


Dimension-Free Parameterized Approximation Schemes for Hybrid Clustering

Ameet Gadekar ✉ 

CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Tanmay Inamdar ✉ 

Indian Institute of Technology Jodhpur, India

Abstract

HYBRID k -CLUSTERING is a model of clustering that generalizes two of the most widely studied clustering objectives: k -CENTER and k -MEDIAN. In this model, given a set of n points P , the goal is to find k centers such that the sum of the r -distances of each point to its nearest center is minimized. The r -distance between two points p and q is defined as $\max\{\text{dist}(p, q) - r, 0\}$ – this represents the distance of p to the boundary of the r -radius ball around q if p is outside the ball, and 0 otherwise. This problem was recently introduced by Fomin et al. [APPROX 2024], who designed a $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria approximation that runs in time $2^{(kd/\varepsilon)^{O(1)}} \cdot n^{O(1)}$ for inputs in \mathbb{R}^d ; such a bicriteria solution uses balls of radius $(1 + \varepsilon)r$ instead of r , and has a cost at most $1 + \varepsilon$ times the cost of an optimal solution using balls of radius r .

In this paper we significantly improve upon this result by designing an approximation algorithm with the same bicriteria guarantee, but with running time that is FPT only in k and ε – crucially, removing the exponential dependence on the dimension d . This resolves an open question posed in their paper. Our results extend further in several directions. First, our approximation scheme works in a broader class of metric spaces, including doubling spaces, minor-free, and bounded treewidth metrics. Secondly, our techniques yield a similar bicriteria FPT-approximation schemes for other variants of HYBRID k -CLUSTERING, e.g., when the objective features the sum of z -th power of the r -distances. Finally, we also design a coresset for HYBRID k -CLUSTERING in doubling spaces, answering another open question from the work of Fomin et al.

2012 ACM Subject Classification Theory of computation → Facility location and clustering; Theory of computation → Fixed parameter tractability

Keywords and phrases Clustering, Parameterized algorithms, FPT approximation, k -Median, k -Center

Digital Object Identifier 10.4230/LIPIcs.STACS.2025.35

Related Version Full Version: <https://arxiv.org/abs/2501.03663>

Funding *Ameet Gadekar*: Partially supported by the Israel Science Foundation (grant No. 1042/22). *Tanmay Inamdar*: Supported by IITJ Research Initiation Grant (grant number I/RIG/T-NI/20240072).

Acknowledgements Tanmay would like to thank his co-authors from [23] – specifically Fedor V. Fomin – for formulating Hybrid k -Clustering problem and introducing him to it. This work was partially carried out while Ameet was at Bar-Ilan University, Israel.

1 Introduction

k -CENTER, k -MEDIAN, and k -MEANS are among the most popular clustering problems both in theory and practice, with numerous applications in areas such as machine learning [36, 28, 6, 11, 4, 9, 38, 25], facility location problems [32, 41, 30, 29, 35, 16], and computational



© Ameet Gadekar and Tanmay Inamdar;

licensed under Creative Commons License CC-BY 4.0

42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025).

Editors: Olaf Beyersdorff, Michał Pilipczuk, Elaine Pimentel, and Nguyễn Kim Thăng;

Article No. 35; pp. 35:1–35:20



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



geometry [7, 34], among others.¹ All of these problems have long been known to be NP-hard, even in the plane [22, 40, 37]. To cope with these intractability results, there has been several decades of research on designing approximation algorithms for these problems, that have lead to polynomial-time constant-factor approximation algorithms for all of them in general metric spaces [27, 20]. Moreover, for more “structured” inputs – such as when the points belong to Euclidean spaces, or planar graph metrics – one can find near-optimal solutions using approximation schemes [24, 17, 19, 5, 34, 7, 2].² Furthermore, given the simplicity and ubiquity of these vanilla clustering objectives, the techniques used for obtaining these results are subsequently used to find clustering with additional constraints, such as capacities [18], fairness [8], and outliers [31, 3].¹ One recurring theme in the literature has been to define a unified clustering objective that also captures classical clustering problems like k -MEDIAN, k -CENTER as special cases [42, 12, 15].

Along this line of research, Fomin et al. [23] recently introduced HYBRID k -CLUSTERING. In this problem, we are given an instance $\mathcal{I} = (M = (P, \mathbb{F}, \text{dist}), k, r)$, where $M = (P, \mathbb{F}, \text{dist})$ is a metric space³ on the set of clients P and a set of facilities \mathbb{F} with distance function dist . Here, k is a positive integer denoting the number of clusters, and r is a non-negative real denoting the radius. The goal is to find a set $X \subseteq \mathbb{F}$ of k centers, such that, $\text{cost}_r(P, X) := \sum_{p \in P} \text{dist}_r(p, X)$ is minimized – here, for any point p , any set $Q \subseteq P \cup \mathbb{F}$, and a real α , $\text{dist}_\alpha(p, Q)$ is defined as $\text{dist}_\alpha(p, Q) := \max\{\text{dist}(p, Q) - \alpha, 0\}$, and $\text{dist}(p, Q) := \min_{q \in Q} \text{dist}(p, q)$. Fomin et al. [23] proposed several motivations for studying this cost function. Indeed, HYBRID k -CLUSTERING can be seen as a *shape fitting* problem, where one wants to find the “best” set of k balls of radius r that fit the given set of points, where the quality of the solution is determined by the sum of distances of each point to the nearest point on the boundary of the ball – this is analogous to the classical regression, where one wants to find the “best” linear function fitting the given set of points. PROJECTIVE CLUSTERING is a well-studied generalization of linear regression, where the aim is to find k affine spaces that minimizes the sum of distances from the points to these spaces (see e.g., [43]), which is also closely related to the model considered in HYBRID k -CLUSTERING. Furthermore, [23] gave another motivation for studying HYBRID k -CLUSTERING – namely, placing k WiFi routers with identical circular coverage, where the clients that lie outside the coverage area need to travel to the boundary of the nearest ball in order to receive coverage. Finally, the name of the problem is motivated from the fact that the objective is a kind of “hybrid” between the k -CENTER and k -MEDIAN costs, and generalizes both of them. Indeed, as observed in [23], HYBRID k -CLUSTERING with $r = 0$ is equivalent to k -MEDIAN, while, when r is set to be the optimal radius for k -CENTER, HYBRID k -CLUSTERING reduces to k -CENTER. These observations immediately rule out *uni-criteria* polynomial-time approximation schemes that violate only the cost, or only the radius by any arbitrary factor $\alpha > 1$ (as formalized in Proposition 1 of [23]). Indeed, such approximation algorithms – even with running times FPT in k – would imply exact FPT algorithms for k -CENTER and k -MEDIAN in low-dimensional continuous Euclidean spaces. To our knowledge, such an algorithm for k -MEDIAN is not known (nor is admittedly a lower bound), whereas [39] shows W[1]-hardness of k -CENTER even in \mathbb{R}^2 . Therefore, given the current state of the art, a *bicriteria* approximation scheme for HYBRID k -CLUSTERING is essentially the best outcome, even if one allows a running time that is FPT in k .

¹ These citations are supposed not to be comprehensive, but representative – an interested reader may use them as a starting point for the relevant literature.

² This includes polynomial-time as well as FPT approximation schemes.

³ Fomin et al. [23] only considered the special case of Euclidean inputs, i.e., when $P \subset \mathbb{F} = \mathbb{R}^d$; however, the underlying problem can be defined for arbitrary metric spaces.

For $\alpha, \beta \geq 1$, an (α, β) -*bicriteria approximation* for HYBRID k -CLUSTERING is an algorithm that returns a solution $X \subseteq \mathbb{F}$ of size at most k satisfying $\text{cost}_{\beta r}(P, X) \leq \alpha \cdot \text{OPT}_r$. Note that the bicriteria solution X is allowed to consider $\text{dist}_{\beta r}(p, X) = \max\{\text{dist}(p, X) - \beta r, 0\}$ instead of $\text{dist}_r(p, X)$ and is also allowed to find a solution of cost αOPT_r , where OPT_r is the cost of an optimal solution w.r.t. radius r , i.e., without any violation of the radius. The main result of Fomin et al. [23] was an $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria approximation for inputs in \mathbb{R}^d in time $2^{(kd/\varepsilon)^{O(1)}} \cdot n^{O(1)}$, where $n = |P|$. An exponential dependence on the dimension d appears inevitable using their approach, although it was not clear whether such a dependence is required. Indeed, for Euclidean k -MEDIAN and k -CENTER, one can obtain *dimension-free* approximation schemes with FPT dependence only on k and ε [7, 33, 1]. This naturally motivates the following question, which was explicitly asked as an open question in [23].

Question 1. “An immediate question is whether improving or removing the FPT dependence on the dimension d is possible[...]”

In this work, we answer this question in the affirmative by designing a (randomized) bicriteria FPT Approximation Scheme (FPT-AS) parameterized by k and ε^4 for the (continuous) Euclidean instances of HYBRID k -CLUSTERING, stated formally in the following theorem.

► **Theorem 1** (Bicriteria FPT-AS for Euclidean Spaces). *There exists a randomized algorithm, that, given an instance of HYBRID k -CLUSTERING in \mathbb{R}^d for any dimension d , runs in time $2^{O(k \log k \cdot (1/\varepsilon^5) \log^2(1/\varepsilon))} \cdot n^{O(1)}$, and returns a $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria approximation with high probability.*

The algorithm of [23] involves a pre-processing step that separately handles “ k -MEDIAN-like”, “ k -CENTER-like” instances, using the techniques specific to these respective problems. Then, the main algorithm handles the remaining instances that cannot be directly reduced to the respective problems. This approach somewhat undermines the goal of defining a unified problem that captures both of these problems as special cases. In contrast, our algorithm adopts a uniform approach by exploiting the intrinsic structure of the hybrid problem, without the need to separately handle some instances. In fact, our algorithm works for a broader class of metric spaces, called metric spaces of *bounded algorithmic scatter dimension* – a notion recently introduced by Abbasi et al. [1], and further studied by Bourneuf and Pilipczuk [10]. These metric spaces capture several interesting and well-studied metric spaces, see below. We give a formal definition of the notion of algorithmic scatter dimension in Section 2; however, a good mental picture to keep is to think of them as essentially doubling spaces, i.e., metric spaces with good “packing-covering” properties.⁵ Our general result is stated below.

► **Theorem 2** (Informal version of Theorem 9). *HYBRID k -CLUSTERING admits a randomized bicriteria FPT-AS in metrics of bounded algorithmic ε -scatter dimension, parameterized by k and ε . In particular, HYBRID k -CLUSTERING admits randomized bicriteria FPT-AS parameterized by k and ε , in continuous Euclidean spaces of any dimension, metrics of bounded doubling dimension, bounded treewidth metrics, and by graphs from any fixed proper minor-closed graph class.*

⁴ Such algorithms run in time $f(k, \varepsilon)n^{O(1)}$ and output a $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria solution. Another term for FPT-AS is EPAS, which stands for Efficient Parameterized Approximation Schemes.

⁵ Although this is a good mental picture, it is ultimately inaccurate since the class of metric spaces of bounded algorithmic scatter dimension is strictly larger than that of doubling spaces. Indeed, continuous Euclidean space of high $(\omega(\log n))$ dimension does not have bounded doubling dimension, yet it does have bounded algorithmic scatter dimension.

We give a technical overview of this result in Section 1.1, and describe the approximation algorithm and its analysis in Section 3.

In their work, [23] also defined HYBRID (k, z) -CLUSTERING problem, which features the z -th power of r -distances, i.e., the objective is to minimize $\text{cost}_r(P, X, z) := \sum_{p \in P} (\text{dist}_r(p, X))^z$, where $z \geq 1$. They designed a bicriteria FPT-AS with similar running time for HYBRID $(k, 2)$ -CLUSTERING, i.e., a hybrid of k -MEANS and k -CENTER; but left the possibility of obtaining a similar result for the general case of $z \geq 1$ conditional on the existence of a certain sampling-based approximation algorithm for the (k, z) -clustering problem for Euclidean inputs. Using our approach, we can obtain a bicriteria FPT-AS for any fixed $z \geq 1$ in a unified way, whose running time is independent of the dimension d . In fact, our approach works for a much more general problem of HYBRID NORM k -CLUSTERING. We discuss these extensions in Section 4.

Next, we turn to another open direction mentioned in the work of Fomin et al. [23].

Question 2. “Another intriguing question is the design of coresets for HYBRID k -CLUSTERING, which could also have some implications for [Question 1].”

In this paper, we also answer this question affirmatively by designing coresets for HYBRID k -CLUSTERING in metric spaces of bounded doubling dimension. Specifically, we prove the following result.

► **Theorem 3 (Coreset for HYBRID k -CLUSTERING).** *There exists an algorithm that takes as input an instance $\mathcal{I} = ((P, \mathbb{F}, \text{dist}), k, r)$ of HYBRID k -CLUSTERING in doubling metric of dimension d and a parameter $\varepsilon \in (0, 1)$, and in time $2^{O(d \log(1/\varepsilon))} |\mathcal{I}|^{O(1)}$ returns a pair (P', w) , where $P' \subseteq P$ has size $2^{O(d \log(1/\varepsilon))} k \log |P|$, and $w : P' \rightarrow \mathbb{N}$, such that the following property is satisfied for any $X \subseteq \mathbb{F}$ of size at most k :*

$$|\text{wcost}_r(P', X) - \text{cost}_r(P, X)| \leq \varepsilon \text{cost}_r(P, X)$$

Here, $\text{wcost}_r(P, X) := \sum_{p \in P'} w(p) \cdot \text{dist}_r(p, X)$.

1.1 Technical Overview

In this section, we highlight our technical and conceptual contributions of our main result (Theorem 2).

A natural direction to design dimension-free parameterized approximation algorithm for HYBRID k -CLUSTERING is to leverage insights from the framework of Abbasi et al. [1], who designed dimension-free parameterized approximations for wide range of clustering problems. However, in HYBRID k -CLUSTERING, the objective is a function of r -distances, instead of true distances, as in [1]. This introduces several technical challenges. For instance, a most obvious roadblock is that the r -distances do not satisfy triangle inequality. Moreover, many of the “nice” properties that are enjoyed by the true distances in “structured” metric spaces (e.g., covering-packing property in doubling spaces, or properties of shortest paths in sparse graphs) do not extend readily to the r -distances. Therefore, we have to overcome several technical and conceptual challenges in order to leverage the ideas presented in the framework of [1]. Let us first recall the key notion from this paper that is crucially used in this framework.⁶

⁶ In fact, the algorithm of our paper and [1] works for a weaker notion called algorithmic scatter dimension. But, for ease of exposition, we work with scatter dimension in this section.

Scatter Dimension. Informally, an ε -scattering sequence in a metric space $M = (P, \mathbb{F}, \text{dist})$ is a sequence of center-point pairs $(x_1, p_1), \dots, (x_\ell, p_\ell)$, where ℓ is some positive integer and for $j \in [\ell], x_j \in \mathbb{F}, p_j \in P$ such that $\text{dist}(p_j, x_i) \leq 1$, for all $1 \leq j < i \leq [\ell]$ and $\text{dist}(p_i, x_i) > (1 + \varepsilon)$, for all $i \in [\ell]$. The ε -scatter dimension of M is the maximum length of ε -scattering sequence contained in M .

Now, consider a natural extension of the framework [1] for HYBRID k -CLUSTERING in a metric space M as follows. Let OPT_r be the optimal cost of the HYBRID k -CLUSTERING instance corresponding to an optimal solution O . The algorithm maintains cluster constraint Q_i for each cluster $i \in [k]$. Each Q_i consists of a collection of requests of the form (p, δ) , where p is a point and δ is a distance, representing the demand that p requires a center within distance δ . The algorithm always maintains a solution X such that $x_i \in X$ satisfy Q_i cluster constraint, for all $i \in [k]$. Now consider the sequence of triples $S_i = (x_i^{(1)}, p_i^{(1)}, \delta_i^{(1)}), \dots, (x_i^{(\ell)}, p_i^{(\ell)}, \delta_i^{(\ell)})$ corresponding to requests in Q_i , where $x_i^{(j)}$ is the i^{th} center maintained by the algorithm just before adding request $(p_i^{(j)}, \delta_i^{(j)})$ to Q_i . If X is a near-optimal solution, then the algorithm terminates successfully and returns X . Otherwise, it identifies a point $p \in P$ whose r -distance to X is much larger than its r -distance to O . Such a point is called a *witness* to X . A simple averaging argument shows that such a witness can be sampled with high probability. The algorithm then guesses the optimal cluster $i \in [k]$ of p and add a new request $(p, \delta_p = \text{dist}(p, X)/(1 + \varepsilon'))$ to Q_i , for some suitable but fixed ε' depending on ε . The center x_i is recomputed to satisfy the updated cluster constraint Q_i , if possible. Otherwise the algorithm reports failure.⁷ The key observation is that the sequence of pairs $(x_i^{(j_1)}, p_i^{(j_1)}), \dots, (x_i^{(j_\ell)}, p_i^{(j_\ell)})$ for a fixed radius $\delta_i^{(j)}$ forms an ε -scattering sequence in M . Thus, if the ε -scatter dimension of M is bounded then the length of these sequences are also bounded. Furthermore, it can be shown that if the aspect ratio of the radii of requests in Q_i for all $i \in [k]$ is bounded, then the number of iterations of the algorithm can be bounded, yielding an FPT-AS.

Working with the inflated radius. A major challenge that arises when the witness is defined in terms of r -distances, but the requests added to the cluster constraints are based on the true distances. Specifically, we need to ensure that a witness whose r -distance to X is significantly larger than its r -distance to O also has a larger true distance to X than to O . This ensures that the request $(p, \text{dist}(p, X)/(1 + \varepsilon'))$ is feasible and the algorithm does not fail. However, this condition may not hold, especially when the true distances are close to r . In fact, the issue is related to a fundamental barrier identified by [23], assuming standard conjectures.⁸ To overcome this fundamental barrier and maintain a sufficient gap between the true distances, we consider the cost of X with respect to $(1 + \varepsilon)r$ radius. In other words, we look for a solution X whose $(1 + \varepsilon)r$ -cost is close to OPT_r . In this case, we redefine a witness as a point in P whose $(1 + \varepsilon)r$ -distance to X is much larger than its r -distance to O . These insights allow us to establish a gap between the true distance of a witness to X and its true distance to O , thereby justifying the requests added by the algorithm.

Bounding the aspect ratio of the radii. Abbasi et al. [1] bound the aspect ratio of the radii by (i) initializing the solution using “good” upper bounds, and (ii) sampling witness from the points that have comparable distance to X with respect to the corresponding upper bounds.

⁷ E.g., if the algorithm failed to sample a witness.

⁸ Such a framework could potentially yield a near-optimal solution that does not violate the radius, contradicting Proposition 1 in [23].

The first step guarantees that the solution X maintained by the algorithm always satisfies the upper bounds within constant factor. This together with the second step allows one to bound the aspect ratio of radii in each Q_i . They show that such witnesses have a good probability mass if the points are sampled proportional to their true distances, and hence the algorithm successfully finds a witness for X with probability that is a function of k and ε . Although, devising feasible and “good” upper bounds for HYBRID k -CLUSTERING can be done with some efforts, the main challenge arises in the second step, which only guarantees to sample a point whose $(1 + \varepsilon)r$ -distance to X is comparable with its upper bound. As mentioned before, these $(1 + \varepsilon)r$ -distances can be much smaller than the corresponding true distances, and hence there is no guarantee that true distance of a witness p to current solution X is comparable to its upper bound. Towards this, we introduce a novel idea of dividing witnesses into two sets – “nearby witness” set, i.e., the set of points within distance $O(r/\varepsilon)$ from X , and “faraway witness” set, which are beyond this distance from X . The observation is that, since the total probability mass on the witness set, now defined using $(1 + \varepsilon)r$ -distances, is still high, it must be that either the nearby or the faraway witnesses have sufficient probability mass. However, since we do not know which of the two sets has enough mass, we perform a randomized branching (i.e., make a guess) – which will be “correct” with probability $1/2$. Then, when a point is sampled proportional to its $(1 + \varepsilon)r$ -distance from either the “nearby” or “faraway” set, it will be a witness with good probability. Now consider each of two cases separately. Since for a nearby witness p , its true distance to X is at least $(1 + \varepsilon)r$ and at most $O(r/\varepsilon)$, the aspect ratio of the radii of requests corresponding to nearby witness set is bounded by $O(1/\varepsilon)$. On the other hand, for requests corresponding to faraway witness set, we show that their radii lie in bounded interval, using ideas similar to [1]. Note that, these two arguments imply that the radii of the requests lie in two (possibly disjoint) intervals that themselves are bounded. However, it is not clear if the length of these requests is bounded, unlike [1], where the radii belonged to a single interval of bounded aspect ratio. Nevertheless, we observe that, using the techniques of [1], the length of requests can be bounded, even when the radii lie in constantly many (here, two) intervals, each with a bounded aspect ratio.

2 Background on Algorithmic Scatter Dimension

In this paper, we consider metric (clustering) space $M = (P, \mathbb{F}, \text{dist})$, where P is a finite set of n points, \mathbb{F} is a (possibly infinite) set of potential cluster centers, and dist is a metric on $(P \cup \mathbb{F})$. A class \mathcal{M} of metric spaces is a (possibly infinite) set of metric spaces.

In this paper, we work with a notion that is weaker (and hence more general) than ε -scatter dimension that was defined in the overview, called *algorithmic ε -scatter dimension*, which we explain next. To this end, we first need the following search problem.

► **Definition 4** (BALL INTERSECTION Problem). *Let \mathcal{M} be a class of metric spaces with possibly infinite set of centers. Given $M = (P, \mathbb{F}, \text{dist}) \in \mathcal{M}$, a finite set $Q \subseteq P \times \mathbb{R}_+$ of distance constraints, and an error parameter $\eta > 0$, the Ball Intersection problem asks to find a center $x \in \mathbb{F}$ that satisfy all distance constraints within η multiplicative error, i.e., $\text{dist}(x, p) \leq (1 + \eta)\delta$, for every $(p, \delta) \in Q$, if such a center exists, and report failure otherwise.*

We say \mathcal{M} admits a BALL INTERSECTION algorithm if it correctly solves the ball intersection problem for every metric space in \mathcal{M} and runs in polynomial time in the size of M and $1/\eta$.

Now, we are ready to define algorithmic scatter dimension.

► **Definition 5** (Algorithmic ε -Scatter Dimension). *Given a class \mathcal{M} of metric spaces with BALL INTERSECTION algorithm $\mathcal{C}_{\mathcal{M}}$, a space $M \in \mathcal{M}$, and $\varepsilon \in (0, 1)$, a $(\mathcal{C}_{\mathcal{M}}, \varepsilon)$ -scattering sequence is a sequence $(x_1, p_1, \delta_1), \dots, (x_\ell, p_\ell, \delta_\ell)$, where ℓ is some positive integer, and for $i \in [\ell]$, $x_i \in \mathbb{F}$, $p_i \in P$ and $\delta_i \in \mathbb{R}_+$ such that*

- (Covering by $\mathcal{C}_{\mathcal{M}}$) $x_i = \mathcal{C}_{\mathcal{M}}(M, \{(p_1, \delta_1), \dots, (p_{i-1}, \delta_{i-1})\}, \varepsilon/2) \quad \forall 2 \leq i \leq \ell$
- (ε -refutation) $\text{dist}(x_i, p_i) > (1 + \varepsilon)\delta_i \quad \forall i \in [\ell]$

The algorithmic $(\varepsilon, \mathcal{C}_{\mathcal{M}})$ -scatter dimension of \mathcal{M} is $\lambda_{\mathcal{M}}(\varepsilon)$ if any $(\mathcal{C}_{\mathcal{M}}, \varepsilon)$ -scattering sequence contains at most $\lambda_{\mathcal{M}}(\varepsilon)$ many triples per radius value. The algorithmic ε -scatter dimension of \mathcal{M} is the minimum algorithmic $(\varepsilon, \mathcal{C}_{\mathcal{M}})$ -scatter dimension over any BALL INTERSECTION algorithm $\mathcal{C}_{\mathcal{M}}$ for \mathcal{M} .

Although, algorithmic scatter dimension restricts the number of triples in the sequence with same radius value, we can use the proof technique from [1] to prove the following stronger guarantee. Results marked with ♠ can be found in the full version of the paper.

► **Lemma 6** (♠). *Let \mathcal{M} be a class of metric spaces of algorithmic ε -scatter dimension $\lambda(\varepsilon)$. Then there exists a BALL INTERSECTION algorithm $\mathcal{C}_{\mathcal{M}}$ with the following property. Given $\varepsilon \in (0, 1)$, a constant $t \geq 1$, and $a_i > 0, \tau_i \geq 2$ for $i \in [t]$, any $(\mathcal{C}_{\mathcal{M}}, \varepsilon)$ -scattering contains $O(\sum_{i \in [t]} \lambda(\varepsilon/2)(\log \tau_i)/\varepsilon)$ many triples whose radii lie in the interval $\cup_{i \in [t]} [a_i, \tau_i a_i]$.*

3 Bicriteria FPT Approximation Scheme

3.1 Algorithm

Our bicriteria FPT-AS for HYBRID k -CLUSTERING is formally stated in Algorithm 1. As an input, we are given an instance $\mathcal{I} = ((P, \mathbb{F}, \text{dist}), k, r)$ of HYBRID k -CLUSTERING, an accuracy parameter ε , access to an algorithm \mathcal{C} for the so-called “BALL INTERSECTION” problem (discussed later), and a guess \mathcal{G} for the optimal cost OPT_r . By a standard exponential search, we will assume that $\text{OPT}_r \leq \mathcal{G} \leq (1 + \varepsilon/3) \cdot \text{OPT}_r$. At a high level, this algorithm can be divided into two steps: *initialization phase* and the *iterative cost-improvement phase*. The initialization phase spans from line 1 to 7.

At a high level, the goal of this phase to compute for each point $p \in P$, an *upper bound* $u(p)$, such that p must have an optimal center within distance $u(p)$. Once we find such upper bounds, we use a subset of them to initialize for each $i \in [k]$, a set of *requests* Q_i , where each request (p, δ_p) demands that the i th center in the solution must be within distance at most δ_p from p in every subsequent solution found by the algorithm.

Now, the algorithm moves to the iterative cost-improvement phase in lines 8 to 20, consisting of a while loop that runs as long as the current solution has not become the bicriteria approximation that we are looking for. Thus, in each iteration of the while loop, we are given a solution X that satisfies all the requests Q_i , and yet satisfies $\text{cost}_{r'}(P, X) > (1 + \varepsilon) \cdot \mathcal{G}$. Then, our algorithm makes a random choice whether there is enough cost-contribution of nearby witnesses, or of faraway witnesses – here a witness is a point p whose distance to X is sufficiently larger than that in the (unknown) optimal solution O . In each of the cases, we sample points from carefully defined sets (cf. N in line 11 and A in 14), proportional to their contribution to the cost to the respective sets (in the analysis, we will argue that with good probability, we will in fact sample witness points). Having sampled such a point p , we guess the index i of the optimal cluster of p in line 17. Finally, assuming p is indeed a witness, we add a request $(p, \frac{\text{dist}(p, X)}{1 + \varepsilon/12})$ to the i th request set Q_i , and in line 19 we recompute x_i using \mathcal{C} . This algorithm either returns a center $x_i \in \mathbb{F}$ that satisfies all the requests (with an error of up to a factor of $\varepsilon/40$), or correctly outputs that there is no point in \mathbb{F} satisfying

■ **Algorithm 1** Approximation Scheme for HYBRID k -CLUSTERING.

Input: Instance $\mathcal{I} = ((P, \mathbb{F}, \text{dist}), k, r)$ of HYBRID k -CLUSTERING, $\varepsilon \in (0, 1)$, and BALL INTERSECTION algorithm \mathcal{C} , and a guess \mathcal{G} for OPT_r

Output: A solution $X \subseteq \mathbb{F}$ such that $\text{cost}_{(1+\varepsilon)r}(P, X) \leq (1 + \varepsilon)\mathcal{G}$, assuming $\text{OPT}_r \leq \mathcal{G} \leq (1 + \varepsilon/3) \cdot \text{OPT}_r$.

- 1: For each $p \in P$, compute $u(p) = 3 \cdot \min\{\alpha > r : |\text{ball}(p, \alpha)| \geq \mathcal{G}/\alpha\}$
- 2: Process P in non-decreasing order of $u(p)$ and mark $p_i \in P$ if $\text{ball}(p_i, u(p_i))$ is disjoint from $\text{ball}(p_j, u(p_j))$ for every marked p_j such that $j < i$
- 3: Let $p^{(1)}, \dots, p^{(k')}$ be the marked points
- 4: **for each** $i \in [k']$, let $Q_i := \{(p^{(i)}, u(p^{(i)}))\}$
- 5: For $k \geq i > k'$, let $Q_i = \emptyset$
- 6: Let $X := (x_1, \dots, x_k)$, where $\forall i \in [k]$, $x_i \in \mathbb{F}$ is any center satisfying requests in Q_i .
- 7: Let $r' := r(1 + \varepsilon/3)$.
- 8: **while** $\text{cost}_{r'}(P, X) > (1 + \varepsilon) \cdot \mathcal{G}$ **do**
- 9: Toss a fair coin to guess whether we are in “nearby witness” or “faraway witness” case
- 10: **if** we guess “nearby witness” case **then**
- 11: $N = \{p \in P : \text{dist}(p, X) \leq \frac{8r}{\varepsilon}\}$
- 12: Sample a point $p \in N$, where $\Pr(p = a) = \frac{\text{dist}_{r'}(a, X_i)}{\sum_{b \in N} \text{dist}_{r'}(b, X_i)}$ for each $a \in P$
- 13: **else** we guess “faraway witness case” **then**
- 14: Let $A := \{p \in P : \text{dist}_{r'}(p, X) > \frac{\varepsilon}{1000k} \cdot u(p)\}$
- 15: Sample a point $p \in A$, where $\Pr(p = a) = \frac{\text{dist}_{r'}(a, X_i)}{\sum_{b \in P} \text{dist}_{r'}(b, X_i)}$ for each $a \in A$
- 16: **end if**
- 17: Sample an integer $i \in [k]$ u.a.r.
- 18: Add (p, δ_p) to Q_i , where $\delta_p = \frac{\text{dist}(p, X)}{1 + \varepsilon/12}$
- 19: $x_i \leftarrow \mathcal{C}(Q_i, \mathbb{F}, \varepsilon/40)$ **if** no x_i was found then **fail**
- 20: **end while**

all the requests. Note that in discrete metric spaces, \mathcal{C} can be simulated by simply scanning each $x \in \mathbb{F}$ and checking whether it satisfies all the requests in Q_i . On the other hand, for continuous Euclidean spaces, such an algorithm was designed in [1]. If the algorithm does not fail at this step, then we update our set X and continue to the next iteration. In the rest of the section, we will prove that the algorithm returns a $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria approximation with good probability.

3.2 Analysis

Throughout the analysis, we assume that we are given a guess \mathcal{G} , such that $\text{OPT}_r \leq \mathcal{G} \leq (1 + \varepsilon/3) \cdot \text{OPT}_r$. We divide the analysis in two parts. In the first part, we bound the running time of the algorithm using the following lemma. The proof of this lemma is present in Section 3.2.1.

► **Lemma 7.** *Algorithm 1 terminates in $O(\frac{k}{\varepsilon} \log(\frac{k}{\varepsilon}) \lambda(\frac{\varepsilon}{40}))$ iterations – with or without failure.*

In the second part, we show the following lemma, which says that the probability that the algorithm terminates without failure is high. The proof the lemma is present in Section 3.2.2.

► **Lemma 8.** *With probability at least $\exp(-O(\frac{k}{\varepsilon} \log(\frac{k}{\varepsilon}) \lambda(\frac{\varepsilon}{40})))$, Algorithm 1 terminates without failure, i.e., returns a solution X satisfying $\text{cost}_{(1+\varepsilon/3)r}(P, X) \leq (1 + \varepsilon)\text{OPT}_r$.*

Using these two lemmas and repeating the algorithm $\exp(O(\frac{k}{\varepsilon} \log(\frac{k}{\varepsilon}) \lambda(\frac{\varepsilon}{40})))$ times, we have our main result.

► **Theorem 9 (Main Theorem).** *Let \mathcal{M} be a class of metric spaces closed under scaling distances by a positive constant. There is a randomized algorithm that takes as input an instance $\mathcal{I} = ((P, \mathbb{F}, \text{dist}), k, r)$ of HYBRID k -CLUSTERING such that $(P, \mathbb{F}, \text{dist}) \in \mathcal{M}$ and $\varepsilon \in (0, 1)$, and outputs a $(1 + \varepsilon, 1 + \varepsilon)$ -bicriteria solution for \mathcal{I} when, for all $\varepsilon' > 0$, the algorithmic ε' -scatter dimension of \mathcal{M} is bounded by $\lambda(\varepsilon')$, for some function λ . The running time of the algorithm is $2^{O(\frac{k}{\varepsilon} \cdot \log(k/\varepsilon) \cdot \lambda(\varepsilon/40))} \cdot |\mathcal{I}|^{O(1)}$.*

3.2.1 Bounding runtime using Algorithmic Scatter Dimension

First, we show some properties of the initial upper bounds (line 4), that we need later in the proof of Lemma 7.

► **Lemma 10 (Feasible upper bounds).** *Consider $Q_i = \{(p^{(i)}, u(p^{(i)}))\}$ initialized in Line 4 of Algorithm 1. Then, $\text{dist}(p^{(i)}, O) \leq u(p^{(i)})$.*

Proof. Suppose $\text{dist}(p^{(i)}, O) > u(p^{(i)})$. Letting $\alpha = u(p^{(i)})/3$, we have that $|\text{ball}(p, \alpha)| \geq \mathcal{G}/\alpha$. Since, $\text{dist}(p^{(i)}, O) > 3\alpha$, we have $\text{dist}(p, O) > 2\alpha$ for $p \in \text{ball}(p^{(i)}, \alpha)$. Using $\alpha > r$, this means $\text{dist}_r(p, O) > \alpha$ for $p \in \text{ball}(p, \alpha)$. Therefore,

$$\text{cost}_r(P, O) \geq \sum_{p \in \text{ball}(p^{(i)}, \alpha)} \text{dist}_r(p, O) = \sum_{p \in \text{ball}(p^{(i)}, \alpha)} (\text{dist}(p, O) - r) > \frac{\mathcal{G}}{\alpha} \cdot \alpha = \mathcal{G},$$

contradicting the the cost of O . ◀

Next, we have the following lemma, whose proof is identical to [1], that says that the initialization of X at line 6 is successful and the solution maintained by the algorithm always satisfies the upper bounds within a factor of 3.1.

► **Lemma 11 (Lemma V.5 of [1]).** *The number of marked points in line 3 is at most k , i.e., $k' \leq k$. Hence, the initialization of X at line 6 is successful. Furthermore, at any iteration, the solution X maintained by the algorithm satisfies that $\text{dist}(p, X) \leq 3.1u(p)$, for every $p \in P$.*

Bounding aspect ratio of radii

Towards proving Lemma 7, we bound the aspect ratio of the radii in the requests.

► **Lemma 12.** *Consider a request set $Q_i = \{(p_i^{(1)}, \delta_i^{(1)}), \dots, (p_i^{(\ell)}, \delta_i^{(\ell)})\}$, for $i \in [k]$. Let $X^{(j)}$, $j \in [\ell]$, be the center maintained by the algorithm just before adding the request $(p_i^{(j)}, \delta_i^{(j)})$ to Q_i . Further, let $x_i^{(j)} \in X^{(j)}$ be the center corresponding to cluster $i \in [k]$. Then, the sequence $S_i = (x_i^{(1)}, p_i^{(1)}, \delta_i^{(1)}), \dots, (x_i^{(\ell)}, p_i^{(\ell)}, \delta_i^{(\ell)})$ is an algorithmic $(\mathcal{C}, \varepsilon/20)$ -scattering. Furthermore, the radii of requests in S_i lie in the interval $[r, 8r/\varepsilon] \cup [r_{\min}, \frac{10^5 k}{\varepsilon^2} r_{\min}]$, where r_{\min} is the smallest radii in S_i that is larger than $8r/\varepsilon$.*

Proof. The proof of the first part is similar to [1].

First note that $\delta_i^{(j)} = \frac{\text{dist}(p_i^{(j)}, X^{(j)})}{1 + \varepsilon/12} \leq \frac{\text{dist}(p_i^{(j)}, x_i^{(j)})}{1 + \varepsilon/12}$. Finally, since $x_i^{(j)}$ is computed by \mathcal{C} on $\{(p_i^{(1)}, \delta_i^{(1)}), \dots, (p_i^{(j-1)}, \delta_i^{(j-1)})\}$ and error parameter $\varepsilon/40$, we have that $\text{dist}(p_i^{(j')}, x_i^{(j)}) \leq (1 + \varepsilon/40)\delta_i^{(j')}$, for $j' < j$. Hence, S_i is $(\mathcal{C}, \varepsilon/20)$ -algorithmic scattering.

35:10 Dimension-Free Parameterized Approximation Schemes for Hybrid Clustering

Now we bound that the radii of sequence $S_i, i \in [k]$. Towards this, we partition S_i into S_i^n and S_i^f , based on a partitioning of Q_i , as follows. We say a request $(p_i^{(j)}, \delta_i^{(j)}) \in Q_i$ belongs to Q_i^n if $p_i^{(j)}$ was sampled when the If condition was satisfied (in Line 12), otherwise it belongs to Q_i^f , which corresponds to the Else condition (in Line 15). Correspondingly, $(x_i^{(j)}, p_i^{(j)}, \delta_i^{(j)}) \in S_i^n$ if $(p_i^{(j)}, \delta_i^{(j)}) \in Q_i^n$ and $(x_i^{(j)}, p_i^{(j)}, \delta_i^{(j)}) \in S_i^f$ if $(p_i^{(j)}, \delta_i^{(j)}) \in Q_i^f$. We bound the aspect ratio of each part, S_i^n and S_i^f , separately. For $(x_i^{(j)}, p_i^{(j)}, \delta_i^{(j)}) \in S_i^n$, we have $r \leq \delta_i^{(j)} \leq 8r/\varepsilon$ and hence, the radii of triples in S_i^n lie in interval $[r, 8r/\varepsilon]$. Now consider $(x_i^{(j)}, p_i^{(j)}, \delta_i^{(j)}) \in S_i^f$. Recall that $X^{(j)}$ is the solution maintained by the algorithm when $(p_i^{(j)}, \delta_i^{(j)})$ is added to Q_i . Then, note that $\text{dist}_{r'}(p_i^{(j)}, X^{(j)}) > \varepsilon u(p)/1000k$ (see Line 14), and hence, we have $\text{dist}(p_i^{(j)}, X^{(j)}) > \varepsilon u(p)/1000k$. The following claim, whose proof is similar to [1], shows that the radii of requests in S_i^f are also bounded, finishing the proof of the lemma.

▷ **Claim 13.** Consider requests $(p, \delta_p), (p', \delta_{p'})$ added (in any order) to Q_i in Line 18 of Algorithm 1 such that $(p, \delta_p), (p', \delta_{p'}) \in Q_i^f$. If $\delta_{p'} < \varepsilon^2 \delta_p / 10^5 k$, then the algorithm fails in Line 19 upon making second of the two requests.

Proof. Suppose, for the sake of contradiction, the algorithm does not fail and finds a center x_i such that $\text{dist}(p, x_i) \leq (1 + \varepsilon/40)\delta_p$ and $\text{dist}(p', x_i) \leq (1 + \varepsilon/40)\delta_{p'}$. Thus, $\text{dist}(p, p') \leq \text{dist}(p, x_i) + \text{dist}(p', x_i) \leq (1 + \varepsilon/40)(\delta_p + \delta_{p'})$.

Therefore, we have

$$\text{dist}(p, X) \leq \text{dist}(p, p') + \text{dist}(p', X) \leq (1 + \varepsilon/40)(\delta_p + \delta_{p'}) + 3.1u(p'),$$

using Observation 11. Let X be the center maintained by the algorithm when the request (p, δ_p) is added to Q_i . Then, we have

$$\delta_p = \frac{\text{dist}(p, X)}{(1 + \varepsilon/12)}, \tag{1}$$

due to Line 18. Similarly, let X' be the center maintained by the algorithm when request $(p', \delta_{p'})$ is added to Q_i . Then, since $(p', \delta_{p'}) \in Q_i^f$, we have that $u(p') \leq 900\text{dist}(p', X')/\varepsilon = 1000k\delta_{p'}/\varepsilon$, using $\delta_{p'} = \frac{\text{dist}(p', X')}{(1 + \varepsilon/12)}$. Therefore, using $\delta_{p'} < \varepsilon^2 \delta_p / 10^5 k$,

$$\begin{aligned} \text{dist}(p, X) &\leq (1 + \varepsilon/40)(\delta_p + \delta_{p'}) + 3000k\delta_{p'}/\varepsilon \leq (1 + \varepsilon/40)\delta_p + 3200k\delta_{p'}/\varepsilon \\ &\leq (1 + \varepsilon/40 + \varepsilon/25)\delta_p < (1 + \varepsilon/12)\delta_p \end{aligned}$$

This means $\delta_p > \frac{\text{dist}(p, X)}{1 + \varepsilon/12}$, contradicting (1). ◁

◀

Now we are ready to finish the proof of Lemma 7.

Proof of Lemma 7. We apply Corollary 6 to S_i and note that the radii in S_i lie in $[r, 8r/\varepsilon] \cup [r_{\min}, \frac{10^5 k}{\varepsilon^2} r_{\min}]$, where r_{\min} is the smallest radii in S_i that is larger than $8r/\varepsilon$, due to Lemma 12. This implies that the length of sequence S_i is bounded by $O(\lambda(\varepsilon/40) \frac{(\log(k/\varepsilon))}{\varepsilon})$. Since in each iteration the algorithm adds one request to some Q_i , the total number of iterations is bounded by $O(\frac{k}{\varepsilon} \lambda(\varepsilon/40) (\log(k/\varepsilon)))$. ◀

3.2.2 Bounding success probability

The goal of this subsection is to prove Lemma 8, i.e., give a lower bound on the success probability of the algorithm. To this end, we first introduce the following notion.

► **Definition 14** (Consistency). *Consider a fixed hypothetical optimal solution $O = (o_1, \dots, o_k)$. We say that the current state of execution (specified by (X, Q_1, \dots, Q_k)) of Algorithm 1 is consistent with O if for any request $(p, \delta) \in Q_i, i \in [k]$, we have $\text{dist}(p, o_i) \leq \delta$.*

Note that, Lemma 11 implies that the initial set of requests (line 4) are feasible. Since the balls $\text{ball}(p^{(i)}, u(p^{(i)})), i \in [k']$ are disjoint, we can relabel the optimum centers $O = \{o_1, \dots, o_k\}$ so that $\text{dist}(p^{(i)}, o_i) \leq u(p^{(i)})$. Thus, Lemma 11 implies that the initial state of the algorithm as computed in lines 1 to 6 is consistent with a fixed optimal solution O which is fixed henceforth. To prove Lemma 8, we will inductively argue that the state of the algorithm remains consistent with O – note that the base case is already shown. Most of this subsection is devoted to show the inductive step, i.e., to show that, given a consistent state at the start of an iteration, there is a good probability that the state of the algorithm at the end of the iteration is also consistent (cf. Lemma 22).

To this end, we introduce the following definitions w.r.t. the current solution X .

► **Definition 15** (Contribution and Witness). *For any set $S \subseteq P$, we use $C_S := \sum_{p \in S} \text{dist}_{r'}(p, X)$ to denote the contribution of the set S to the cost of the current solution.*

We say that a point $p \in P$ is an $\varepsilon/3$ -witness (or simply, a witness) w.r.t. the solution X if it satisfies $\text{dist}_{(1+\varepsilon/3)r}(p, X) > (1 + \varepsilon/3) \cdot \text{dist}_r(p, O)$.

The following claim can be proved by a simple averaging argument.

▷ **Claim 16** (♠). $C_W \geq \frac{\varepsilon C_P}{10}$.

Next, we introduce several different classifications of witnesses.

► **Definition 17** (Different subsets of witnesses). *For each $x_j \in X$, let W_j denote the set of witnesses for which x_j is a nearest center in X (breaking ties arbitrarily). Then, $W_{j,\text{near}} := \{p \in W_j : \text{dist}(p, x_j) \leq 8r/\varepsilon\}$, and $W_{j,\text{far}} := W_j \setminus W_{j,\text{near}}$. Further, let $W_{\text{near}} := \bigcup_{j \in [k]} W_{j,\text{near}}$, and $W_{\text{far}} := \bigcup_{j \in [k]} W_{j,\text{far}}$. We will refer to a witness in W_{near} as a nearby witness and a witness in W_{far} as a faraway witness.*

Now, we consider two different cases regarding the behavior of the algorithm.

Case 1: $C_{W_{\text{near}}} \geq \frac{\varepsilon}{100} C_P$. In this case, when we sample a point from $N := \{p \in P : \text{dist}(p, X) \leq 8r/\varepsilon\}$ proportional to their $\text{dist}_{r'}(\cdot, X)$ values, the probability of sampling a nearby witness is at $\frac{\varepsilon}{100}$. This will correspond to the “good event”. We prove this formally in the following lemma.

► **Lemma 18** (Nearby witness lemma). *Suppose the current solution X at the start of an iteration satisfies $\text{cost}_{r'}(P, X) > (1 + \varepsilon) \cdot \mathcal{G}$. Further, suppose $C_{W_{\text{near}}} \geq \frac{\varepsilon}{100} C_P$. Then, with probability at least $\frac{\varepsilon}{200k}$, the point $p \in P$, the index $i \in [k]$, and value δ_p defined in the iteration satisfy the following properties.*

1. $o_i \in O$ is the closest center to $p \in P$, i.e., $\text{dist}(p, o_i) = \text{dist}(p, O)$,
2. $p \in W_{\text{near}}$, i.e., (i) $\text{dist}_{r'}(p, X) > (1 + \varepsilon/3)\text{dist}_r(p, o_i)$, and (ii) $\text{dist}(p, X) \leq \frac{8r}{\varepsilon}$.
3. $\text{dist}(p, o_i) < \frac{\text{dist}(p, X)}{1 + \varepsilon/12} =: \delta_p \leq \frac{8r}{\varepsilon}$.

35:12 Dimension-Free Parameterized Approximation Schemes for Hybrid Clustering

Proof. First, in line 10 with probability $1/2$, we correctly move to the “nearby witness” (if) case of line 10. We condition on this event. Next, note that $W_{\text{near}} \subseteq N$, and $C_{W_{\text{near}}} \geq \frac{\varepsilon}{100}C_P \geq \frac{\varepsilon}{100}C_N$, where the first inequality is due to the case assumption. Therefore, in line 12, the point p sampled from N , will belong to W_{near} with probability at least $\frac{\varepsilon}{100}$. Let $i \in [k]$ denote the index such that $o_i \in O$ is the closest center in the optimal solution O , and we correctly guess the index i in line 17. Note that the probability of the algorithm making the “correct” random choices is at least $\frac{1}{2} \cdot \frac{\varepsilon}{100} \cdot \frac{1}{k} = \frac{\varepsilon}{200k}$.

We condition on the said events. Since p is a witness, we know that $\text{dist}_{r'}(p, X) > (1 + \varepsilon/3) \cdot \text{dist}_r(p, o_i)$. We first observe that $\text{dist}_{r'}(p, X)$ must be positive, which implies that $\text{dist}(p, X) > r' = (1 + \varepsilon/3)r$. Now, we consider two cases based on the value of $\text{dist}(p, O)$.

If $\text{dist}(p, o_i) < r$, then $\text{dist}_r(p, o_i) = 0$, in which case,

$$\text{dist}(p, o_i) < r \leq \frac{r'}{1 + \varepsilon/12} \leq \frac{\text{dist}(p, X)}{1 + \varepsilon/12} \quad (2)$$

Otherwise, $\text{dist}(p, o_i) \geq r$, in which case $\text{dist}_r(p, o_i) = \text{dist}(p, o_i) - r$. Then, consider

$$\begin{aligned} \text{dist}(p, X) - (1 + \frac{\varepsilon}{3})r &= \text{dist}_{r'}(p, X) > (1 + \frac{\varepsilon}{3}) \cdot \text{dist}_r(p, o_i) = (1 + \frac{\varepsilon}{3})(\text{dist}(p, o_i) - r) \\ \implies \text{dist}(p, X) - (1 + \frac{\varepsilon}{3})r &> (1 + \frac{\varepsilon}{3}) \cdot \text{dist}(p, o_i) - (1 + \frac{\varepsilon}{3})r \\ \implies \text{dist}(p, o_i) < \frac{\text{dist}(p, X)}{1 + \varepsilon/3} &\leq \frac{\text{dist}(p, X)}{1 + \varepsilon/12} \end{aligned} \quad (3)$$

Thus, regardless of the value of $\text{dist}(p, o_i)$, we have established the third item, hence completing the proof of the lemma. \blacktriangleleft

Case 2: $C_{W_{\text{near}}} < \frac{\varepsilon}{100}C_P$. In this case, most of contribution of witnesses is coming from “faraway witnesses”. In this case, the “correct choice” corresponds to the case when we sample points from the set $A = \{p \in P : \text{dist}_{r'}(p, X) > \frac{\varepsilon}{1000k} \cdot u(p)\}$ as defined in line 14. In this case, we will show that with good probability, the sampled point is a “faraway witness”. Specifically, we show the following lemma.

► **Lemma 19 (Faraway witness lemma).** *Suppose the current solution X at the start of an iteration satisfies $\text{cost}_{r'}(P, X) > (1 + \varepsilon) \cdot \mathcal{G}$. Further, suppose $C_{W_{\text{near}}} < \frac{\varepsilon}{100}C_P$. Then, with probability at least $\frac{\varepsilon}{16k}$, the point $p \in P$, the index $i \in [k]$, and value δ_p defined in the iteration satisfy the following properties.*

1. $o_i \in O$ is the closest center to p , i.e., $\text{dist}(p, o_i) = \text{dist}(p, O)$,
2. $p \in W_{\text{far}}$, i.e., (i) $\text{dist}_{r'}(p, X) > (1 + \varepsilon/3)\text{dist}_r(p, o_i)$, and (ii) $\text{dist}(p, X) > \frac{8r}{\varepsilon}$,
3. $\text{dist}_{r'}(p, X) > \frac{\varepsilon}{1000k}u(p)$, and
4. $\text{dist}(p, o_i) < \frac{\text{dist}(p, X)}{1 + \varepsilon/12} =: \delta_p \leq \frac{8r}{\varepsilon}$.

Proof. First, in line 10 with probability $1/2$, we correctly move to the “faraway witness” (else) case of line 13. We condition on this event. Now, by combining Claim 16 and the case assumption we obtain,

$$C_{W_{\text{far}}} = C_W - C_{W_{\text{near}}} \geq C_P \cdot \left(\frac{\varepsilon}{10} - \frac{\varepsilon}{100}\right) = \frac{\varepsilon}{9}C_P. \quad (4)$$

Next, let $H := \{j \in [k] : C_{W_{j, \text{far}}} \geq \frac{\varepsilon C_P}{100k}\}$. It follows that,

$$\begin{aligned} \sum_{j \in [k] \setminus H} C_{W_{j, \text{far}}} &\leq k \cdot \frac{\varepsilon C_P}{100k} \\ \implies \sum_{j \in H} C_{W_{j, \text{far}}} &\geq C_{W_{\text{far}}} - \sum_{j \in [k] \setminus H} C_{W_{j, \text{far}}} \geq \frac{\varepsilon C_P}{9} - \frac{\varepsilon C_P}{100} \geq \frac{\varepsilon C_P}{8} \end{aligned} \quad (5)$$

Fix a $j \in H$, and let us index the points in $W_{j,\text{far}} = \{z_1, z_2, \dots, z_\ell\}$ in the non-decreasing order of their distances $\text{dist}(z, x_j)$ (and equivalently, $\text{dist}_r(z, x_j)$). First, we state the following simple consequences of various definitions for future reference.

► **Observation 20** (♠). *For each $z \in W_{j,\text{far}}$, the following bounds hold.*

1. $\text{dist}(z, x_j) \geq \frac{8r}{\varepsilon}$,
2. $\text{dist}(z, O) > (1 + \frac{\varepsilon}{3}) \cdot \text{dist}(z, x_j)$
3. $\text{dist}_{r'}(z, x_j) \geq \frac{6r}{\varepsilon}$, and
4. $\text{dist}_r(z, x_j) \leq \text{dist}(z, x_j) \leq (1 + \frac{\varepsilon}{5}) \cdot \text{dist}_{r'}(z, x_j) \leq (1 + \frac{\varepsilon}{5}) \cdot \text{dist}_r(z, x_j)$.

Let us return to the points in $W_{j,\text{far}} = \{z_1, z_2, \dots, z_\ell\}$. Let $q \in [\ell]$ denote the minimum index such that, the contribution of the set $W_{j,\text{far}}^- = \{z_1, z_2, \dots, z_q\}$, is at least $\frac{C_{W_{j,\text{far}}}}{2}$. Note that by a simple argument, this implies that the contribution of the set $W_{j,\text{far}}^+ = W_{j,\text{far}} \setminus W_{j,\text{far}}^-$ is also at least $\frac{C_{W_{j,\text{far}}}}{3}$ (see, e.g., Lemma 7.20 in [21]). Hence, we have the following lower bounds on the contribution of both the sets, by recalling that $j \in H$.

$$C_{W_{j,\text{far}}^-} \geq \frac{\varepsilon C_P}{200k}, \quad C_{W_{j,\text{far}}^+} \geq \frac{\varepsilon C_P}{300k} \quad (6)$$

Next, we first prove the following technical claim.

▷ **Claim 21.** Let $j \in H$. For all $p \in W_{j,\text{far}}^+$, it holds that $u(p) \leq \frac{900k \text{dist}(p, x_j)}{\varepsilon}$. Therefore, $W_{j,\text{far}}^+ \subseteq A$.

Proof. For this proof, we use the following shorthand: $W^+ := W_{j,\text{far}}^+$, and $W^- := W_{j,\text{far}}^-$. Now, fix an arbitrary point $p \in W^+$. For any $q \in W^-$, $\text{dist}(p, x_j) \geq \text{dist}(q, x_j)$, which implies that

$$\text{dist}(p, q) \leq \text{dist}(p, x_j) + \text{dist}(q, x_j) \leq 2 \cdot \text{dist}(p, x_j) \leq 2(1 + \frac{\varepsilon}{5}) \cdot \text{dist}_{r'}(p, x_j) \leq 4 \cdot \text{dist}_{r'}(p, x_j) \quad (7)$$

Here, we use the property 4 from Observation 20 in the third inequality in the above. On the other hand, note that,

$$\frac{\varepsilon \text{OPT}_r}{200k} \leq \frac{\varepsilon C_P}{200k} \leq C_{W^-} = \sum_{q \in W^-} \text{dist}_{r'}(q, x_j) \leq \text{dist}_{r'}(p, x_j) \cdot |W^-| \quad (8)$$

Then, if we set $\alpha = 12 \text{dist}_{r'}(p, x_j)$, from (7), we obtain that $W^- \subseteq \text{ball}(p, \alpha/3)$. Now, combining this with (8), we obtain that,

$$|\text{ball}(p, \alpha/3)| \geq |W^-| \geq \frac{\varepsilon \text{OPT}_r}{300k \cdot \text{dist}_{r'}(p, x_j)} \geq \frac{\varepsilon \text{OPT}_r}{25k\alpha} = \frac{\varepsilon \text{OPT}_r}{75k(\alpha/3)} \quad (9)$$

Hence, we have that $|\text{ball}(p, \frac{75k\alpha}{3\varepsilon})| \geq \frac{\varepsilon \text{OPT}_r}{75k(\alpha/3)}$. Using $\text{dist}(p, x_j) \geq \frac{8r}{\varepsilon}$ from Observation 20 and the fact $\text{dist}_r(p, x_j) \geq 5 \text{dist}(p, x_j)/6$, we have $\frac{75k\alpha}{3\varepsilon} = \frac{300k \text{dist}_{r'}(p, x_j)}{\varepsilon} > r$. Therefore, $u(p) \leq \frac{900k \text{dist}_{r'}(p, x_j)}{\varepsilon} \leq \frac{900k \text{dist}(p, x_j)}{\varepsilon}$, as desired. ◁

Recall from (5) that $C_{W_{\text{far}} \cap A} \geq \sum_{j \in H} C_{W_{j,\text{far}}^+} \geq \frac{\varepsilon C_P}{8} \geq \frac{\varepsilon C_A}{8}$, therefore, when we sample a point from A , the probability that the sampled point p belongs to $\bigcup_{j \in H} W_{j,\text{far}}^+$ is at least $\frac{\varepsilon}{8}$. Now, we condition on this event. Let $o_i \in O$ be the nearest center to p , and in line 17, the probability that we sample the correct index is $\frac{1}{k}$. Thus, the total probability of the algorithm making the “correct choices” in this case is at least $\frac{1}{2} \cdot \frac{\varepsilon}{8} \cdot \frac{1}{k} = \frac{\varepsilon}{16k}$. We condition on these choices. Note that, item 1 is thus satisfied due to the correct sampling of i , item 2 is satisfied due $p \in \bigcup_{j \in H} W_{j,\text{far}}^+ \subseteq W_{\text{far}} \subseteq W$, and item 3 is satisfied since $p \in A$ by construction. Thus, we are only left with showing item 4, to which end, we consider different cases for the distance between p and its closest optimal center, o_i .

If $\text{dist}(p, o_i) < \frac{6r}{\varepsilon}$, then since $\text{dist}(p, X) \geq \frac{8r}{\varepsilon}$, it easily follows that,

$$\text{dist}(p, o_i) < \frac{3}{4} \cdot \text{dist}(p, X) \leq \frac{\text{dist}(p, X)}{1 + \varepsilon/12} \quad (10)$$

Otherwise, if $\text{dist}(p, o_i) \geq \frac{6r}{\varepsilon}$, then similar item 4 of Observation 20, we can show that,

$$\text{dist}_r(p, o_i) < \text{dist}(p, o_i) \leq (1 + \frac{\varepsilon}{3}) \cdot \text{dist}_r(p, o_i) \quad (11)$$

However, since p is an $\varepsilon/3$ -witness, it follows that, $\text{dist}_{r'}(p, X) > (1 + \frac{\varepsilon}{3}) \cdot \text{dist}_r(p, O)$. Therefore, we obtain that,

$$\text{dist}(p, x_j) \geq \text{dist}_{r'}(p, x_j) > (1 + \frac{\varepsilon}{3}) \cdot \text{dist}_r(p, o_i) \geq \frac{1 + \varepsilon/3}{1 + \varepsilon/5} \cdot \text{dist}(p, o_i) \geq (1 + \frac{\varepsilon}{12}) \cdot \text{dist}(p, o_i) \quad (12)$$

Thus, in each of the sub-cases, in (10) and (12), we have shown that $\text{dist}(p, o_i) < \frac{\text{dist}(p, X)}{1 + \varepsilon/12}$, thus completing the proof of the lemma. \blacktriangleleft

Lemma 18 and Lemma 19 can be combined in a straightforward way to obtain the following lemma completing the inductive step.

► Lemma 22. *Consider an iteration of the algorithm such that the state of execution (X, Q_1, \dots, Q_k) at the start of the iteration is consistent with a fixed optimal solution O , and further $\text{cost}_{r'}(P, X) > (1 + \varepsilon) \cdot \mathcal{G}$. Then, with probability at least $\frac{\varepsilon}{200k}$, the state of the algorithm at the end of the iteration $(X', Q'_1, Q'_2, \dots, Q'_k)$ is also consistent with O .*

Proof. Consider the scenario described in the premise of the lemma. Then, the solution X at the start of the iteration either satisfies that $C_{W_{\text{near}}} \geq \frac{\varepsilon}{100} C_P$ (nearby witness case), or $C_{W_{\text{near}}} \geq \frac{\varepsilon}{100} C_P$ (faraway witness case). Then, Lemmas 18 and 19; along with the correctness of \mathcal{C} , together imply that, conditioned on the respective case assumption, the probability that the state of the algorithm is consistent with O is at least $\frac{\varepsilon}{200k}$.⁹ \blacktriangleleft

Armed with Lemma 22, it is easy to conclude the proof of Lemma 8.

Proof of Lemma 8. As argued initially, the state of the algorithm at line 6 is consistent with the optimal solution O . Then, Lemma 22 implies that, for any $\ell \geq 1$, if the algorithm runs for ℓ iterations, then with probability at least $(\frac{\varepsilon}{200k})^\ell$, the state of the algorithm remains consistent. Further, note that as long as the state of the algorithm remains consistent with O , then the algorithm cannot fail. Finally, Lemma 7 implies that the algorithm terminates within $O(\frac{k}{\varepsilon} \lambda(\frac{\varepsilon}{40}) \log(\frac{k}{40}))$ iterations. Therefore, the probability that the algorithm terminates without failure – i.e., by successfully breaking the **while** loop by finding a solution X satisfying $\text{cost}_{r'}(P, X) \leq (1 + \varepsilon) \text{OPT}_r$, is at least $(\frac{\varepsilon}{k})^{O(\frac{k}{\varepsilon} \lambda(\frac{\varepsilon}{40}) \log(\frac{k}{40}))} = \exp(-O(\frac{k}{\varepsilon} \log(\frac{k}{\varepsilon}) \lambda(\frac{\varepsilon}{40})))$. \blacktriangleleft

4 Extensions of the Bicriteria FPT-AS

Fomin et al. [23] defined a generalization of HYBRID k -CLUSTERING, which they call HYBRID (k, z) -CLUSTERING, wherein the objective is $\sum_{p \in P} (\text{dist}_r(p, X))^z$, for fixed $z \geq 1$, which simultaneously generalizes (k, z) -CLUSTERING and k -CENTER. They generalized their algorithm for $z = 2$, i.e., HYBRID k -MEANS, but left the possibility of extending to an arbitrary value of z conditional on the existence of a certain kind of sampling algorithm. In this paper, we consider a much more general HYBRID NORM k -CLUSTERING problem that captures all

⁹ Note that this probability also accounts for the result of the coin toss in line 10.

(k, z) -CLUSTERING problems, and much more. Here, the objective is to find a set $X \subseteq \mathbb{F}$ that minimizes $f(\mathbf{d}_r(P, X))$, where $\mathbf{d}_r(P, X) = (\text{dist}_r(p_1, X), \text{dist}_r(p_2, X), \dots, \text{dist}_r(p_n, X))$ is the vector of r -distances to the solution X , and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a *monotone norm*¹⁰ – we refer the reader to [1] for a detailed background on norms and related notions.

Algorithm 1 and its analysis can be extended to the general norm objective following the approach of [1]. Here, we only sketch the modifications required to the algorithm and its proof, since this is not the main focus of this work.

The initial upper bounds $u(p)$ can be found in a similar manner. The while loop runs as long as the solution X in hand satisfies that $f(\mathbf{d}_{r'}(P, X)) > (1 + \varepsilon/3) \cdot \text{OPT}_r$, where $\mathbf{d}_{r'}$ denotes the vector of distances $(\text{dist}_{r'}(p, X))_{p \in X}$, and $r' = (1 + \varepsilon/3)r$. If this is the case, then the first step is to compute¹¹ a *subgradient* of f at $\mathbf{d}_{r'}$, which is, loosely speaking, a weight function $w : P \rightarrow \mathbb{R}_{\geq 0}$, such that it suffices to focus the attention in the current iteration on $\text{wcost}_{r'}(P, X) := \sum_{p \in P} w(p) \text{dist}_r(p, X)$ that satisfies $\text{wcost}_{r'}(P, X) = f(\mathbf{d}_{r'}(P, X)) > (1 + \varepsilon/3) \cdot \text{OPT}_r$. Our algorithm proceeds in a similar fashion, except that whenever points are sampled from the set N (in line 12) or set A (in line 15), the probability of sampling a point p is proportional to $w(p) \cdot \text{dist}_r(p, X)$. The rest of the algorithm remains unchanged.

There are a few minor modifications required in the analysis – mainly in Section 3.2.2. First, while the definition of an $(\varepsilon/3)$ -witness (or simply a *witness*) remains unchanged, we redefine the *contribution* of a subset $S \subseteq P$ to be $C_S := \sum_{p \in S} w(p) \cdot \text{dist}_r(p, X)$. The nearby and faraway witness cases are then considered exactly as in Section 3.2.2. The proof of the analogue of Lemma 18 goes through without any modifications; whereas we need to be slightly more careful in the proof of the analogue of Lemma 19. Here, in inequalities (8 and 9) we need to take the weighted distances into account and relate C_P to $f(\mathbf{d}_{r'}(P, X))$, which requires a slight worsening of constants. With these minor changes in place, we can conclude that the state of the algorithm after ℓ iterations is consistent with a fixed optimal solution O with probability at least $\Omega((\frac{\varepsilon}{k}))^\ell$. The analysis for bounding the number of iterations remains unchanged, and hence we obtain an FPT-AS with a similar running time, modulo the constants hidden in the big-Oh notation. We omit the details.

5 Coreset

In this section, we design a coreset for HYBRID k -CLUSTERING in doubling metric of bounded dimension. More specifically, we prove the following (restated for convenience).

► **Theorem 3** (Coreset for HYBRID k -CLUSTERING). *There exists an algorithm that takes as input an instance $\mathcal{I} = ((P, \mathbb{F}, \text{dist}), k, r)$ of HYBRID k -CLUSTERING in doubling metric of dimension d and a parameter $\varepsilon \in (0, 1)$, and in time $2^{O(d \log(1/\varepsilon))} |\mathcal{I}|^{O(1)}$ returns a pair (P', w) , where $P' \subseteq P$ has size $2^{O(d \log(1/\varepsilon))} k \log |P|$, and $w : P' \rightarrow \mathbb{N}$, such that the following property is satisfied for any $X \subseteq \mathbb{F}$ of size at most k :*

$$|\text{wcost}_r(P', X) - \text{cost}_r(P, X)| \leq \varepsilon \text{cost}_r(P, X)$$

Here, $\text{wcost}_r(P, X) := \sum_{p \in P'} w(p) \cdot \text{dist}_r(p, X)$.

Proof. A formal description of our algorithm can be found in the full version of the paper. Here, we provide an overview of the overview, which is based on grid construction approach of [26, 2]. Let $T \subseteq P \cup \mathbb{F}$ be an approximate solution that satisfies the following properties:

¹⁰ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a *norm* if it satisfies following properties for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and any $\lambda \in \mathbf{R}$, (i) $f(\mathbf{x}) = 0$ iff $\mathbf{x} = \mathbf{0}$, (ii) $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$, (iii) $f(\lambda \mathbf{x}) = |\lambda| f(\mathbf{x})$; furthermore, f is *monotone* if $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\mathbf{x} \leq \mathbf{y}$, where \leq denotes point-wise inequality.

¹¹ Actually, it is sufficient to compute an approximate subgradient, as done in [1].

(i) $|T| \leq 2^{O(d)}k$, and (ii) $\text{cost}_r(P, T) \leq 36 \cdot \text{OPT}_r$ ¹². For each $t_i \in T$, consider the balls $B_i^j = \text{ball}(t_i, 2^j R)$, for $j \in \{0, 1, \dots, 2 \log \lceil \alpha n \rceil\}$. Note that, since $\text{dist}(p, T) \leq \text{cost}(P, T) = \alpha n \cdot R$, we have that p lies in some ball B_i^j , for $i \in [\gamma k]$ and $j \in \{0, 1, \dots, 2 \log \lceil \alpha n \rceil\}$.

The idea is to decompose each B_i^j into smaller balls each of radius $\frac{\varepsilon 2^j R}{4\alpha}$, and associate each point $p \in P$ to a unique smallest ball containing p , breaking ties arbitrary. We say a small ball b is non-empty if there is a point in P associated with b . Next, for each non-empty ball b , pick an arbitrary point p associated with b and add p to P' with weight $w(p)$ equal to the total number of points associated with b ; we call such a point p as the representative of points associated with b .

To bound the size of P' , note that, in doubling metric of dimension d , a unit ball can be covered by $2^{O(d \log(1/\varepsilon))}$ balls, each of radius ε . Hence, we have $|P'| = O(2^{O(d \log(1/\varepsilon))} \gamma k \log(\alpha n))$.

Recall that, for $X \subseteq \mathbb{F}$, $|X| \leq k$, we define the weighted cost of X with respect to $(P', \{w(p')\}_{p' \in P'})$ as $\text{wcost}_r(P', X) = \sum_{p' \in P'} w(p') d_r(p', X)$. Now we show that the cost of X with respect to $(P', \{w(p')\}_{p' \in P'})$ approximately preserves the cost of X with respect to P . Consider a point $p \in P$ and let $p' \in P'$ be its representative in P' . Now, the contribution of p towards $\text{cost}_r(P, X)$ is $\text{dist}_r(p, X)$. On the other hand, its contribution towards $\text{wcost}_r(P', X)$ is $\text{dist}_r(p', X)$. Hence,

$$|\text{wcost}_r(P', X) - \text{cost}_r(P, X)| \leq \sum_{p \in P} |\text{dist}_r(p, X) - \text{dist}_r(p', X)|.$$

Now if $p \in \bigcup_{i \in [\gamma k]} B_i^0$, then $\text{dist}(p, X) - \frac{\varepsilon R}{4\alpha} \leq \text{dist}(p', X) \leq \text{dist}(p, X) + \frac{\varepsilon R}{4\alpha}$. Hence,

$$\max\{\text{dist}(p, X) - \varepsilon R/4\alpha - r, 0\} \leq \text{dist}_r(p', X) \leq \max\{\text{dist}(p, X) + \varepsilon R/4\alpha - r, 0\}$$

Therefore, $\text{dist}_r(p, X) - \varepsilon R/4\alpha \leq \text{dist}_r(p', X) \leq \text{dist}_r(p, X) + \varepsilon R/4\alpha$. Hence, we have

$$\sum_{p \in P_0} |\text{dist}_r(p, X) - \text{dist}_r(p', X)| \leq \frac{\varepsilon R n}{4\alpha} \leq \frac{\varepsilon \text{OPT}_r}{2} \leq \frac{\varepsilon \text{cost}_r(P, X)}{2}.$$

Now, suppose $p \in P \setminus \bigcup_{i \in [\gamma k]} B_i^0$, then $\text{dist}(p, T) > R$. Let $j \geq 1$ be such that $2^{j-1} R \leq \text{dist}(p, T) \leq 2^j R$. Hence, we have that $2^j R \leq 2 \text{dist}(p, T)$. Therefore, using $\text{dist}(p, X) - \frac{\varepsilon d(p, T)}{2\alpha} \leq \text{dist}(p', X) \leq \text{dist}(p, X) + \frac{\varepsilon d(p, T)}{2\alpha}$, we have

$$\max\{\text{dist}(p, X) - \varepsilon d(p, T)/2\alpha - r, 0\} \leq \text{dist}_r(p', X) \leq \max\{\text{dist}(p, X) + \varepsilon d(p, T)/2\alpha - r, 0\}$$

This implies, $\text{dist}_r(p, X) - \varepsilon d(p, T)/2\alpha \leq \text{dist}_r(p', X) \leq \text{dist}_r(p, X) + \varepsilon d(p, T)/2\alpha$. Thus, we have

$$\sum_{p \in P \setminus P_0} |\text{dist}_r(p, X) - \text{dist}_r(p', X)| \leq \sum_{p \in P \setminus P_0} \frac{\varepsilon}{2\alpha} \text{dist}(p, T) \leq \frac{\text{OPT}_r}{2} \leq \frac{\varepsilon \text{cost}_r(P, X)}{2}.$$

Finally, we have $|\text{wcost}_r(P', X) - \text{cost}_r(P, X)| \leq \varepsilon \text{cost}_r(P, X)$, as desired.

To finish the proof, we invoke the coreset construction algorithm using the approximate set T obtained from the following lemma. At a high level, to obtain this lemma, we start from an (18, 6)-bicriteria approximation for HYBRID k -CLUSTERING from [23, 13], and use the fact that, a ball of radius $O(r)$ can be decomposed into $2^{O(d)}$ balls of radius r , converting the guarantee from $\text{dist}_{6r}(\cdot, \cdot)$ to $\text{dist}_r(\cdot, \cdot)$.

¹²Note the set T is a bicriteria solution, but in a different sense, since it approximates the size and the cost while using r -distances.

► **Lemma 23** (Approximate solution T for coresets construction). *There is a polynomial-time algorithm that, given an instance $\mathcal{I} = (P, \mathbb{F}, \text{dist}, r)$ of HYBRID k -CLUSTERING in doubling metric of dimension d , computes $T \subseteq P \cup \mathbb{F}$, $|T| \leq 2^{O(d)}k$ such that $\text{cost}_r(P, T) \leq 36 \cdot \text{OPT}_r$, where OPT_r is the optimal cost for \mathcal{I} . ◀*

Proof. Let A be $(18, 6)$ -bicriteria solution for \mathcal{I} obtained from [14], as mentioned in [23]. That is $\text{cost}_{6r}(P, A) \leq 18\text{OPT}_r$. Consider $B_a := \text{ball}(a, 12r)$ for $a \in A$, and decompose B_a into $2^{O(d)}$ smaller balls, each of radius $r/2$ – note that this follows from the fact that the metric space has doubling dimension d . For each smaller ball b such that $b \cap (P \cup \mathbb{F}) \neq \emptyset$, add an arbitrary $t \in b \cap (P \cup \mathbb{F})$ to T . Finally, add A to T . Hence, $|T| = |A| + 2^{O(d)}|A| = k2^{O(d)}$.

It is easy to see that, for every $p \in \bigcup_{c \in A} \text{ball}(c, 12r)$, there is some $q \in T$, such that $\text{dist}(p, q) \leq r$. Now, consider some $p' \notin \bigcup_{c \in A} \text{ball}(c, 12r)$. Note that $\text{dist}_{6r}(p', A) \geq 6r$, which implies that $\text{dist}(p', A) \leq 2 \cdot \text{dist}_{6r}(p', A)$. Therefore, $\text{dist}_r(p', T) \leq \text{dist}_r(p', A) \leq \text{dist}(p', A) \leq 2\text{dist}_{6r}(p', A)$. Therefore, it follows that for all $p \in P$, $\text{dist}_r(p, T) \leq 2 \cdot \text{dist}_{6r}(p, A)$. Which implies that,

$$\text{cost}_r(p, T) \leq 2 \cdot \text{cost}_r(p, A) \leq 36 \cdot \text{OPT}_r.$$

This concludes the proof of the lemma. ◀

6 Conclusion

In this paper, we revisit HYBRID k -CLUSTERING, which was introduced and studied recently by Fomin et al. [23]. We resolve two open questions explicitly asked in their work, namely, for continuous Euclidean instances from \mathbb{R}^d , (1) we obtain an FPT-AS for HYBRID k -CLUSTERING that does not have an FPT dependence on the dimension d (and in fact, the dependence on k is $2^{O(k \log k)}$ instead $2^{\text{poly}(k)}$ as in [23]), and (2) we design coresets for the same. Indeed, our technique also generalizes to the (k, z) -clustering variant of the problem, which was implicitly considered in [23], but was not completely resolved. To obtain our algorithmic result, we build upon insights from the recent framework of Abbasi et al. [1] for clustering in metric spaces of bounded *algorithmic scatter dimension* that encapsulates a broad class of metric spaces. Thus, our result shows that the potential of the framework introduced in [1] extends to clustering problems beyond that captured by the monotone norm setting, thus paving the way for obtaining FPT-ASes for other clustering problems using the similar technique. However, several interesting questions remain.

Firstly, the framework of [1] is inherently randomized due to key sampling steps, and its derandomization remains an intriguing open question. In particular, derandomizing the step that samples a witness in our algorithm is an interesting challenge. In another direction, now that the approximation landscape of the vanilla version of HYBRID k -CLUSTERING is beginning to be well-understood, it is natural to explore *constrained* variants of the problem such as those involving fairness or capacity constraints.

References

- 1 Fateme Abbasi, Sandip Banerjee, Jaroslav Byrka, Parinya Chalermsook, Ameet Gadekar, Kamyar Khodamoradi, Dániel Marx, Roohani Sharma, and Joachim Spoerhase. Parameterized approximation schemes for clustering with general norm objectives. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 1377–1399. IEEE, 2023. doi:10.1109/FOCS57990.2023.00085.

- 2 Fateme Abbasi, Sandip Banerjee, Jaroslaw Byrka, Parinya Chalermsook, Ameet Gadekar, Kamyar Khodamoradi, Dániel Marx, Roohani Sharma, and Joachim Spoerhase. Parameterized approximation for robust clustering in discrete geometric spaces. In Karl Bringmann, Martin Grohe, Gabriele Puppis, and Ola Svensson, editors, *51st International Colloquium on Automata, Languages, and Programming, ICALP 2024, July 8-12, 2024, Tallinn, Estonia*, volume 297 of *LIPICs*, pages 6:1–6:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPICs.ICALP.2024.6.
- 3 Akanksha Agrawal, Tanmay Inamdar, Saket Saurabh, and Jie Xue. Clustering what matters: Optimal approximation for clustering with outliers. *J. Artif. Intell. Res.*, 78:143–166, 2023. doi:10.1613/JAIR.1.14883.
- 4 Barbara Anthony, Vineet Goyal, Anupam Gupta, and Viswanath Nagarajan. A plant location guide for the unsure: Approximation algorithms for min-max location problems. *Mathematics of Operations Research*, 35(1):pages 79–101, 2010.
- 5 Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for euclidean k -medians and related problems. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98*, pages 106–113, New York, NY, USA, 1998. ACM. doi:10.1145/276698.276718.
- 6 Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for k -median and facility location problems. *SIAM J. Comput.*, 33(3):544–562, 2004. doi:10.1137/S0097539702416402.
- 7 Mihai Badoiu, Sarel Har-Peled, and Piotr Indyk. Approximate clustering via core-sets. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, (STOC)*, pages 250–257. ACM, 2002. doi:10.1145/509907.509947.
- 8 Sayan Bandyapadhyay, Fedor V. Fomin, and Kirill Simonov. On coresets for fair clustering in metric and euclidean spaces and their applications. *J. Comput. Syst. Sci.*, 142:103506, 2024. doi:10.1016/J.JCSS.2024.103506.
- 9 Sayan Bhattacharya, Parinya Chalermsook, Kurt Mehlhorn, and Adrian Neumann. New approximability results for the robust k -median problem. In *Scandinavian Workshop on Algorithm Theory (SWAT'14)*, pages 50–61. Springer, 2014. doi:10.1007/978-3-319-08404-6_5.
- 10 Romain Bourneuf and Marcin Pilipczuk. Bounding ε -scatter dimension via metric sparsity. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (to appear)*, 2025.
- 11 Vladimir Braverman, Shaofeng H-C Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for ordered weighted clustering. In *International Conference on Machine Learning (ICML'19)*, pages 744–753. PMLR, 2019. URL: <http://proceedings.mlr.press/v97/braverman19a.html>.
- 12 Jaroslaw Byrka, Krzysztof Sornat, and Joachim Spoerhase. Constant-factor approximation for ordered k -median. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 620–631. ACM, 2018. doi:10.1145/3188745.3188930.
- 13 Deeparnab Chakrabarty, Prachi Goyal, and Ravishankar Krishnaswamy. The non-uniform k -center problem. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPICs*, pages 67:1–67:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPICs.ICALP.2016.67.
- 14 Deeparnab Chakrabarty and Chaitanya Swamy. Interpolating between k -median and k -center: Approximation algorithms for ordered k -median. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 107 of *LIPICs*, pages 29:1–29:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPICs.ICALP.2018.29.
- 15 Deeparnab Chakrabarty and Chaitanya Swamy. Approximation algorithms for minimum norm and ordered optimization problems. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 126–137. ACM, 2019. doi:10.1145/3313276.3316322.

- 16 Vincent Cohen-Addad, Fabrizio Grandoni, Euiwoong Lee, and Chris Schwiegelshohn. Breaching the 2 LMP approximation barrier for facility location with applications to k -median. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 940–986. SIAM, 2023. doi:10.1137/1.9781611977554.CH37.
- 17 Vincent Cohen-Addad, Philip N. Klein, and Claire Mathieu. Local search yields approximation schemes for k -means and k -median in euclidean and minor-free metrics. *SIAM J. Comput.*, 48(2):644–667, 2019. doi:10.1137/17M112717X.
- 18 Vincent Cohen-Addad and Jason Li. On the fixed-parameter tractability of capacitated clustering. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPICs*, pages 41:1–41:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPICs.ICALP.2019.41.
- 19 Vincent Cohen-Addad, Michał Pilipczuk, and Marcin Pilipczuk. A polynomial-time approximation scheme for facility location on planar graphs. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 560–581. IEEE, 2019.
- 20 Vincent Cohen-Addad, David Saulpic, and Chris Schwiegelshohn. A new coresets framework for clustering. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 169–182. ACM, 2021. doi:10.1145/3406325.3451022.
- 21 Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshantov, Dániel Marx, Marcin Pilipczuk, Michał Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. doi:10.1007/978-3-319-21275-3.
- 22 Tomás Feder and Daniel H. Greene. Optimal algorithms for approximate clustering. In Janos Simon, editor, *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 434–444. ACM, 1988. doi:10.1145/62212.62255.
- 23 Fedor V. Fomin, Petr A. Golovach, Tanmay Inamdar, Saket Saurabh, and Meirav Zehavi. Hybrid k -clustering: Blending k -median and k -center. In Amit Kumar and Noga Ron-Zewi, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2024, August 28-30, 2024, London School of Economics, London, UK*, volume 317 of *LIPICs*, pages 4:1–4:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPICs.APPROX/RANDOM.2024.4.
- 24 Zachary Friggstad, Mohsen Rezapour, and Mohammad R. Salavatipour. Local search yields a PTAS for k -means in doubling metrics. *SIAM J. Comput.*, 48(2):452–480, 2019. doi:10.1137/17M1127181.
- 25 Mehrdad Ghadiri, Samira Samadi, and Santosh Vempala. Socially fair k -means clustering. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 438–448, 2021. doi:10.1145/3442188.3445906.
- 26 Sarel Har-Peled and Soham Mazumdar. On coresets for k -means and k -median clustering. In László Babai, editor, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 291–300. ACM, 2004. doi:10.1145/1007352.1007400.
- 27 Dorit S. Hochbaum and Wolfgang Maass. Approximation schemes for covering and packing problems in image processing and VLSI. *J. ACM*, 32(1):130–136, 1985. doi:10.1145/2455.214106.
- 28 Dorit S. Hochbaum and David B. Shmoys. A best possible heuristic for the k -center problem. *Mathematics of Operations Research*, 10(2):180–184, 1985. URL: <http://www.jstor.org/stable/3689371>, doi:10.1287/MOOR.10.2.180.
- 29 Kamal Jain, Mohammad Mahdian, and Amin Saberi. A new greedy approach for facility location problems. In John H. Reif, editor, *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 731–740. ACM, 2002. doi:10.1145/509907.510012.

- 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. *J. ACM*, 48(2):274–296, 2001. doi:10.1145/375827.375845.
- 31 Ravishankar Krishnaswamy, Shi Li, and Sai Sandeep. Constant approximation for k -median and k -means with outliers via iterative rounding. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 646–659. ACM, 2018. doi:10.1145/3188745.3188882.
- 32 Alfred A Kuehn and Michael J Hamburger. A heuristic program for locating warehouses. *Management science*, 9(4):643–666, 1963.
- 33 Amit Kumar, Yogish Sabharwal, and Sandeep Sen. Linear time algorithms for clustering problems in any dimensions. In *32nd International Colloquium on Automata, Languages and Programming (ICALP)*, volume 3580 of *Lecture Notes in Computer Science*, pages 1374–1385. Springer, 2005. doi:10.1007/11523468_111.
- 34 Amit Kumar, Yogish Sabharwal, and Sandeep Sen. Linear-time approximation schemes for clustering problems in any dimensions. *J. ACM*, 57(2):5:1–5:32, 2010. doi:10.1145/1667053.1667054.
- 35 Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Inf. Comput.*, 222:45–58, 2013. doi:10.1016/J.IC.2012.01.007.
- 36 S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982. doi:10.1109/TIT.1982.1056489.
- 37 Meena Mahajan, Prajakta Nimbhorkar, and Kasturi R. Varadarajan. The planar k -means problem is np-hard. *Theor. Comput. Sci.*, 442:13–21, 2012. doi:10.1016/J.TCS.2010.05.034.
- 38 Yury Makarychev and Ali Vakilian. Approximation algorithms for socially fair clustering. In *Conference on Learning Theory (COLT'21)*, pages 3246–3264. PMLR, 2021. URL: <http://proceedings.mlr.press/v134/makarychev21a.html>.
- 39 Dániel Marx. Parameterized complexity of independence and domination on geometric graphs. In Hans L. Bodlaender and Michael A. Langston, editors, *Parameterized and Exact Computation, Second International Workshop, IWPEC 2006, Zürich, Switzerland, September 13-15, 2006, Proceedings*, volume 4169 of *Lecture Notes in Computer Science*, pages 154–165. Springer, 2006. doi:10.1007/11847250_14.
- 40 Nimrod Megiddo and Kenneth J Supowit. On the complexity of some common geometric location problems. *SIAM journal on computing*, 13(1):182–196, 1984. doi:10.1137/0213014.
- 41 John F. Stollsteimer. A working model for plant numbers and locations. *Journal of Farm Economics*, 45(3):631–645, 1963. URL: <http://www.jstor.org/stable/1235442>.
- 42 Arie Tamir. The k -centrum multi-facility location problem. *Discret. Appl. Math.*, 109(3):293–307, 2001. doi:10.1016/S0166-218X(00)00253-5.
- 43 Murad Tukan, Xuan Wu, Samson Zhou, Vladimir Braverman, and Dan Feldman. New coresets for projective clustering and applications. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*, volume 151 of *Proceedings of Machine Learning Research*, pages 5391–5415. PMLR, 2022. URL: <https://proceedings.mlr.press/v151/tukan22a.html>.