Improved Coresets for Euclidean k-Means

Vincent Cohen-Addad
Google Research

Kasper Green Larsen
Aarhus University

David Saulpic
University of Vienna

Chris Schwiegelshohn
Aarhus University

Omar Ali Sheikh-Omar
Aarhus University

Abstract

Given a set of n points in d dimensions, the Euclidean k-means problem (resp. the Euclidean k-median problem) consists of finding k centers such that the sum of squared distances (resp. sum of distances) from every point to its closest center is minimized. The arguably most popular way of dealing with this problem in the big data setting is to first compress the data by computing a weighted subset known as a coreset and then run any algorithm on this subset. The guarantee of the coreset is that for any candidate solution, the ratio between coreset cost and the cost of the original instance is less than a $(1 \pm \varepsilon)$ factor. The current state of the art coreset size is $\tilde{O}(\min(k^2 \cdot \varepsilon^{-2}, k \cdot \varepsilon^{-4}))$ for Euclidean k-means and $\tilde{O}(\min(k^2 \cdot \varepsilon^{-2}, k \cdot \varepsilon^{-3}))$ for Euclidean k-median. The best known lower bound for both problems is $\Omega(k\varepsilon^{-2})$. In this paper, we improve the upper bounds $\tilde{O}(\min(k^{3/2} \cdot \varepsilon^{-2}, k \cdot \varepsilon^{-4}))$ for k-means and $\tilde{O}(\min(k^{4/3} \cdot \varepsilon^{-2}, k \cdot \varepsilon^{-3}))$ for k-median. In particular, ours is the first provable bound that breaks through the k^2 barrier while retaining an optimal dependency on ε.

1 Introduction

Coresets have become a staple in the design of algorithms for large data sets. In the most general setting, a coreset compresses the data set in such a way that for any set of previously specified candidate queries, the cost of evaluating the query and the cost of the coreset are similar, up to an arbitrary small distortion.

A popular subject in coreset literature is the Euclidean k-means problem. Here, we are given n points P in d dimensions and our task is to find a set of k points C called centers minimizing $\text{cost}(P, C) := \sum_{p \in P} \min_{c \in C} \|p - c\|^2$, where $\|p - c\| = \sqrt{\sum_{i=1}^{d} (p_i - c_i)^2}$ denotes the Euclidean distance. In this case, a coreset is a weighted subset of the input such that difference between the cost for any set of k centers C on the coreset and the cost on the original point set P is at most $\varepsilon \cdot \text{cost}(P, C)$. Since its initial study by Har-Peled and Mazumdar [2004], the Euclidean k-means problem has received arguably the most attention out of any coreset problem. The current state of the art by Cohen-Addad et al. [2022] yields coresets of size $\tilde{O}(k\varepsilon^{-2} \min(k, \varepsilon^{-2}))$, where $\tilde{O}(x)$ hides multiplicative factors that are polylogarithmic in x. Unfortunately, there is still a gap towards the best known lower bound of $\Omega(k\varepsilon^{-2})$ by Cohen-Addad et al. [2022].

We thus have the option of obtaining either an optimal dependency on k, at the cost of a suboptimal dependency on ε^{-1}, or an optimal dependency on ε^{-1}, at the cost of a suboptimal dependency on k. While these bounds suggest that the lower bound is the correct answer, things are not as clear on closer inspection. Quadratic dependencies on k become necessary for many forms of analysis and so far, it is unknown how to avoid this loss while retaining an optimal dependency on the remaining

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

parameters[1]. Moreover, the trade-off between the dependency on k and the dependency on ε^{-2} is natural. Specifically, if $k \approx \varepsilon^{-2}$ the two previous alternatives $\tilde{O}(k^2/\varepsilon^2)$ and $\tilde{O}(k/\varepsilon^4)$ are equal.

In this paper we break through this barrier. Specifically, we show that coresets of size $\tilde{O}(k^{1.5}\varepsilon^{-2})$ exist. In our view, this is further and arguably stronger evidence that the $k\varepsilon^{-2}$ bound will be the correct answer. Another contribution exists on a technical level. Previously, most coreset constructions for high dimensions heavily relied on terminal embeddings to facilitate the analysis. In this paper, we present a novel method that avoids terminal embeddings. We expect that our technique may have further applications for coreset constructions in Euclidean spaces.

1.1 Techniques

Starting point of our work is the framework introduced by Cohen-Addad et al. [2021a] and specifically Cohen-Addad et al. [2022]. Prior to Cohen-Addad et al. [2022], all coreset analyses required a dependency of at least $k \cdot d \cdot \varepsilon^{-2}$. To illustrate why, suppose we are sampling points from some distribution and wish to use the sampled points as an estimator for the true cost of any candidate solution. The analysis then consists of (1) a bound on the variance σ^2 of the estimator and (2) a bound on the number of solutions $|\mathbb{N}|$ to be approximated. This typically results in coresets of size $O(\varepsilon^{-2} \cdot \sigma^2 \cdot \log |\mathbb{N}|)$. When enumerating all (discretized) candidate solutions (henceforth called a *net*) in d dimensions, virtually all known techniques result in $|\mathbb{N}| \approx \exp(k \cdot d)$. The dependency on d may be reduced to $\log(k/\varepsilon)\varepsilon^{-2}$ using dimension reduction techniques.

To bypass this, Cohen-Addad et al. [2022] used a chaining-based analysis to define a sequence of discretized candidate solutions. Specifically, they showed that there exist discretizations $\mathbb{N}_\alpha$ of size $\exp(k\alpha^{-2})$ such that for any point p and any solution $\mathcal{S}$, there exists a solution $\mathcal{S}_\alpha$ with

$$|\text{cost}(p, \mathcal{S}_\alpha) - \text{cost}(p, \mathcal{S})| \leq \alpha \cdot \text{cost}(p, \mathcal{S}),$$

where $\text{cost}(p, \mathcal{S}_\alpha) = \min_{s \in \mathcal{S}_\alpha} \|p - s\|^2$ and $\text{cost}(p, \mathcal{S}) = \min_{s \in \mathcal{S}} \|p - s\|^2$. The idea of the analysis is then to write $\text{cost}(p, \mathcal{S})$ as a telescoping sum

$$\text{cost}(p, \mathcal{S}) = \sum_{h=1}^{\infty} \text{cost}(p, \mathcal{S}_{2^{-(h+1)}}) - \text{cost}(p, \mathcal{S}_{2^{-h}})$$

and show that the sampled points achieve concentration for each summand. The number of candidate solutions is now $|\mathbb{N}_{2^{-(h+1)}}| \cdot |\mathbb{N}_{2^{-h}}| \approx \exp(k \cdot 2^{2h})$ but the difference in cost is $2^{-h}\text{cost}(p, \mathcal{S})$. This directly leads to a variance of the order $2^{-2h}\sigma^2$, where σ^2 is the variance of the basic estimator. Thus, the increase in net size is countered by the decrease and variance. Using this technique, Cohen-Addad et al. [2022] obtained a coreset of size roughly $\varepsilon^{-2}k\sigma^2$.

To obtain a bound on σ^2, the framework by Cohen-Addad et al. [2021a] proposed an algorithm that first computes a solution $\mathcal{A}$ with strong structural properties and then samples any point p proportionate $\text{cost}(p, \mathcal{A})$. To simplify the exposition, we assume that $\mathcal{A}$ is the optimum, every cluster has identical cost and every point has identical cost. For this special case, the distribution of Cohen-Addad et al. [2021a] turns out to be equivalent to uniform sampling. For any given solution $\mathcal{S}$, let k_i be the number of clusters of $\mathcal{A}$ whose points are served at cost 2^i times their cost in $\mathcal{A}$ for $i > 2^2$, i.e. $2^i = \frac{\text{cost}(p,\mathcal{S})}{\text{cost}(p,\mathcal{A})}$. Cohen-Addad et al. [2021a] showed that their sampling distribution leads to a variance of the order $\sigma^2 \approx \left(\frac{k \cdot k_i 2^i}{(k + k_i \cdot 2^i)^2}\right) \cdot \min(\varepsilon^{-2}, 2^i, k)$, which yields the $\tilde{O}(k\varepsilon^{-2}\min(k, \varepsilon^{-2}))$ bound.

To improve either the variance or the net size has to reduced. Unfortunately, is unlikely that reducing σ^2 will be possible as the bounds on σ^2 obtained by Cohen-Addad et al. [2022] are tight up to constant factors. Our main goal is to find a net with an finer error of

$$|\text{cost}(p, \mathcal{S}_\alpha) - \text{cost}(p, \mathcal{S})| \leq 2^{-h} \cdot \sqrt{\text{cost}(p, \mathcal{S}), \text{cost}(p, \mathcal{A})}.$$

In the case of $\min(2^i, \varepsilon^{-2}, k) = 2^i$, this leads to a reduced variance for the h-th summand of the order $2^{-2h} \cdot \frac{(\sqrt{\text{cost}(p,\mathcal{S})\text{cost}(p,\mathcal{A})})^2}{\text{cost}(p,\mathcal{A})^2} \cdot \left(\frac{k \cdot k_i}{(k + k_i \cdot 2^i)^2}\right) \leq 2^{-2h} \cdot \left(\frac{k \cdot k_i \cdot 2^i}{(k + k_i \cdot 2^i)^2}\right)$. To find such a net, we now

[1]Coresets of size $\tilde{O}(kd/\varepsilon^2)$ are also known (Cohen-Addad et al. [2021a]). This offers improvements in low dimensions, but generally a dependency on d is considered worse than a dependency on k or ε^{-2}.

[2]Observe that if one points of a cluster C_j of $\mathcal{A}$ costs 2^i times more for $i \geq 3$, then all points from C_j do likewise (up to constant factors).

have two options. First, we can essentially use the previous nets and rescale 2^{-h} by a factor $2^{-i/2}$. Unfortunately, this leads to nets of size $\exp(k \cdot 2^{2h} 2^i)$, so any gain in reducing the variance is countered by an increase in the net size.

The novelty in our approach now lies in showing that a net of size $\exp(k \cdot k_i \cdot 2^{2h})$ exists. Combining the two net sizes with the improved variance bound results in a coreset of size roughly $\varepsilon^{-2} \log\left(\exp(k \cdot \min(k_i, 2^i) \cdot 2^{2h})\right) \cdot 2^{-2h} \cdot \left(\frac{k \cdot k_i \cdot 2^i}{(k+k_i \cdot 2^i)^2}\right)$, which after some calculation is shown to be of the order $O(k^{1.5}\varepsilon^{-2})$.

For the remainder of this section we will illustrate how these improved nets can be obtained. The key idea already appears in the case of a single center. Suppose that s is a candidate center. Then we can show that there always exists a subspace with orthogonal basis U spanned by a set T of at most $O(\alpha^{-2})$ points of P such that

$$|p^T (I - UU^T)s| \leq \alpha \cdot \|(I - UU^T)p\| \cdot \|(I - UU^T)s\|.$$

We can then write

$$\|p - s\|^2 = \|U(p - s)\|^2 + \|(I - UU^T)p\|^2 + \|(I - UU^T)s\|^2 - 2p^T(I - UU^T)s.$$

The error for the terms $\|Up - s\|^2$ and $\|(I - UU^T)s\|^2$ can be made negligibly small using a net of size $\exp(\alpha^{-2})$. Thus the main loss in error comes from the $p^T(I - UU^T)s$ term. The key insight is that when adding the center c_i of point p in solution $\mathcal{A}$, we have $\|(I - UU^T)p\|^2 \leq \sqrt{\text{cost}(p, \mathcal{A})}$. Thus, when adding all the k_i centers of clusters that cost 2^i more in $\mathcal{S}$ than in $\mathcal{A}$ to T the net size becomes $\exp(\alpha^{-2} + k_i)$. By composing these nets for k candidate centers, we then obtain our desired bound.

1.2 Related Work

There has been a tremendous amount of work on coresets for Euclidean k-means following the work in Bachem et al. [2018a,b], Baker et al. [2020], Bandyapadhyay et al. [2021], Becchetti et al. [2019], Braverman et al. [2022, 2021b], Chen [2009], Cohen-Addad and Li [2019], Cohen-Addad et al. [2021a,b], Feldman and Langberg [2011], Feldman et al. [2020], Feng et al. [2021], Fichtenberger et al. [2013], Har-Peled and Kushal [2007], Har-Peled and Mazumdar [2004], Huang and Vishnoi [2020], Huang et al. [2018, 2019], Langberg and Schulman [2010], Schmidt et al. [2019], Schwiegelshohn and Sheikh-Omar [2022], Sohler and Woodruff [2018]. Almost as prolific is the catalogue of work on dimension reduction for clustering problems in Euclidean spaces, see Boutsidis et al. [2009, 2010, 2015], Charikar and Waingarten [2022a,b], Cohen et al. [2015], Cohen-Addad and Schwiegelshohn [2017], Drineas et al. [2004], Feng et al. [2019], Makarychev et al. [2019]. The arguably most important dimension reduction technique for coresets are terminal embeddings, see Cherapanamjeri and Nelson [2021], Elkin et al. [2017], Mahabadi et al. [2018], Narayanan and Nelson [2019].

Further work on coresets considering objects other than points as centers Braverman et al. [2021a], Feldman et al. [2010], Huang et al. [2021] or other objectives all together Boutsidis et al. [2013], Huang et al. [2020], Jiang et al. [2021], Karnin and Liberty [2019], Mai et al. [2021], Molina et al. [2018], Munteanu et al. [2018], Phillips and Tai [2020], Tukan et al. [2020]. For further reading, we refer the interested reader to recent surveys Feldman [2020], Munteanu and Schwiegelshohn [2018].

2 Preliminaries and Setup

First, we require the following basic notions. For a point $p \in \mathbb{R}^d$, we denote $\|p\|_2 = \sqrt{\sum_{i=1}^d p_i^2}$ to be the Euclidean norm of p and $\|p\|_1 = \sum_{i=1}^d |p_i|$. The distinct number of points in a point set P is denoted by $\|P\|_0$. Note that the true number of points $|P|$ may be larger than $\|P\|_0$ as different points may lie on the same coordinates. Given a solution $\mathcal{S}$ consisting of at most k centers, and any subset $P' \subset P$ we use $\text{cost}(P', \mathcal{S}) := \sum_{p \in P'} \text{cost}(p, \mathcal{S}) = \sum_{p \in P} \min_{s \in \mathcal{S}} w_p \text{cost}(p, s)$, where $\text{cost}(p, s) = \|p - s\|^2$ for Euclidean k-means and $\text{cost}(p, s) = \|p - s\|$ for Euclidean k-median and w_p is a non-negative weight (in the basic case this simply 1 whereas for the coreset it can be any non-negative number). To unify the notation, we will often write $\text{cost}(p, s) = \|p - s\|^z$ where $z = 1$ corresponds to k-median and $z = 2$ corresponds to k-means. We also denote by $v^{\mathcal{S}} \in \mathbb{R}^{\|P\|_0}$ the cost vector associated with the point set P and solution $\mathcal{S}$, that is $v_p^{\mathcal{S}} := w_p \text{cost}(p, s)$. Note that

$\|v^{\mathcal{S}}\|_1 = \text{cost}(P, \mathcal{S})$. The classic coreset guarantee is to show that for any solution $\mathcal{S}$ the designated coreset Ω satisfies

$$|\text{cost}(\Omega, \mathcal{S}) - \text{cost}(P, \mathcal{S})| \leq \varepsilon \cdot \text{cost}(P, \mathcal{S}).$$

We will later introduce an equivalent statement that uses cost vectors. It will also be convenient to consider coresets with an additive error E which satisfy

$$|\text{cost}(\Omega, \mathcal{S}) - \text{cost}(P, \mathcal{S})| \leq \varepsilon \cdot \text{cost}(P, \mathcal{S}) + E.$$

Cohen-Addad et al. [2022] showed that any coreset algorithm that works for instances with the following assumptions can be extended to general instances:

Assumption 1: $\|P\|_0 \in \text{poly}(k, \varepsilon^{-1})$.

Assumption 2: $d \in O(\log(k/\varepsilon) \cdot \varepsilon^{-2})$.

Assumption 3: $w_p = 1$, for all $p \in P$. Note that this only applies to the weights of the original points; the coreset points will have different weights.

Assumption 4: There exists a solution $\mathcal{A}$ such that

1. $|\mathcal{A}| \in O(k)$.
2. For any two clusters C_i, C_j induced by $\mathcal{A}$, $\text{cost}(C_i, \mathcal{A}) \leq 2 \cdot \text{cost}(C_j, \mathcal{A})$.
3. For any cluster C_j induced by $\mathcal{A}$ and any two points $p, p' \in C_j$, $\text{cost}(p, \mathcal{A}) \leq 2 \cdot \text{cost}(p', \mathcal{A})$

To keep this paper self contained, we will detail the validity of these assumptions at the end of this section.

The sampling procedure is now very simple. Given that these aforementioned assumptions hold, we sample a points $p \in C_j$ with probability $\mathbb{P}_p := \frac{1}{|C_j|} \cdot \frac{\text{cost}(C_j, \mathcal{A})}{\text{cost}(P, \mathcal{A})}$ and add it to the designated coreset Ω. Furthermore, p receives the weight $w_p = \frac{1}{\mathbb{P}_p}$. Overall, our basic cost estimator for any candidate solution $\mathcal{S}$ is therefore

$$\text{cost}(\Omega, \mathcal{S}) := \frac{1}{|\Omega|} \sum_{p \in \Omega} \text{cost}(p, \mathcal{S}) \cdot w_p.$$

It is routine to check that $\mathbb{E}[\text{cost}(\Omega, \mathcal{S})] = \text{cost}(P, \mathcal{S})$. The remainder of this section will be devoted to showing that Ω satisfies for all $\mathcal{S}$

$$|\text{cost}(\Omega, \mathcal{S}) - \text{cost}(P, \mathcal{S})| \leq \frac{\varepsilon}{\log^2 \varepsilon^{-1}} \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A})) \tag{1}$$

Using the framework from Cohen-Addad et al. [2021a], this implies an $O(\varepsilon)$ coreset in general.

2.1 Justification of the Assumptions

To obtain the first assumption, we compute any coreset of size $\text{poly}(k, \varepsilon^{-1})$ in preprocessing. Constructions of these coresets are abundant in literature and any one would serve our needs.

To obtain the second assumption, we apply a terminal embedding on the coreset. A terminal embedding guarantees that for any point $p \in P$ and any point $q \in \mathbb{R}^d$, where d is the dimension of the points of P, we have a mapping f s.t.

$$\|p - q\|^2 = (1 \pm \varepsilon)\|f(p) - f(q)\|^2.$$

Narayanan and Nelson [2019] showed that for any n-point set a terminal embedding of target dimension $\tilde{O}(\varepsilon^{-2} \log n)$ exists, which, combined with the first assumption, yields the desired target dimension.

To obtain the third assumption, we merely have to ensure that the weights of the coreset points are integers. A number of constructions satisfy this but a simple way of always enforcing this is to scale and round the weights (see Corollary 2 of Cohen-Addad et al. [2021a]).

The fourth assumption follows from the preprocessing of Cohen-Addad et al. [2021a], see Sections 3.3 and 4.1 of that reference. Similarly, the same preprocessing, given that $\mathcal{A}$ is an $O(1)$-approximation, also shows that Eq 1 implies that the overall construction will be a coreset (subject to rescaling ε by constant factors), see Section 4.2 of the aforementioned reference. We must point out that a point set cannot always be decomposed into only sets that satisfy the aforementioned assumption. Nevertheless Cohen-Addad et al. [2022] showed that every other case require only $\tilde{O}(k/\varepsilon^2)$ many sampled points (compared Lemmas 15 and 17 of that reference.)

Finally, we remark that these steps and assumptions immediately also apply to the k-median problem.

3 Analysis

In this section we prove the following theorems.

Theorem 1. *For any set of points in d dimensional Euclidean space, there exists a coreset for k-means clustering of size $\tilde{O}(k^{3/2}\varepsilon^{-2})$.*

Theorem 2. *For any set of points in d dimensional Euclidean space, there exists a coreset for k-median clustering of size $\tilde{O}(k^{4/3}\varepsilon^{-2})$.*

If not remarked upon, the analysis will holds for both problems.

We first describe the random process used to show concentration of the estimator.

3.1 Setting up the Chaining Analysis

First, we observe that Eq.1 is equivalent to showing

$$\sup_{\mathcal{S}} \frac{|\text{cost}(\Omega, \mathcal{S}) - \|v\|_1|}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \leq \frac{\varepsilon}{\log^2 \varepsilon^{-1}}.$$

Our goal is to show that

$$\mathbb{E}_\Omega \left[\sup_{\mathcal{S}} \frac{|\text{cost}(\Omega, \mathcal{S}) - \|v\|_1|}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))}\right] \leq \frac{\varepsilon}{\log^2 \varepsilon^{-1}},$$

where $\mathbb{E}_\Omega$ is meant to denote the expectation over the randomness of Ω. This implies that the desired guarantee holds with constant probability.

We now apply a standard symmetrization argument.

Lemma 1 (Appendix B.3 of Rudra and Wootters [2014])**.** *Let g_p be independent standard Gaussian random variables. Then.*

$$\mathbb{E}_\Omega \mathit{sup}_{\mathcal{S}} \left[\left|\frac{\frac{1}{|\Omega|}\sum_{p\in\Omega} \mathit{cost}(p, \mathcal{S}) \cdot w_p - \|v\|_1}{(\mathit{cost}(P, \mathcal{S}) + \mathit{cost}(P, \mathcal{A}))}\right|\right] \leq \sqrt{2\pi}\mathbb{E}_\Omega \mathbb{E}_g \mathit{sup}_{\mathcal{S}} \left[\left|\frac{\frac{1}{|\Omega|}\sum_{p\in\Omega} \mathit{cost}(p, \mathcal{S}) \cdot w_p \cdot g_p}{(\mathit{cost}(P, \mathcal{S}) + \mathit{cost}(P, \mathcal{A}))}\right|\right].$$

It is therefore sufficient to show

$$\mathbb{E}_\Omega \mathbb{E}_g \sup_{\mathcal{S}} \left[\left|\frac{\frac{1}{|\Omega|}\sum_{p\in\Omega} \text{cost}(p, \mathcal{S}) \cdot w_p \cdot g_p}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))}\right|\right] \leq \frac{\varepsilon}{\sqrt{2\pi}\log^2 \varepsilon^{-1}}. \tag{2}$$

Let $z = 1$ for Euclidean k-median and 2 for Euclidean k-means. We partition the clusters of any solution $\mathcal{S}$ by type. We consider a cluster C_j of type T_i if for

$$2^i \min_{p\in C_j} \text{cost}(p, \mathcal{A}) \leq \min_{p\in C_j} \min_{s\in\mathcal{S}} \text{cost}(p, \mathcal{S}) \leq 2^{i+1} \min_{c\in\mathcal{A}} \text{cost}(p, \mathcal{A}).$$

The number of clusters $C_j \in T_i$ are denoted by k_i. If C_j is of type $i \leq 3$, we say C_j is of type T_{small} and if C_j is of type $i \geq \log \gamma\varepsilon^{-z}$, for a sufficiently large absolute constant γ, we say that C_j is of type T_{large}. Then, we show

$$\mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{\mathcal{S}} \left|\frac{\frac{1}{|\Omega|}\sum_{C_j\in T_{small}} \sum_{p\in C_j\cap\Omega} \text{cost}(p, \mathcal{S}) w_p \cdot g_p}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))}\right|\right] \leq \frac{\varepsilon}{\sqrt{2\pi}\log^3 \varepsilon^{-1}} \tag{3}$$

$$\mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{\mathcal{S}} \left|\frac{\frac{1}{|\Omega|}\sum_{C_j\in T_i} \sum_{p\in C_j\cap\Omega} \text{cost}(p, \mathcal{S}) w_p \cdot g_p}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))}\right|\right] \leq \frac{\varepsilon}{\sqrt{2\pi}\log^3 \varepsilon^{-1}} \tag{4}$$

$$\mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{\mathcal{S}} \left|\frac{\frac{1}{|\Omega|}\sum_{C_j\in T_{large}} \sum_{p\in C_j\cap\Omega} \text{cost}(p, \mathcal{S}) w_p \cdot g_p}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))}\right|\right] \leq \frac{\varepsilon}{\sqrt{2\pi}\log^3 \varepsilon^{-1}} \tag{5}$$

Note that if Equation 4 holds for $i \in \{3, \ldots, \log 1/\varepsilon\}$, this also implies Equation 2, as the error from each type can only sum up in the worst case and there are at most $O(\log \varepsilon^{-1})$ many types.

The small and large types are comparatively simple to handle.

Lemma 2 (Lemmas 15 and 16 of Cohen-Addad et al. [2022])**.** *Let $|\Omega| \geq \kappa \frac{k}{\varepsilon^2} \log^{10}(k/\varepsilon)$ for some absolute constant κ. Then Equations 3 and 5 hold.*

Our main objective will be to prove the following lemma.

Lemma 3. *Let $|\Omega| \geq \kappa_1 \frac{k^{1+z/(z+2)}}{\varepsilon^2} \log^{10}(k/\varepsilon) \geq \kappa_2 \frac{k}{\varepsilon^2} \log^{10}(k/\varepsilon) \cdot \left(\frac{\min(k_i, 2^i) \cdot 2^i k \cdot k_i}{(k + k_i \cdot 2^i)^2}\right)$ for some absolute constants κ_1 and κ_2. Then Equation 4 holds.*

Combining Lemma 2 and Lemma 3 then implies Theorem 1.

3.2 Proof of Lemma 3

The proof of Lemma 3 mainly consists of defining a nested sequence of nets over cost vectors over which we apply a union bound. Roughly speaking, for any cost vector $v^{\mathcal{S}}$, we aim to find an approximating cost vector v' such that

$$|v_p^{\mathcal{S}} - v'_p| \leq \varepsilon \cdot \sqrt{\text{cost}(p, \mathcal{S})^{z-1} \cdot \text{cost}(p, \mathcal{A})^{3-z}}.$$

Thus, on closer inspection, we have an error proportionate to $\varepsilon \cdot \sqrt{\text{cost}(p, \mathcal{S}) \cdot \text{cost}(p, \mathcal{A})}$ for k-means and $\varepsilon \cdot \text{cost}(p, \mathcal{S})$ for k-median.

This analysis differs from the terminal-embedding-based nets one used in Cohen-Addad et al. [2022], which aimed for an error of the order $\varepsilon \cdot \text{cost}(p, \mathcal{S})$.

Suppose we have, for every ε, a suitable collection of approximating cost vectors $\mathbb{N}_{\log 1/\varepsilon}$ with this guarantee for any candidate $\mathcal{S}$[3]. Let $v^{\mathcal{S},\varepsilon}$ be the cost vector approximating $v^{\mathcal{S}}$ in the net $\mathbb{N}_{\log 1/\varepsilon}$. Then we can write

$$v_p^{\mathcal{S}} = \sum_{h=0}^{\infty} v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}},$$

with $v_p^{\mathcal{S},1} = 0$. Our goal is to now bound

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{\mathcal{S}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} \text{cost}(p, \mathcal{S}) w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] \\
= \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S}}} \left| \frac{\sum_{h=0}^{\infty} \sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}}) w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] \\
\leq \; & \sum_{h=0}^{\infty} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1} - v^{\mathcal{S},h} \in \mathbb{N}_{2^{-(h+1)}} \times \mathbb{N}_{2^{-h}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}}) w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] \\
= \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} v_p^{\mathcal{S}, 2^{-1}} w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] && (6) \\
+ \; & \sum_{h=1}^{\log \varepsilon^{-2}} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1} - v^{\mathcal{S},h} \in \mathbb{N}_{h+1} \times \mathbb{N}_h} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}}) w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] && (7) \\
+ \; & \sum_{h=\log \varepsilon^{-2}}^{\infty} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1} - v^{\mathcal{S},h} \in \mathbb{N}_{h+1} \times \mathbb{N}_h} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}}) w_p \cdot g_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} \right| \right] && (8)
\end{aligned}$$

We will bound Equations 6 and 8 directly. For the $O(\log \varepsilon^{-1})$ equations in term 7, we prove a bound on each. Thus, we aim for a bound of the order $O(\frac{\varepsilon}{\log^3 \varepsilon^{-1}})$; the overall bound then follows by

[3] The reason for indexing the net by $\mathbb{N}_{\log 1/\varepsilon}$ and not by $\mathbb{N}_\varepsilon$ is to conveniently sum over $\sum_{i=1}^{\infty} \log |N_i|$, rather than $\sum_{i=1}^{\infty} \log |N_{2^i}|$.

summing up the errors and rescaling by constant factors. Technically, bounding each of the terms in Equations 6, 7 and 8 requires somewhat different arguments. For the sake of illustrating the key new ideas we first focus on Eq. 7.

The next section presents the nets for the cost vectors. The subsequent section bounds the variance. The final section combines these results and completes the proof of Lemma 3.

Cost Vector Nets

Definition 1. *Let I be a metric space, P a set of points, k a positive integer, and let $\alpha > 0$ be a precision parameters and let $\mathcal{A}$ be some solution with at most k' centers. Let $\mathbb{C} \subset I^k$ be a (potentially infinite) set of candidate k-clusterings. We say that a set of cost vectors $\mathbb{N} \subset \mathbb{R}^{|P|}$ is an (α, k)-clustering net if for every $\mathcal{S} \in \mathbb{C}$ there exists a vector $v' \in \mathbb{N}$ such that the following condition holds. For all $p \in P$,*

$$|v_p^{\mathcal{S}} - v'_p| \leq \alpha \cdot \sqrt{cost(p, \mathcal{S})^{z-1} \cdot cost(p, \mathcal{A})^{3-z}}.$$

These clustering nets have a substantially smaller error than those proposed in Cohen-Addad et al. [2022], which had an error of the order $\alpha \cdot (\text{cost}(p, \mathcal{S}) + \text{cost}(p, \mathcal{A}))$.

Given a set of points X in Euclidean space, an ε-net is a subset $S \subset X$ such that for every $p \in X$ there exists a q in S with $\|p - q\| \leq \varepsilon$. Throughout this section, we will frequently use the fact that in d dimensions, there exists an ε-net of cardinality $(1 + 2/\varepsilon)^d$ (see for example Pisier [1999]). Our main goal in this section is to prove the following lemma.

Lemma 4. *Let P be a set of points in d dimensional Euclidean space, k a positive integer, $\mathcal{A}$ be a candidate solution with k_i clusters and γ and absolute constant. Define $\mathbb{C}$ to be the set of possible candidate centers such that the clusters induced by $\mathcal{A}$ are of type i, with $3 \leq i \leq \log 1/\varepsilon^z$. For all $\alpha \leq 1/2$, there exists an (α, k)-clustering net $\mathbb{N}$ of $\mathbb{C}$ with*

$$|\mathbb{N}| \leq \exp\left(\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i + \alpha^{-2}, \alpha^{-2} \cdot 2^i) \cdot i \log \frac{1}{\alpha})\right).$$

Proof. We first show that given a set of vectors P and any vector s, there always exists a small subset U of P such that all inner products between $p \in P$ and s are preserved by the span of U.

Lemma 5. *Let $P = \{p_1, ..., p_n\} \subseteq \mathbb{R}^d$ and let $s \in \mathbb{R}^d$. Then there exists $U \subseteq P$, with $|U| = O(\varepsilon^{-2})$ and orthogonal basis Π_U, such that*

$$\forall p \in P, |p^T (I - \Pi_U \Pi_U^T) s| \leq \varepsilon \|(I - \Pi_U \Pi_U^T) p\| \cdot \min_{p \in P} \|p - s\| \tag{9}$$

Proof. Start with $U_0 = \text{argmin}_{p \in P} \|p - s\|$, and proceed in rounds. Note that $\|(I - \Pi_{U_0} \Pi_{U_0}^T) s\| \leq \|p - s\|$ for all $p \in P$.

In each round i, denote the current set of vectors U_i with orthogonal basis Π_{U_i}. We add a vector p_i if the following equation holds

$$|p^T (I - \Pi_{U_i} \Pi_{U_i}^T) s| \geq \varepsilon \|(I - \Pi_{U_i} \Pi_{U_i}^T) p\| \cdot \|(I - \Pi_{U_0} \Pi_{U_0}^T) s\|.$$

We observe that if this equation holds for all $p \in P$, then Equation 9 must also hold. Note that $(I - \Pi_{U_i} \Pi_{U_i}^T) p$ is orthogonal to the span of U_i of all previously added vectors. Thus, due to the Pythagorean theorem, we have

$$\sum_i^t \left(\frac{(p^T (I - \Pi_{U_{i-1}} \Pi_{U_{i-1}}^T) s)}{\|(I - \Pi_{U_{i-1}} \Pi_{U_{i-1}}^T) p\| \cdot \|(I - U_0 U_0^T) s\|} \right)^2 \geq t \cdot \varepsilon^2.$$

Therefore, after $t = \varepsilon^{-2}$ many rounds $(I - \Pi_U \Pi_U^T) s = 0$, which implies that after at most ε^{-2} rounds Eq. 9 has to hold. □

With this lemma, we can prove our net bound. Our objective is to generate a small set of cost vectors that satisfy the desired guarantee. Throughout this proof, let $\text{dist}(p, \mathcal{A}) = \text{cost}(p, \mathcal{A})^{1/z}$ be the distance of p to its center in $\mathcal{A}$. We first define the cost vectors. For each subset U of size

$O(\min(\alpha^{-2}2^i, \alpha^{-2} + k_i)$, we consider the the subspace Π_U spanned by U. In this subspace we consider $(\alpha/2^i) \cdot \text{dist}(p, \mathcal{A})$-nets of every ball centered around $\Pi_U p$ with radius $60 \cdot 2^i/2 \cdot \text{dist}(p, \mathcal{A})$ for all $p \in P$. Such a net has size $\exp(\gamma \cdot \textbf{rank}(U) i \log \alpha)$, for some constant γ and there exist at most $\|P\|_0 \cdot \exp(\gamma \cdot |U| i \log \alpha)$ many such nets. Furthermore, there are at most $\binom{\|P\|_0}{|U|} \leq \|P\|_0^{|U|}$ such subsets.

Now, for every point p, define an exponential sequence $\alpha^2(1 + \alpha/2^i)^j$ for $j \in \{0, \dots \log 10 \cdot 2^i\}$. There exist at most $\|P\|_0$ such sequences and every such sequence consists of at most $O(\alpha^{-1} \cdot 2^i \cdot i)$ many values. We combine every net point in ever ball of every subspace with all values in the exponential sequence to obtain the evaluation for a single candidate center. The overall number of candidate centers is therefore of the order $\|P\|_0^{|U|} \cdot \exp(\gamma \cdot |U| i \log \alpha)$, for a sufficiently large γ. The overall number of candidate cost vectors is now the number of k subsets of candidate centers, i.e. $\|P\|_0^{k \cdot |U|} \cdot \exp(\gamma \cdot k \cdot |U| i \log \alpha)$. Combined with the bounds on U, this yields the desired size. What remains to be shown is that the thus constructed cost vectors are a (α, k)-clustering net.

Here, we use that for any center s in some candidate solution $\mathcal{S}$

$$\|p - s\|^2 = \|\Pi_U(p - s)\|^2 + \|(I - \Pi_U \Pi_U^T)p\|^2 + \|(I - \Pi_U \Pi_U^T)s\|^2 - 2p^T(I - \Pi_U \Pi_U^T)s.$$

The nets for the span of Π_U are so fine that the distance $\|\Pi_U s - s'\|^2$ is essentially negligible compared to the maximum error incurred by $2p^T(I - \Pi_U \Pi_U^T)s$, where s' is the point in the span of Π_U closest to $\Pi_U s$ and the same holds for the exponential sequence approximating the term $\|(I - \Pi_U \Pi_U^T)s\|^2$. Thus, the error is dominated by $2p^T(I - \Pi_U \Pi_U^T)s$. Now, we can assume that the input point closest to s is included in U. Then $\min_{p \in P} \|p - s\| \leq O(1) \cdot 2^{i/z} \cdot \text{cost}(p, \mathcal{A})^{1/z}$ and $\|(I - \Pi_U \Pi_U^T)p\| \leq \text{cost}(p, \mathcal{S})^{1/z} \leq O(1) \cdot 2^{i/z} \text{cost}(p, \mathcal{A})^{1/z}$. If $\alpha^{-2} \cdot 2^i < k_i + \alpha^{-2}$, we have

$$|p^T(I - \Pi_U \Pi_U^T)s| \leq \alpha \cdot 2^{-i/2} \cdot \|(I - \Pi_U \Pi_U^T)p\| \cdot \min_{p \in P} \|p - s\| \leq O(1) \cdot \alpha \cdot \text{cost}(p, \mathcal{S})^{z-1} \text{cost}(p, \mathcal{A})^{3-z}$$

otherwise we have $\|(I - \Pi_U \Pi_U^T)p\| \leq \text{cost}(p, \mathcal{A})^{1/z}$ which implies

$$|p^T(I - \Pi_U \Pi_U^T)s| \leq \alpha \cdot \|(I - \Pi_U \Pi_U^T)p\| \cdot \min_{p \in P} \|p - s\| \leq \alpha \cdot \text{cost}(p, \mathcal{S})^{z-1} \text{cost}(p, \mathcal{A})^{3-z}.$$

Rescaling α by constant factors yields the claim. □

We also require an additional net that works for low dimensions.

Lemma 6 (Compare Lemma 22 of Cohen-Addad et al. [2022]). *Let P be a set of points in d dimensional Euclidean space, k a positive integer and $\mathcal{A}$ be a candidate solution. Define $\mathbb{C}$ to be the set of possible candidate centers such that the clusters induced by $\mathcal{A}$ are of type i, with $3 \leq i \leq \log 1/\varepsilon^2$. For all $\alpha \leq 1/2$, there exists an (α, k)-clustering net $\mathbb{N}$ of $\mathbb{C}$ with*

$$|\mathbb{N}| \leq \exp\left(\gamma \cdot k \cdot d \cdot i \log(4/\alpha)\right),$$

where γ is an absolute constant.

Proof. The only difference to Lemma 22 of Cohen-Addad et al. [2022] is that the nets are required to have an error of $\alpha \cdot \sqrt{\text{cost}(p, \mathcal{S}) \text{cost}(p, \mathcal{A})}$ rather than $\alpha \cdot (\text{cost}(p, \mathcal{S}) + \text{cost}(p, \mathcal{A}))$. This can be done by rescaling ε by 2^{-i}, which in turn is absorbed by the constant γ as $2^i \leq O(1) \cdot \varepsilon^{-2}$. □

Bounding the Variance

We now use the cost vectors to obtain an improved variance for the estimator $\frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S}, 2^{-(h+1)}} - v_p^{\mathcal{S}, 2^{-h}}) w_p}{(\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} g_p$. The bounds on variance for any random variable $\sum a_p g_p$ with standard Gaussians g_p is Gaussian distributed with mean 0 and variance $\sum a_p^2$.

Before we do this, we require an additional notion. Let $\mathcal{E}$ denote the event that $\frac{1}{|\Omega|} \sum_{p \in C_j \cap \Omega} w_p = (1 \pm \varepsilon) \cdot |C_j|$. The following lemma bounds the probability of $\mathcal{E}$ occurring.

Lemma 7. *[Compare Lemma 19 of Cohen-Addad et al. [2022]] If Assumption 4 holds, then event $\mathcal{E}$ holds with probability $1 - k^{-2}$ if $|\Omega| > \kappa \cdot k \varepsilon^{-2} \log k$ for a sufficiently high absolute constant κ.*

Lemma 8. *Given Assumption 4, the variance of* $\frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p\cdot g_p}{|\Omega|\cdot(cost(P,\mathcal{S})+cost(P,\mathcal{A}))}$ *is at most*

$$\gamma\cdot\frac{2^{-2h}}{|\Omega|}\cdot\frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}\qquad \textit{conditioned on event } \mathcal{E}$$
$$\gamma\cdot\frac{2^{-2h}\cdot k}{|\Omega|}\cdot\frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}\qquad \textit{conditioned on event } \overline{\mathcal{E}}$$

for an absolute constant γ.

Proof. We first observe that since the g_p are standard normal Gaussians, the entire estimator is Gaussian distributed with variance

$$\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\frac{1}{|\Omega|^2}\left(\frac{(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))}\right)^2.$$

We have $|v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}}| = |v_p^{\mathcal{S},2^{-(h+1)}}-\text{cost}(p,\mathcal{S})+\text{cost}(p,\mathcal{S})-v_p^{\mathcal{S},2^{-h}}| \leq 2\cdot 2^{-h}\cdot\sqrt{\text{cost}(p,\mathcal{S})^{z-1}\cdot\text{cost}(p,\mathcal{A})^{3-z}}$ due to Lemma 4. Furthermore, by definition $w_p = \frac{\text{cost}(P,\mathcal{A})|C_j|}{\text{cost}(C_j,\mathcal{A})}$. Finally, by definition of type i, we have $\text{cost}(p,\mathcal{S})\cdot|C_j| = O(1)\cdot\text{cost}(C_j,\mathcal{S})$ and by Assumption 4 we have $\text{cost}(p,\mathcal{A})\cdot|C_j| = O(1)\cdot\text{cost}(C_j,\mathcal{A})$ for all $p\in C_j$.

$$\begin{aligned}
&\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\frac{1}{|\Omega|^2}\left(\frac{(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))}\right)^2\\
\leq\ & O(1)\cdot\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\frac{1}{|\Omega|^2}\left(\frac{2^{-h}\cdot\sqrt{\text{cost}(p,\mathcal{S})^{z-1}\cdot\text{cost}(p,\mathcal{A})^{3-z}}\cdot\text{cost}(P,\mathcal{A})|C_j|}{\text{cost}(C_j,\mathcal{A})\cdot(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))}\right)^2\\
\leq\ & O(1)\cdot\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\frac{1}{|\Omega|^2}\left(\frac{2^{-2h}\cdot\text{cost}(C_j,\mathcal{A})^{1-z}\cdot\text{cost}(C_j,\mathcal{S})^{z-1}\cdot\text{cost}(P,\mathcal{A})^2}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))^2}\right)
\end{aligned}$$

Now, let k_i be the number of clusters of type i. Then due to Assumption 4 $\text{cost}(C_j,\mathcal{S})\cdot k_i \leq O(1)\text{cost}(P,\mathcal{S})$, for all C_j of type i. Finally, note that $\frac{\text{cost}(P,\mathcal{A})}{\text{cost}(C_j,\mathcal{A})}\leq O(1)\cdot k$, also due to Assumption 4. Combining this, we then have

$$\begin{aligned}
&\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\left(\frac{(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))}\right)^2\\
\leq\ & O(1)\cdot\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\left(\frac{2^{-2h}\cdot 2^{i(z-1)}k^2}{|\Omega|^2\cdot(k+k_i\cdot 2^i)^2}\right)\\
\leq\ & O(1)\cdot\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}\left(\frac{2^{-2h}\cdot k}{k_i\cdot|\Omega|^2}\right)\cdot\frac{k_i\cdot k\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}
\end{aligned}$$

Assuming event $\mathcal{E}$, this may now be bounded by $O(1)\cdot\frac{2^{-2h}}{|\Omega|}\cdot\frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}$. If event $\mathcal{E}$ does not hold, we may bound the term by $\frac{2^{-2h}\cdot k}{k_i\cdot|\Omega|}\cdot\frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}\leq\frac{2^{-2h}\cdot k}{|\Omega|}\frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}$. □

Completing the Proof for Eq. 7

Throughout this section, we use the bound on the expected maximum of independent Gaussians.

Lemma 9 (Lemma 2.3 of Massart [2007])**.** *Let* $g_i\sim\mathcal{N}(0,\sigma_i^2)$, $i\in[n]$ *be Gaussian random variables and suppose* $\sigma_i\leq\sigma$ *for all* i*. Then* $\mathbb{E}[\max_{i\in[n]}|g_i|]\leq 2\sigma\cdot\sqrt{2\ln n}$.

The number of cost vectors in $\mathbb{N}_{h+1}\times\mathbb{N}_h$ is at most $\exp\left(\gamma\cdot k\cdot\log\|P\|_0\cdot\min(k_i+2^{2h},2^{2h}\cdot 2^i)\cdot i\cdot h\right)$ for some absolute constant γ due to Lemma 4. With the bound on the variance (Lemma 8 and conditioned on event $\mathcal{E}$), we then have

$$\begin{aligned}
&\sum_{h=1}^{\log \varepsilon^{-2}} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h} \in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}} - v_p^{\mathcal{S},2^{-h}})w_p\cdot g_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} \right| \middle| \mathcal{E} \right] \\
\leq\ & \sum_{h=1}^{\log \varepsilon^{-2}} \sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i + 2^{2h}, 2^{2h}\cdot 2^i)\cdot i \cdot h \cdot \frac{2^{-2h}}{|\Omega|}\cdot \frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & 2\sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i)\cdot i \cdot \log^3 \varepsilon^{-1} \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}}. \qquad (10)
\end{aligned}$$

Conditioned on event $\mathcal{E}$ not holding, we then have using a similar calculation

$$\begin{aligned}
&\sum_{h=1}^{\log \varepsilon^{-2}} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h} \in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}} - v_p^{\mathcal{S},2^{-h}})w_p\cdot g_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} \right| \middle| \overline{\mathcal{E}} \right] \\
\leq\ & \sum_{h=1}^{\log \varepsilon^{-2}} \sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i + 2^{2h}, 2^{2h}\cdot 2^i)\cdot i \cdot h \cdot \frac{2^{-2h}\cdot k}{|\Omega|}\cdot \frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & 2\sqrt{\gamma \cdot k^2 \cdot \log \|P\|_0 \cdot \min(k_i, 2^i)\cdot i \cdot \log^3 \varepsilon^{-1} \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}}. \qquad (11)
\end{aligned}$$

We have $\mathbb{P}[\overline{\mathcal{E}}] \leq 1/k^2$ due to Lemma 7. Since $\|P\|_0 \leq \text{poly}(k, \varepsilon^{-1})$, $2^i \leq O(1)\cdot \varepsilon^{-2}$, we can combine Equations 10 and 11 with the law of total expectation to obtain

$$\begin{aligned}
&\sum_{h=1}^{\log \varepsilon^{-2}} \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h} \in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}} - v_p^{\mathcal{S},2^{-h}})w_p\cdot g_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} \right| \right] \\
\leq\ & 2\sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i)\cdot i \cdot \log^3 \varepsilon^{-1} \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i \cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
&+2\sqrt{\gamma \cdot k^2 \cdot \log \|P\|_0 \cdot \min(k_i, 2^i)\cdot i \cdot \log^3 \varepsilon^{-1} \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i \cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}\cdot\frac{1}{k^2}} \\
\leq\ & 4\sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i)\cdot i \cdot \log^3 \varepsilon^{-1} \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i \cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \qquad (12)
\end{aligned}$$

Using a straightforward, but tedious calculation, we have $\min(k_i, 2^i)\frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2} \in O(k^{z/(z+2)})$. Specifically, if $\min(k_i, 2^i) = k_i$, the term may be bounded by $\frac{k_i^2\cdot k\cdot 2^{i(z-1)}}{k^{3-z}\cdot(k_i\cdot 2^i)^{z-1}}$. If $\min(k_i, 2^i) = 2^i$, the term may be bounded by $\frac{k_i\cdot k\cdot 2^{i\cdot z}}{k^{2-z}\cdot(k_i\cdot 2^i)^z}$. Setting both terms to be equal, solving for k_i yields $k_i = k^{(z+1)/(z+2)}$. Inserting that value of k_i back into $\frac{k_i^2\cdot k\cdot 2^{i(z-1)}}{k^{3-z}\cdot(k_i\cdot 2^i)^{z-1}}$ then yields the upper bound $k^{z/(z+2)}$. Therefore, by our choice of $|\Omega|$, we can bound Eq. 12 by $O(1)\frac{\varepsilon^{-2}}{\log^3 \varepsilon^{-1}}$.

Completing the Proof for Eq. 8 Here, we use Lemma 6 and Assumption 2 to show that the number of cost vectors in $\mathbb{N}_{h+1}\times\mathbb{N}_h$ is at most $\exp\left(\gamma\cdot k\cdot\log\|P\|_0\cdot\varepsilon^{-2}\log h/\varepsilon\right)$. Conditioned on event $\mathcal{E}$, we therefore have

$$
\begin{aligned}
&\sum_{\log \varepsilon^{-2}}^{\infty} \mathbb{E}_{\Omega}\mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h}\in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} g_p \right| \middle| \mathcal{E} \right] \\
\leq\ & \sum_{\log \varepsilon^{-2}}^{\infty} O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \varepsilon^{-2}\cdot \log h/\varepsilon \cdot \frac{2^{-2h}}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & \sum_{1}^{\infty} O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \log h/\varepsilon \cdot \frac{2^{-2h}}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \log 1/\varepsilon \cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}}
\end{aligned}
$$

Similarly, if $\mathcal{E}$ does not hold, we have

$$
\begin{aligned}
&\sum_{\log \varepsilon^{-2}}^{\infty} \mathbb{E}_{\Omega}\mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h}\in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} g_p \right| \middle| \overline{\mathcal{E}} \right] \\
\leq\ & \sum_{\log \varepsilon^{-2}}^{\infty} O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \varepsilon^{-2}\cdot \log h/\varepsilon \cdot \frac{2^{-2h}k}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & \sum_{1}^{\infty} O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \log h/\varepsilon \cdot \frac{2^{-2h}k}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}} \\
\leq\ & O(1)\cdot \sqrt{\gamma\cdot k\cdot \log \|P\|_0 \cdot \log 1/\varepsilon \cdot \frac{k}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}}
\end{aligned}
$$

We have $\mathbb{P}[\overline{\mathcal{E}}] \leq 1/k^2$ due to Lemma 7. Since $\|P\|_0 \leq \text{poly}(k, \varepsilon^{-1})$, $2^i \leq O(1)\cdot \varepsilon^{-2}$ and by our choice of $|\Omega|$, we can combine the last two equations with the law of total expectation to obtain

$$
\begin{aligned}
&\sum_{\log \varepsilon^{-2}}^{\infty} \mathbb{E}_{\Omega}\mathbb{E}_g \left[\sup_{v^{\mathcal{S},h+1}-v^{\mathcal{S},h}\in \mathbb{N}_{h+1}\times\mathbb{N}_h} \left| \frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-(h+1)}}-v_p^{\mathcal{S},2^{-h}})w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} g_p \right| \right] \\
\leq\ & O(1)\cdot \sqrt{k\cdot \log k\cdot \min(k_i, 2^i)\cdot \log^5 \varepsilon^{-1}\cdot \frac{1}{|\Omega|}\cdot \frac{k\cdot k_i\cdot 2^{i(z-1)}}{(k+k_i\cdot 2^i)^2}}. \qquad (13)
\end{aligned}
$$

Observe that Eq. 13 and Eq. 12 are essentially identical up to lower order terms.

Completing the Proof for Eq. 6 Here, we first split Eq. 6 into two estimators that will be easier to handle. We split the estimator into two parts as follows. First, let $q_j := \frac{\sum_{p\in C_j} v_p^{\mathcal{S},2^{-1}}}{|C_j|}$. Now we consider

$$
\frac{1}{|\Omega|}\frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega}(v_p^{\mathcal{S},2^{-1}}-q_j)w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} g_p \qquad (14)
$$

$$
+\frac{1}{|\Omega|}\frac{\sum_{C_j\in T_i}\sum_{p\in C_j\cap\Omega} q_j\cdot w_p}{(\text{cost}(P,\mathcal{S})+\text{cost}(P,\mathcal{A}))} g_p \qquad (15)
$$

Thus, Equation 6 becomes

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} v_p^{\mathcal{S},2^{-1}} w_p}{|\Omega| \cdot (\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} g_p \right| \right] \\
= \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} |v_p^{\mathcal{S},2^{-1}} - q_j| w_p}{|\Omega| \cdot (\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} g_p \right| \right] \qquad (16) \\
+ \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot (\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} g_p \right| \right] \qquad (17)
\end{aligned}$$

Due to Assumption 4, we have $\text{cost}(T_i, \mathcal{S}) = O(1) \cdot k_i \text{cost}(C_j, \mathcal{S})$, for any $C_j \in T_i$ Thus

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot (\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} g_p \right| \right] \\
\leq \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(T_i,\mathcal{S})} g_p \right| \right] \\
\leq \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \max_{C_j \in T_i} \frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j,\mathcal{S})} g_p \right| \right] \qquad (18)
\end{aligned}$$

For the variance of the estimator used for Eq. 16, we use the following lemma.

Lemma 10. *If Assumption 4 holds, the variance of* $\frac{1}{|\Omega|} \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} (v_p^{\mathcal{S},2^{-1}} - q_j) w_p}{(\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} g_p$ *is at most*

$$\begin{aligned}
& \gamma \cdot \frac{1}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2} \qquad \textit{conditioned on event } \mathcal{E} \\
& \gamma \cdot \frac{k}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2} \qquad \textit{conditioned on event } \overline{\mathcal{E}}
\end{aligned}$$

for an absolute constant γ.

Proof. The proof of this is very close to the proof of Lemma 9 from Cohen-Addad et al. [2021a]. For k-median, this is a straightforward application of the triangle inequality. For k-means, the analysis is slightly more involved and included for completeness. Thus, throughout this proof, we have $z = 2$.

We will bound $|v_p^{\mathcal{S},2^{-1}} - q_j|$ for any point $p \in C_j$. Due to the triangle inequality and by Assumption 4 which states that all points have roughly equal distance to their center in $\mathcal{A}$, we have

$$|\sqrt{v_p^{\mathcal{S},2^{-1}}} - \sqrt{q_j}| \leq O(1) \cdot \sqrt{\text{cost}(p,\mathcal{A})}.$$

Futhermore, again due to the triangle inequality, $C_j \in T_i$ with $i > 3$ and Assumption 4, we have $(\sqrt{v_p^{\mathcal{S},2^{-1}}} + \sqrt{q_j}) = O(1)\sqrt{\text{cost}(p,\mathcal{S}}$. Therefore

$$|v_p^{\mathcal{S},2^{-1}} - q_j| \;=\; |\sqrt{v_p^{\mathcal{S},2^{-1}}} - \sqrt{q_j}| \cdot (\sqrt{v_p^{\mathcal{S},2^{-1}}} + \sqrt{q_j}) = O(1)\sqrt{\text{cost}(p,\mathcal{S})\text{cost}(p,\mathcal{A})}$$

Using this bound and the same steps as in Lemma 8, we then have

$$\begin{aligned}
& \sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} \frac{1}{|\Omega|^2} \left(\frac{(v_p^{\mathcal{S},2^{-1}} - q_j) w_p}{(\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))} \right)^2 \\
\leq \; & O(1) \cdot \sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} \frac{1}{|\Omega|^2} \frac{\text{cost}(p,\mathcal{S})\text{cost}(p,\mathcal{A})\text{cost}(P,\mathcal{S})^2 |C_j|^2}{\text{cost}(C_j,\mathcal{A})^2 \cdot (\text{cost}(P,\mathcal{S}) + \text{cost}(P,\mathcal{A}))^2} \\
\leq \; & O(1) \cdot \sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} \left(\frac{k}{k_i \cdot |\Omega|^2} \right) \cdot \frac{k \cdot k_i \cdot 2^i}{(k + k_i \cdot 2^i)^2}
\end{aligned}$$

Conditioned on event $\mathcal{E}$, this now becomes $O(1) \cdot \frac{1}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^i}{(k+k_i \cdot 2^i)^2}$ and similarly, if event $\mathcal{E}$ does not hold, we have the bound $\frac{k}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^i}{(k+k_i \cdot 2^i)^2}$. □

We now focus on the variance of the estimator used for Eq. 17. Due to Assumption 4, we have $\text{cost}(T_i, \mathcal{S}) = O(1) \cdot k_i \text{cost}(C_j, \mathcal{S})$, for any $C_j \in T_i$ Thus

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\left| \sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} g_p \right| \right] \\
\leq \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\left| \sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(T_i, \mathcal{S})} g_p \right| \right] \\
\leq \; & \mathbb{E}_\Omega \mathbb{E}_g \left[\left| \sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \max_{C_j \in T_i} \frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j, \mathcal{S})} g_p \right| \right] \qquad (19)
\end{aligned}$$

We now obtain the following variance for the estimator used in Equation 19.

Lemma 11. *If Assumption 4 holds, the variance of* $\frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot cost(C_j, \mathcal{S})} g_p$*, given that* $C_j \in T_i$ *with* $i \in \{3, \ldots, \log \varepsilon^{-2}\}$ *is at most*

$$\begin{aligned}
& \gamma \cdot \frac{k}{|\Omega|} \qquad \textit{conditioned on event } \mathcal{E} \\
& \gamma \cdot \frac{k^2}{|\Omega|} \qquad \textit{conditioned on event } \overline{\mathcal{E}}
\end{aligned}$$

for an absolute constant γ*.*

Proof. Recall by Assumption 4 $\text{cost}(P, \mathcal{S}) = O(1) \cdot k \cdot \text{cost}(C_j, \mathcal{S})$. We have

$$\begin{aligned}
& \sum_{p \in C_j \cap \Omega} \left(\frac{q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j, \mathcal{S})} \right)^2 \\
= \; & \sum_{p \in C_j \cap \Omega} \left(\frac{q_j \cdot \text{cost}(P, \mathcal{A}) \cdot |C_j|}{|\Omega| \cdot \text{cost}(C_j, \mathcal{A}) \cdot \text{cost}(C_j, \mathcal{S})} \right)^2 \\
= \; & O(1) \cdot \sum_{p \in C_j \cap \Omega} \left(\frac{k}{|\Omega|} \right)^2
\end{aligned}$$

Conditioned on event $\mathcal{E}$, $|C_j \cap \Omega| = \frac{1}{k} \cdot |\Omega|$ and this now becomes $O(1) \cdot \frac{k}{|\Omega|}$. Otherwise, we have the bound $\frac{k^2}{|\Omega|}$. □

We now bound Equations 16 and 19. For the former, we have $|\mathbb{N}_{2^{-1}}| \leq \exp\left(\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i) \cdot i\right)$. Thus, combined with Lemma 10 and conditioning on event $\mathcal{E}$, we have

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} |v_p^{\mathcal{S}, 2^{-1}} - q_j| w_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} g_p \right| |\mathcal{E} \right] \\
= \; & O(1) \sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i) \cdot i) \cdot \frac{1}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2}}
\end{aligned}$$

Similarly, not conditioning on event $\mathcal{E}$ implies

$$\begin{aligned}
& \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} |v_p^{\mathcal{S}, 2^{-1}} - q_j| w_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} g_p \right| |\overline{\mathcal{E}} \right] \\
\leq \; & O(1) \sqrt{\gamma \cdot k \cdot \log \|P\|_0 \cdot \min(k_i, 2^i) \cdot i) \cdot \frac{k}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2}}
\end{aligned}$$

We have $\mathbb{P}[\overline{\mathcal{E}}] \leq 1/k^2$ due to Lemma 7. Plugging in $\|P\|_0 \leq \text{poly}(k, \varepsilon^{-1})$ and our choice of $|\Omega|$, we can combine the last two equations with the law of total expectation to obtain

$$\begin{aligned} & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \left| \frac{\sum_{C_j \in T_i} \sum_{p \in C_j \cap \Omega} |v_p^{\mathcal{S},2^{-1}} - q_j| w_p}{|\Omega| \cdot (\text{cost}(P, \mathcal{S}) + \text{cost}(P, \mathcal{A}))} g_p \right| \right] \\ \leq \quad & O(1) \cdot \sqrt{k \cdot \log k \cdot \min(k_i, 2^i) \cdot \log^5 \varepsilon^{-1} \cdot \frac{1}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2}} \end{aligned} \tag{20}$$

For the term in Equation 19, we note that $\frac{q_j \cdot w_p}{\text{cost}(C_j, \mathcal{S})} = \text{cost}(P, \mathcal{A})$. Thus, for every cluster, we have a net of size 1, which means we have an overall net of size k. We thus obtain

$$\begin{aligned} & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \max_{C_j \in T_i} \left| \frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j, \mathcal{S})} g_p \right| |\mathcal{E} \right] \\ \leq \quad & O(1) \sqrt{\log k \frac{k}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^i}{(k + k_i \cdot 2^i)^2}} \end{aligned}$$

Similarly, conditioning on event $\overline{\mathcal{E}}$ implies

$$\begin{aligned} & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \max_{C_j \in T_i} \left| \frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j, \mathcal{S})} g_p \right| |\overline{\mathcal{E}} \right] \\ \leq \quad & O(1) \sqrt{\log k \frac{k^2}{|\Omega|} \cdot \frac{k \cdot k_i \cdot 2^{i(z-1)}}{(k + k_i \cdot 2^i)^2}} \end{aligned}$$

Combining both terms, using $\mathbb{P}[\overline{\mathcal{E}}] \leq 1/k^2$ due to Lemma 7 and the law of total expectation, we obtain

$$\begin{aligned} & \mathbb{E}_\Omega \mathbb{E}_g \left[\sup_{v^{\mathcal{S},1} \in \mathbb{N}_{2^{-1}}} \max_{C_j \in T_i} \left| \frac{\sum_{p \in C_j \cap \Omega} q_j \cdot w_p}{|\Omega| \cdot \text{cost}(C_j, \mathcal{S})} g_p \right| \right] \\ \leq \quad & O(1) \cdot \sqrt{\frac{k \log k}{|\Omega|}} \end{aligned} \tag{21}$$

Combining the bounds in Equations 12, 13, 20 and 21 for the respective terms in Equations 7, 8, 16 and 19 now yields the claim.

4 Disclosure of Funding Acknowledgements

Kapser Green Larsen was partially supported by the Independent Research Fund Denmark (DFF) under a Sapere Aude Research Leader grant No 9064-00068B.

David Saulpic has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101019564 "The Design of Modern Fully Dynamic Data Structures (MoDynStruct)".

Chris Schwiegelshohn was partially supported by the Independent Research Fund Denmark (DFF) under a Sapere Aude Research Leader grant No 1051-00106B and the Innovation Fund Denmark under grant agreement No 0153-00233A.

Omar Ali Sheikh-Omar was partially supported by the Innovation Fund Denmark under grant agreement No 0153-00233A.

References

David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007*, pages 1027–1035, 2007. URL `http://dl.acm.org/citation.cfm?id=1283383.1283494`.

Olivier Bachem, Mario Lucic, and Andreas Krause. Scalable k -means clustering via lightweight coresets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1119–1127, 2018a. doi: 10.1145/3219819.3219973. URL `https://doi.org/10.1145/3219819.3219973`.

Olivier Bachem, Mario Lucic, and Silvio Lattanzi. One-shot coresets: The case of k-clustering. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pages 784–792, 2018b. URL `http://proceedings.mlr.press/v84/bachem18a.html`.

Daniel N. Baker, Vladimir Braverman, Lingxiao Huang, Shaofeng H.-C. Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for clustering in graphs of bounded treewidth. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 569–579. PMLR, 2020. URL `http://proceedings.mlr.press/v119/baker20a.html`.

Sayan Bandyapadhyay, Fedor V. Fomin, and Kirill Simonov. On coresets for fair clustering in metric and euclidean spaces and their applications. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 23:1–23:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPIcs.ICALP.2021.23. URL `https://doi.org/10.4230/LIPIcs.ICALP.2021.23`.

Luca Becchetti, Marc Bury, Vincent Cohen-Addad, Fabrizio Grandoni, and Chris Schwiegelshohn. Oblivious dimension reduction for k-means: beyond subspaces and the johnson-lindenstrauss lemma. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 1039–1050, 2019. URL `https://doi.org/10.1145/3313276.3316318`.

Jock A. Blackard and Denis J. Dean. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture*, 24(3):131–151, 1999. ISSN 0168-1699. doi: https://doi.org/10.1016/S0168-1699(99)00046-0. URL `https://www.sciencedirect.com/science/article/pii/S0168169999000460`.

Christos Boutsidis, Michael W. Mahoney, and Petros Drineas. Unsupervised feature selection for the \$k\$-means clustering problem. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, pages 153–161, 2009. URL `http://papers.nips.cc/paper/3724-unsupervised-feature-selection-for-the-k-means-clustering-problem`.

Christos Boutsidis, Anastasios Zouzias, and Petros Drineas. Random projections for \$k\$-means clustering. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada.*, pages 298–306, 2010. URL `http://papers.nips.cc/paper/3901-random-projections-for-k-means-clustering`.

Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail. Near-optimal coresets for least-squares regression. *IEEE Trans. Inf. Theory*, 59(10):6880–6892, 2013. doi: 10.1109/TIT.2013.2272457. URL `https://doi.org/10.1109/TIT.2013.2272457`.

Christos Boutsidis, Anastasios Zouzias, Michael W. Mahoney, and Petros Drineas. Randomized dimensionality reduction for k-means clustering. *IEEE Trans. Information Theory*, 61(2):1045–1062, 2015. doi: 10.1109/TIT.2014.2375327. URL `https://doi.org/10.1109/TIT.2014.2375327`.

Vladimir Braverman, Shaofeng H.-C. Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for clustering with missing values. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 17360–17372, 2021a. URL `https://proceedings.neurips.cc/paper/2021/hash/90fd4f88f588ae64038134f1eeaa023f-Abstract.html`.

Vladimir Braverman, Shaofeng H.-C. Jiang, Robert Krauthgamer, and Xuan Wu. Coresets for clustering in excluded-minor graphs and beyond. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 2679–2696. SIAM, 2021b. doi: 10.1137/1.9781611976465.159. URL `https://doi.org/10.1137/1.9781611976465.159`.

Vladimir Braverman, Vincent Cohen-Addad, Shaofeng H.-C. Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Toftrup, and Xuan Wu. The power of uniform sampling for coresets. *CoRR*, abs/2209.01901, 2022. doi: 10.48550/arXiv.2209.01901. URL `https://doi.org/10.48550/arXiv.2209.01901`.

Moses Charikar and Erik Waingarten. The johnson-lindenstrauss lemma for clustering and subspace approximation: From coresets to dimension reduction. *CoRR*, abs/2205.00371, 2022a. doi: 10.48550/arXiv.2205.00371. URL `https://doi.org/10.48550/arXiv.2205.00371`.

Moses Charikar and Erik Waingarten. Polylogarithmic sketches for clustering. In Mikolaj Bojanczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022, July 4-8, 2022, Paris, France*, volume 229 of *LIPIcs*, pages 38:1–38:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022b. doi: 10.4230/LIPIcs.ICALP.2022.38. URL `https://doi.org/10.4230/LIPIcs.ICALP.2022.38`.

Ke Chen. On coresets for k-median and k-means clustering in metric and Euclidean spaces and their applications. *SIAM J. Comput.*, 39(3):923–947, 2009.

Yeshwanth Cherapanamjeri and Jelani Nelson. Terminal embeddings in sublinear time. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 1209–1216. IEEE, 2021. doi: 10.1109/FOCS52979.2021.00118. URL `https://doi.org/10.1109/FOCS52979.2021.00118`.

Michael B. Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 163–172, 2015.

Vincent Cohen-Addad and Jason Li. On the fixed-parameter tractability of capacitated clustering. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, pages 41:1–41:14, 2019. doi: 10.4230/LIPIcs.ICALP.2019.41. URL `https://doi.org/10.4230/LIPIcs.ICALP.2019.41`.

Vincent Cohen-Addad and Chris Schwiegelshohn. On the local structure of stable clustering instances. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 49–60, 2017. doi: 10.1109/FOCS.2017.14. URL `https://doi.org/10.1109/FOCS.2017.14`.

Vincent Cohen-Addad, David Saulpic, and Chris Schwiegelshohn. A new coreset framework for clustering. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 169–182. ACM, 2021a. doi: 10.1145/3406325.3451022. URL `https://doi.org/10.1145/3406325.3451022`.

Vincent Cohen-Addad, David Saulpic, and Chris Schwiegelshohn. Improved coresets and sublinear algorithms for power means in euclidean spaces. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 21085–21098, 2021b. URL `https://proceedings.neurips.cc/paper/2021/hash/b035d6563a2adac9f822940c145263ce-Abstract.html`.

Vincent Cohen-Addad, Kasper Green Larsen, David Saulpic, and Chris Schwiegelshohn. Towards optimal lower bounds for k-median and k-means coresets. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1038–1051. ACM, 2022. doi: 10.1145/3519935.3519946. URL https://doi.org/10.1145/3519935.3519946.

Petros Drineas, Alan M. Frieze, Ravi Kannan, Santosh Vempala, and V. Vinay. Clustering large graphs via the singular value decomposition. *Machine Learning*, 56(1-3):9–33, 2004. doi: 10.1023/B:MACH.0000033113.59016.96. URL https://doi.org/10.1023/B:MACH.0000033113.59016.96.

Michael Elkin, Arnold Filtser, and Ofer Neiman. Terminal embeddings. *Theor. Comput. Sci.*, 697: 1–36, 2017. doi: 10.1016/j.tcs.2017.06.021. URL https://doi.org/10.1016/j.tcs.2017.06.021.

Dan Feldman. Core-sets: An updated survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 10 (1), 2020. doi: 10.1002/widm.1335. URL https://doi.org/10.1002/widm.1335.

Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 569–578, 2011.

Dan Feldman, Morteza Monemizadeh, Christian Sohler, and David P. Woodruff. Coresets and sketches for high dimensional subspace approximation problems. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 630–649, 2010.

Dan Feldman, Melanie Schmidt, and Christian Sohler. Turning big data into tiny data: Constant-size coresets for k-means, pca, and projective clustering. *SIAM J. Comput.*, 49(3):601–657, 2020. doi: 10.1137/18M1209854. URL https://doi.org/10.1137/18M1209854.

Zhili Feng, Praneeth Kacham, and David P. Woodruff. Strong coresets for subspace approximation and k-median in nearly linear time. *CoRR*, abs/1912.12003, 2019. URL http://arxiv.org/abs/1912.12003.

Zhili Feng, Praneeth Kacham, and David P. Woodruff. Dimensionality reduction for the sum-of-distances metric. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 3220–3229. PMLR, 2021. URL http://proceedings.mlr.press/v139/feng21a.html.

Hendrik Fichtenberger, Marc Gillé, Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. BICO: BIRCH meets coresets for k-means clustering. In *Algorithms - ESA 2013 - 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, pages 481–492, 2013.

Gereon Frahling and Christian Sohler. A fast k-means implementation using coresets. In Nina Amenta and Otfried Cheong, editors, *Proceedings of the 22nd ACM Symposium on Computational Geometry, Sedona, Arizona, USA, June 5-7, 2006*, pages 135–143. ACM, 2006. doi: 10.1145/1137856.1137879. URL https://doi.org/10.1145/1137856.1137879.

Sariel Har-Peled and Akash Kushal. Smaller coresets for k-median and k-means clustering. *Discrete & Computational Geometry*, 37(1):3–19, 2007.

Sariel Har-Peled and Soham Mazumdar. On coresets for k-means and k-median clustering. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 291–300, 2004.

Lingxiao Huang and Nisheeth K. Vishnoi. Coresets for clustering in euclidean spaces: importance sampling is nearly optimal. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, 2020. doi: 10.1145/3357713.3384296. URL https://doi.org/10.1145/3357713.3384296.

Lingxiao Huang, Shaofeng H.-C. Jiang, Jian Li, and Xuan Wu. Epsilon-coresets for clustering (with outliers) in doubling metrics. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 814–825, 2018. doi: 10.1109/FOCS.2018.00082. URL https://doi.org/10.1109/FOCS.2018.00082.

Lingxiao Huang, Shaofeng H.-C. Jiang, and Nisheeth K. Vishnoi. Coresets for clustering with fairness constraints. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 7587–7598, 2019.

Lingxiao Huang, K. Sudhir, and Nisheeth K. Vishnoi. Coresets for regressions with panel data. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/03287fcce194dbd958c2ec5b33705912-Abstract.html.

Lingxiao Huang, K. Sudhir, and Nisheeth K. Vishnoi. Coresets for time series clustering. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 22849–22862, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/c115ba9e04ab27fbbb664f932112246d-Abstract.html.

Shaofeng H.-C. Jiang, Robert Krauthgamer, Jianing Lou, and Yubo Zhang. Coresets for kernel clustering. *CoRR*, abs/2110.02898, 2021. URL https://arxiv.org/abs/2110.02898.

Zohar S. Karnin and Edo Liberty. Discrepancy, coresets, and sketches in machine learning. In Alina Beygelzimer and Daniel Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 1975–1993. PMLR, 2019. URL http://proceedings.mlr.press/v99/karnin19a.html.

Ron Kohavi. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In Evangelos Simoudis, Jiawei Han, and Usama M. Fayyad, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA*, pages 202–207. AAAI Press, 1996. URL http://www.aaai.org/Library/KDD/1996/kdd96-033.php.

Michael Langberg and Leonard J. Schulman. Universal ε-approximators for integrals. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 598–607, 2010.

Sepideh Mahabadi, Konstantin Makarychev, Yury Makarychev, and Ilya P. Razenshteyn. Nonlinear dimension reduction via outer bi-lipschitz extensions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1088–1101, 2018.

Tung Mai, Cameron Musco, and Anup Rao. Coresets for classification - simplified and strengthened. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 11643–11654, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/6098ed616e715171f0dabad60a8e5197-Abstract.html.

Konstantin Makarychev, Yury Makarychev, and Ilya P. Razenshteyn. Performance of johnson-lindenstrauss transform for k-means and k-medians clustering. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 1027–1038, 2019.

Pascal Massart. Concentration inequalities and model selection. 2007.

Alejandro Molina, Alexander Munteanu, and Kristian Kersting. Core dependency networks. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI*

Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 3820–3827. AAAI Press, 2018. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16847.

Alexander Munteanu and Chris Schwiegelshohn. Coresets-methods and history: A theoreticians design pattern for approximation and streaming algorithms. *Künstliche Intell.*, 32(1):37–53, 2018.

Alexander Munteanu, Chris Schwiegelshohn, Christian Sohler, and David P. Woodruff. On coresets for logistic regression. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 6562–6571, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/63bfd6e8f26d1d3537f4c5038264ef36-Abstract.html.

Shyam Narayanan and Jelani Nelson. Optimal terminal dimensionality reduction in euclidean space. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1064–1069. ACM, 2019. doi: 10.1145/3313276.3316307. URL https://doi.org/10.1145/3313276.3316307.

Jeff M. Phillips and Wai Ming Tai. Near-optimal coresets of kernel density estimates. *Discret. Comput. Geom.*, 63(4):867–887, 2020. doi: 10.1007/s00454-019-00134-6. URL https://doi.org/10.1007/s00454-019-00134-6.

Gilles Pisier. *The volume of convex bodies and Banach space geometry.* Cambridge Tracts in Mathematics. 94, 1999.

Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 764–773. ACM, 2014. doi: 10.1145/2591796.2591797. URL https://doi.org/10.1145/2591796.2591797.

Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. Fair coresets and streaming algorithms for fair k-means. In *Approximation and Online Algorithms - 17th International Workshop, WAOA 2019, Munich, Germany, September 12-13, 2019, Revised Selected Papers*, pages 232–251, 2019. doi: 10.1007/978-3-030-39479-0_16. URL https://doi.org/10.1007/978-3-030-39479-0_16.

Chris Schwiegelshohn and Omar Ali Sheikh-Omar. An empirical evaluation of k-means coresets. In Shiri Chechik, Gonzalo Navarro, Eva Rotenberg, and Grzegorz Herman, editors, *30th Annual European Symposium on Algorithms, ESA 2022, September 5-9, 2022, Berlin/Potsdam, Germany*, volume 244 of *LIPIcs*, pages 84:1–84:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi: 10.4230/LIPIcs.ESA.2022.84. URL https://doi.org/10.4230/LIPIcs.ESA.2022.84.

Christian Sohler and David P. Woodruff. Strong coresets for k-median and subspace approximation: Goodbye dimension. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 802–813, 2018. doi: 10.1109/FOCS.2018.00081. URL https://doi.org/10.1109/FOCS.2018.00081.

Murad Tukan, Alaa Maalouf, and Dan Feldman. Coresets for near-convex functions. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/0afe095e81a6ac76ff3f69975cb3e7ae-Abstract.html.