 271–286.

[26] T. H. Chan, A. Gupta, and K. Talwar, "Ultra-low-dimensional embeddings for doubling metrics," *J. ACM*, vol. 57, no. 4, pp. 21:1–21:26, 2010.

[27] K. L. Clarkson, "Nearest neighbor queries in metric spaces," *Discrete Comput. Geom.*, vol. 22, no. 1, pp. 63–93, 1999.

[28] P. Indyk and A. Naor, "Nearest neighbor preserving embeddings," *ACM Transactions on Algorithms*, To appear.

[29] S. Har-Peled and M. Mendel, "Fast construction of nets in low-dimensional metrics and their applications," *SIAM J. Comput.*, vol. 35, no. 5, pp. 1148–1184, 2006.

[30] K. Talwar, "Bypassing the embedding: algorithms for low dimensional metrics," in *STOC*. ACM, 2004, pp. 281–290.

[31] Y. Bartal, L. Gottlieb, and R. Krauthgamer, "The traveling salesman problem: Low-dimensionality implies a polynomial time approximation scheme," *SIAM J. Comput.*, vol. 45, no. 4, pp. 1563–1581, 2016.

[32] T. H. Chan and K. M. Elbassioni, "A QPTAS for TSP with fat weakly disjoint neighborhoods in doubling metrics," *Discrete & Computational Geometry*, vol. 46, no. 4, pp. 704–723, 2011.

[33] T. H. Chan and S. H. Jiang, "Reducing curse of dimensionality: Improved PTAS for TSP (with neighborhoods) in doubling metrics," *ACM Trans. Algorithms*, vol. 14, no. 1, pp. 9:1–9:18, 2018.

[34] T. H. Chan, S. Hu, and S. H. Jiang, "A PTAS for the steiner forest problem in doubling metrics," in *FOCS*. IEEE Computer Society, 2016, pp. 810–819.

[35] N. H. Bshouty, Y. Li, and P. M. Long, "Using the doubling dimension to analyze the generalization of learning algorithms," *J. Comput. Syst. Sci.*, vol. 75, no. 6, pp. 323–335, 2009.

[36] L. Gottlieb, A. Kontorovich, and R. Krauthgamer, "Efficient classification for metric data," *IEEE Trans. Information Theory*, vol. 60, no. 9, pp. 5750–5759, 2014.

[37] M. Langberg and L. J. Schulman, "Universal ϵ-approximators for integrals," in *SODA*, 2010, pp. 598–607.

[38] K. Varadarajan and X. Xiao, "On the sensitivity of shape fitting problems," in *FSTTCS*, 2012, pp. 486–497.

[39] N. Alon, S. Dar, M. Parnas, and D. Ron, "Testing of clustering," *SIAM Journal on Discrete Mathematics*, vol. 16, no. 3, pp. 393–417, 2003.

[40] Y. Li, P. M. Long, and A. Srinivasan, "Improved bounds on the sample complexity of learning," *J. Comput. Syst. Sci.*, vol. 62, no. 3, pp. 516–527, 2001.

[41] M. J. Kearns and U. V. Vazirani, *An introduction to computational learning theory*. MIT press, 1994.

[42] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, "Approximating extent measures of points," *Journal of the ACM (JACM)*, vol. 51, no. 4, pp. 606–635, 2004.

[43] P. K. Agarwal and C. M. Procopiuc, "Exact and approximation algorithms for clustering," *Algorithmica*, vol. 33, no. 2, pp. 201–226, 2002.

[44] S. Har-Peled, "Clustering motion," *Discrete & Computational Geometry*, vol. 31, no. 4, pp. 545–565, 2004.

[45] S. Har-Peled and A. Kushal, "Smaller coresets for k-median and k-means clustering," *Discrete & Computational Geometry*, vol. 37, no. 1, pp. 3–19, 2007.

[46] J. M. Phillips, "Coresets and sketches," *CoRR*, vol. abs/1601.00617, 2016.

[47] M. Balcan, S. Ehrlich, and Y. Liang, "Distributed k-means and k-median clustering on general communication topologies," in *NIPS*, 2013, pp. 1995–2003.

[48] V. Braverman, G. Frahling, H. Lang, C. Sohler, and L. F. Yang, "Clustering high dimensional dynamic data streams," in *ICML*, ser. Proceedings of Machine Learning Research, vol. 70. PMLR, 2017, pp. 576–585.

[49] J. Matoušek, "On approximate geometric k-clustering," *Discrete & Computational Geometry*, vol. 24, no. 1, pp. 61–84, 2000.