Moderate Dimension Reduction for k-Center Clustering

Shaofeng H.-C. Jiang
Peking University, Beijing, China

Robert Krauthgamer
Weizmann Institute of Science, Rehovot, Israel

Shay Sapir
Weizmann Institute of Science, Rehovot, Israel

Abstract

The Johnson-Lindenstrauss (JL) Lemma introduced the concept of dimension reduction via a random linear map, which has become a fundamental technique in many computational settings. For a set of n points in $\mathbb{R}^d$ and any fixed $\epsilon > 0$, it reduces the dimension d to $O(\log n)$ while preserving, with high probability, all the pairwise Euclidean distances within factor $1 + \epsilon$. Perhaps surprisingly, the target dimension can be lower if one only wishes to preserve the optimal value of a certain problem on the pointset, e.g., Euclidean max-cut or k-means. However, for some notorious problems, like diameter (aka furthest pair), dimension reduction via the JL map to below $O(\log n)$ does not preserve the optimal value within factor $1 + \epsilon$.

We propose to focus on another regime, of *moderate dimension reduction*, where a problem's value is preserved within factor $\alpha > 1$ using target dimension $\log n/\mathrm{poly}(\alpha)$. We establish the viability of this approach and show that the famous k-center problem is α-approximated when reducing to dimension $O(\frac{\log n}{\alpha^2} + \log k)$. Along the way, we address the diameter problem via the special case $k = 1$. Our result extends to several important variants of k-center (with outliers, capacities, or fairness constraints), and the bound improves further with the input's doubling dimension.

While our $\mathrm{poly}(\alpha)$-factor improvement in the dimension may seem small, it actually has significant implications for streaming algorithms, and easily yields an algorithm for k-center in dynamic geometric streams, that achieves $O(\alpha)$-approximation using space $\mathrm{poly}(kdn^{1/\alpha^2})$. This is the first algorithm to beat $O(n)$ space in high dimension d, as all previous algorithms require space at least $\exp(d)$. Furthermore, it extends to the k-center variants mentioned above.

2012 ACM Subject Classification Theory of computation → Random projections and metric embeddings; Theory of computation → Sketching and sampling; Theory of computation → Streaming models

Keywords and phrases Johnson-Lindenstrauss transform, dimension reduction, clustering, streaming algorithms

Digital Object Identifier 10.4230/LIPIcs.SoCG.2024.64

Related Version *Full Version*: `https://arxiv.org/abs/2312.01391`

Funding *Shaofeng H.-C. Jiang*: Research partially supported by a national key R&D program of China No. 2021YFA1000900 and a startup fund from Peking University.

Robert Krauthgamer: Work partially supported by the Israel Science Foundation grant #1336/23, and by the Weizmann Data Science Research Center.

Shay Sapir: This research was partially supported by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center.

1 Introduction

The seminal work of Johnson and Lindenstrauss [39] introduced the technique of dimension reduction via an (oblivious) random linear map, and this technique has become fundamental in many computational settings, from offline to streaming and distributed algorithms, especially nowadays that high-dimensional data is ubiquitous. Their so-called JL Lemma asserts (roughly) that for any fixed $\epsilon > 0$, a random mapping (e.g., projection) of a set $P \subset \mathbb{R}^d$ of n points to target dimension $t = O(\log n)$ preserves the (Euclidean) distances between all points in P within $1 + \epsilon$ factor. Furthermore, the JL Lemma is known to be tight [45], see also the recent survey [50].

Perhaps surprisingly, the target dimension can sometimes be reduced below that $O(\log n)$ bound, particularly when one only wants to preserve the optimal value of a specific objective function rather than all pairwise distances. Indeed, for several optimization problems on the pointset P, previous work has shown that the target dimension may be much smaller than $O(\log n)$ or even independent of n, e.g., for Euclidean max-cut [44, 43, 19], k-means [11, 22, 10, 46] and subspace approximation [16]. However, for some notorious problems, like facility location, minimum spanning tree [49] and diameter (aka furthest pair), dimension reduction via the JL map to below $O(\log n)$ does not preserve the optimal value within factor $1 + \epsilon$.

In light of this, we consider a new regime of *moderate dimension reduction*, where a problem's value is approximated within factor $O(1)$ using target dimension slightly below $O(\log n)$.[1] More precisely, we aim at a tradeoff of achieving α-estimation,[2] for any desired $\alpha > 1$, when reducing to dimension $\log n/\operatorname{poly}(\alpha)$. This relaxation of the approximation factor, from $1 + \epsilon$ to α, may be effective in combating the "curse of dimensionallity" phenomenon, because when an algorithm's efficiency is exponential in the dimension (e.g., space complexity in streaming algorithms), bounds of the form $2^{O(\log n)} = n^{O(1)}$ improve to $2^{\log n/\operatorname{poly}(\alpha)} = n^{1/\operatorname{poly}(\alpha)}$. This tradeoff can even yield target dimension $t = O(1)$, by using approximation $\alpha = \operatorname{polylog}(n)$, and this can lead to new results (e.g., streaming algorithms).

We study this regime of moderate dimension reduction for the fundamental problem of k-center clustering: The input is a set $P \subset \mathbb{R}^d$ of n points, and the goal is to find a set of centers $C \subset \mathbb{R}^d$ of size k that minimizes the objective $\max_{p\in P} \operatorname{dist}(p, C)$, where $\operatorname{dist}(p, C) = \min_{c\in C} \|p - c\|$ (throughout, we use ℓ_2 norm). The special case $k = 1$ is exactly the minimum enclosing ball problem, which is within factor 2 of the diameter problem, and for both problems, we show that reducing the dimension to $o(\log n)$ is unlikely to achieve $(1 + \epsilon)$-estimation. Our main contribution is a general framework for moderate dimension reduction that works even for more challenging and widely studied variants of k-center, such as the outliers [15], capacitated [9, 40], and fair [20] variants.

1.1 Main Results

Our main result is a moderate dimension reduction for k-center via an (oblivious) random linear map. For simplicity, consider a map defined via a matrix $G \in \mathbb{R}^{d\times t}$ of iid Gaussians (scaled appropriately), although our result is more general and holds for several (but not all) known JL maps, similarly to prior work in this context [46, 19]. Specifically, we need G to satisfy the JL Lemma and have a *sub-Gaussian tail*, as described in Section 1.4.

---

[1] A related approach called low-quality dimension reduction was proposed in [5]. Similarly to ours, it refers to a map with target dimension below $O(\log n)$, but its utility is to $(1+\epsilon)$-approximate Nearest-Neighbor Search (NNS), and since NNS is not an optimization problem, their definitions and techniques are rather different from ours.

[2] We say that dimension reduction achieves α-estimation if it preserves the problem's optimal value within factor α; and we say it achieves α-approximation if it also preserves solutions, in a sense that we define formally later. Our results extend to $O(\alpha)$-approximation, but we focus here on $O(\alpha)$-estimation for clarity.

Table 1 Dimension reduction bounds for $O(\alpha)$-estimation of k-center variants in terms of $D = \min(d, \log n)$. The lower bound (marked by *) holds only for a Gaussian Matrix.

k-center variant	target dimension	reference
all variants	$O(D)$ $\Omega(\frac{D}{\alpha^2} + \frac{\log k}{\log \alpha})$ *	JL Lemma/trivial the full version
vanilla	$O(\frac{D}{\alpha^2} + \log k)$	Theorem 2.1
with z outliers	$O(\frac{D}{\alpha^2} + \log(kz))$	Theorem 3.1
assignment constrained	$O(\frac{D}{\alpha} + \log k)$	Theorem 4.3

▶ **Theorem 1.1** (Main Result, informal). *For every α, d, k and n, there is a random linear map $G : \mathbb{R}^d \to \mathbb{R}^t$ with target dimension $t = O(\frac{\log n}{\alpha^2} + \log k)$, such that for every set $P \subset \mathbb{R}^d$ of n points, with high probability, G preserves the k-center value of P within $O(\alpha)$ factor. This result extends to k-center variants as listed in Table 1.*

▶ Remark 1.2. We actually prove this theorem for target dimension $t = O(\frac{d}{\alpha^2} + \log k)$, which is only stronger; indeed, one can assume that $d = O(\log n)$ by the JL Lemma.

Our bound is nearly optimal when G is a matrix of iid Gaussians (and plausibly for all JL maps). Concretely, even for $k = 1$, the target dimension must be $t = \Omega(\frac{\log n}{\alpha^2})$, which matches the leading term in our bound; and the second term is nearly matched by an $\Omega(\frac{\log k}{\log \alpha})$ bound. For more details, see the full version.

To demonstrate that moderate dimension reduction is a general approach that may be applied more broadly, our Theorem 1.1 includes non-trivial extensions to several important variants of k-center. From here on, we call the classic k-center mentioned above the "vanilla" variant, essentially to distinguish it from other variants that we now discuss. In the variant *with outliers* (aka robust k-center), the input specifies also $z \geq 0$, and the goal is to find, in addition to the set of centers C, a set of at most z outliers $Z \subset P$, that minimizes the objective $\max_{p \in P \setminus Z} \operatorname{dist}(p, C)$ [15]. Another variant, called k-center *with an assignment constraint*, asks to assign each input point to one of the k clusters (not necessarily to the closest center), given a constraint on the entire assignment, and the goal is to minimize the maximum cluster radius. This formulation via a constrained assignment has been studied before for other clustering problems [51, 34, 8, 12] but not for k-center. It is useful as a generalization that captures both the capacitated variant, where every cluster has a bounded capacity [9, 40], and the fair variant, where input points have colors and the relative frequency of colors in every cluster must be similar to that of the entire input [20]. Perhaps surprisingly, the dimension reduction satisfies a strong "for all" guarantee with respect to the assignment constraints, i.e., with high probability the optimal value is preserved simultaneously for all possible constraints (whose number can be extremely large). This is in contrast to the weaker "for each" guarantee, where the high-probability bound applies separately to each constraint. The extensions to k-center variants come at the cost of a slightly increased target dimension, as listed in Table 1.

Another important feature of our approach is that it actually preserves solutions, i.e., Theorem 1.1 extends from $O(\alpha)$-estimation to $O(\alpha)$-approximation, as follows. For vanilla k-center, an *α-approximate solution* of P is a set $C \subset P$ of size k whose objective value $\max_{p \in P} \operatorname{dist}(p, C)$ is within factor α from the optimal. The proof of Theorem 1.1 can be extended to show that a set $C \subset P$ is an $O(\alpha)$-approximate solution of P whenever $G(C)$ is an

$O(1)$-approximate solution of $G(P)$.[3] The above restriction to $C \subset P$ is needed to guarantee that $C \mapsto G(C)$ is one-to-one, but one can easily relax it to our true requirement $C \subset \mathbb{R}^d$ by introducing a factor of 2 in the approximation ratio. This extension to $O(\alpha)$-approximation holds also for the aforementioned variants of k-center.

Still, an important question remains – how useful is this theorem, and more generally, this regime of moderate dimension reduction? It may seem that for fixed $\alpha > 1$, the target dimension $\frac{\log n}{\text{poly}(\alpha)} = \Theta(\log n)$ offers only negligible improvement over the JL Lemma. The crux is that many algorithms depend exponentially in the dimension, in which case decreasing the dimension *by a constant factor* amounts to a *polynomial improvement* in efficiency. We indeed show such an application to streaming algorithms, as discussed next.

1.2 Application: Dynamic Geometric Streams

In dynamic geometric streams, a model introduced by Indyk [36], the input P is presented as a stream of insertions and deletions of points from $[\Delta]^d = \{1, 2, ..., \Delta\}^d$. Algorithms in this model read the stream in one pass and their space complexity (aka storage requirement) is limited. We assume that $\Delta \leq \text{poly}(n)$, which is common in this model. Ideally, algorithms for k-center should use at most $\text{poly}(kd \log n)$ bits of space, which is polynomial in the bit representation of a solution consisting of k points in $[\Delta]^d$. We focus throughout on the general case of high dimension (say $d \geq \log n$), and mostly ignore algorithms whose space complexity is $\geq 2^d$ bits (which are suitable only for low dimension).

For k-center in insertion-only streams (i.e., without deletions), the tradeoff between approximation and space complexity is well understood, and the regime of $O(1)$-approximation seems to be the most useful and interesting. Indeed, there is an $O(1)$-approximation algorithm using $O(kd \log n)$ bits of space [15], with extensions to the outliers variant [48] and also to the sliding-window model [23]. In contrast, all known $(1 + \epsilon)$-approximation algorithms have space bound that grows like $(1/\epsilon)^d$ [2, 14, 29, 28], which is not sublinear in n for high dimension. Furthermore, approximation below $\frac{1+\sqrt{2}}{2}$ provably requires $\Omega(\min\{n, \exp(d^{1/3})\})$ bits of space, even for $k = 1$ [3]. Another indication is a $(1.8 + \epsilon)$-approximation for vanilla k-center using space complexity bigger by factor $k^{O(k)}$ [41].

The setting of dynamic streams (i.e., with deletions) seems much harder. Here, $O(1)$-estimation using $\text{poly}(kd \log n)$ bits of space is widely open, as we seem to lack effective algorithmic techniques. For example, in insertion-only streams, 2-approximation can be easily achieved for $k = 1$, by just storing the first point (in the stream) and then the furthest point from it. It is unknown whether this algorithm extends to dynamic streams, because it relies on access to a point from P right as the stream begins. We provide an algorithm for dynamic streams and all k, by simply applying our moderate dimension reduction result and then using an algorithm for low dimension by [28].

▶ **Theorem 1.3** (Streaming Algorithm for k-Center, informal). *There is a randomized algorithm that, given α, d, k, n and a set $P \subset \mathbb{R}^d$ of size at most n presented as a stream of* $\text{poly}(n)$ *insertions and deletions of points, returns an $O(\alpha)$-approximation to the k-center problem on P. The algorithm uses n^{1/α^2}* $\text{poly}(kd \log n)$ *bits of space, and extends to k-center variants as listed in Table 2.*

▶ Remark 1.4. This theorem can also achieve $2^{d/\alpha^2} \text{poly}(kd \log n)$ bits of space, which is only stronger; indeed, one can assume that $d = O(\log n)$ by the JL Lemma. By setting α appropriately, this algorithm achieves $O(\sqrt{d/\log(kd \log n)})$-estimation using our ideal space bound of $\text{poly}(kd \log n)$ bits.

---

[3] For a set $X \subset \mathbb{R}^d$ and matrix $G \in \mathbb{R}^{t \times d}$, let $G(X) := \{Gx : x \in X\}$.

Table 2 Space complexity upper bounds for $O(\alpha)$-estimation in dynamic streams of k-center variants, listed separately as a function of d and of n, but omitting $\mathrm{poly}(kd\log n)$ factors. The results of [28] actually achieve $(1+\epsilon)$-estimation.

k-center variant	dynamic streaming space upper bounds		reference
vanilla	2^d	–	[28]
	–	n^{1/α^2}, only for $k=1$	[35]
	$2^{d/\alpha}$	$n^{1/\alpha}$	derived from [24]
	$2^{d/\alpha^2}$	n^{1/α^2}	Corollary 2.3
with z outliers	2^d+z	–	[28]
	$2^{d/\alpha^2}\,\mathrm{poly}(z)$	$n^{1/\alpha^2}\,\mathrm{poly}(z)$	Corollary 3.4
capacitated	2^d	–	[28]
	$2^{d/\alpha}$	$n^{1/\alpha}$	the full version
fair	2^d	–	[28]
	$2^{d/\alpha}$	$n^{1/\alpha}$	the full version

We do not know whether our bounds are tight, and in fact, proving lower bounds for dynamic geometric streams is a challenging open problem (apparently, even for deterministic algorithms). One would expect these lower bounds to exceed those known for insertion-only streams, however, current techniques seem unable to exploit deletions.

Our result improves and significantly generalizes the known bounds for k-center in dynamic streams. Currently, the best algorithm that can handle deletions works only for $k=1$ and the vanilla variant, and achieves $O(\alpha)$-estimation using $\tilde{O}(n^{1/\alpha^2}d)$ bits of space[4] (i.e., same bound as in Theorem 1.3, but only for a special case). This result is unpublished but known to experts, and follows by adapting a dynamic algorithm from [35] to the streaming setting. Known algorithms for insertion-only streams are largely not relevant, as they rarely extend to handle deletions. Perhaps the only exception is a simple approach based on ϵ-nets (see e.g. [2]), that has been extended to handle deletions by employing sparse-recovery techniques [28]. It was further extended to the outliers variant [28], and may possibly extend to the capacitated and fair variants as well. This algorithm achieves $(1+\epsilon)$-approximation and uses space complexity that is bigger by factor $(O(\frac{1}{\epsilon}))^d$, which, as mentioned earlier, is not sublinear in n for high dimension. Another possible approach is to employ a recent technique called consistent hashing [24], to obtain (quite easily, details omitted) a streaming algorithm for vanilla k-center that uses $2^{d/\alpha}\,\mathrm{poly}(kd\log n)$ bits of space. The dependence on α here is inferior to Theorem 1.3, and is known to be optimal for consistent hashing [24]. We remark that the tree-embedding technique of [36], which has been useful for several dynamic streaming problems in high dimension, is ineffective for k-center, or even the diameter problem, as it bounds only the *expected stretch* of every pair of points.

Recent research on streaming algorithms has uncovered an intriguing tradeoff between approximation and space complexity for different geometric problems. Interestingly, the tradeoff we obtain for vanilla k-center is better than the one known for these other problems. For earth mover's distance (EMD) in the plane (i.e., $\mathbb{R}^2$), we know of $O(\alpha)$-estimation

[4] Throughout, the notation $\tilde{O}(f)$ hides $\mathrm{poly}(\log n)$ factors.

using $\tilde{O}(n^{1/\alpha})$ bits of space [6].[5] For minimum spanning tree (MST), we know of $O(\alpha)$-estimation using $n^{\sqrt{\log \alpha/\alpha}}\,\mathrm{poly}(d \log n)$ bits of space [17]. For facility location, we know of $O(\alpha)$-estimation using $n^{1/\alpha}\,\mathrm{poly}(d \log n)$ bits of space [24]. We currently have no satisfactory explanation for these gaps, but since these four results rely on rather different methods, developing unified techniques (possibly via dimension reduction) may potentially improve some of these bounds.

1.3 Extension: Inputs of Small Doubling Dimension

When the input $P \subset \mathbb{R}^d$ has small doubling dimension, one can achieve even better dimension reduction than Theorem 1.1, eliminating the dependence (of t) on d and n. Following [32] (see also [21]), the *doubling dimension* of a set $P \subset \mathbb{R}^d$, denoted $\mathrm{ddim}(P)$, is the smallest number such that every ball (in P) can be covered by $2^{\mathrm{ddim}(P)}$ balls of half the radius. Dimension reduction for inputs of small doubling dimension has been studied before for three problems: For facility location, one can achieve $O(1)$-estimation using target dimension $t = O(\mathrm{ddim}(P))$ [49]. For Nearest-Neighbour Search (NNS), one can obtain $(1+\epsilon)$-approximation using $t = O(\frac{\log(1/\epsilon)}{\epsilon^2}\,\mathrm{ddim}(P))$ [38], and for minimum spanning tree (MST), one obtains $(1+\epsilon)$-estimation using a similar target dimension albeit with another additive term of $O(\frac{\log \log n}{\epsilon^2})$ [49]. The following theorem, which we prove in the full version, shows an analogous result for $(1+\epsilon)$-estimation of k-center.

▶ **Theorem 1.5** (Dimension Reduction for Doubling Sets, informal)**.** *For every ϵ, d, k and ddim, suppose G is as in Section 1.1 with target dimension $t = O(\frac{\log(1/\epsilon)}{\epsilon^2}\,\mathrm{ddim} + \frac{\log k}{\epsilon^2})$. Then, for every set $P \subset \mathbb{R}^d$ whose doubling dimension is at most ddim, with high probability, G preserves the k-center value of P within $1+\epsilon$ factor.*[6] *This result extends to the k-center variants listed in Table 1.*

Observe that by composing the maps in Theorems 1.1 and 1.5, we can achieve $O(\alpha)$-approximation of k-center when reducing to target dimension $t = O(\frac{\mathrm{ddim}(P)}{\alpha^2} + \log k)$. This bound is better than the one in Theorem 1.1 since always $\mathrm{ddim}(P) = O(\min(d, \log n))$. Consequently, the space of the streaming algorithm in Theorem 1.3 improves to $2^{\mathrm{ddim}(P)/\alpha^2}\,\mathrm{poly}(kd \log n)$ bits. As the result extends to the k-center variants (with outliers and with assignment constraint) in a natural way, we can replace d with $\mathrm{ddim}(P)$ also in Tables 1 and 2.

1.4 Technical Overview

We discuss below dimension reduction that maps from dimension d to $t = O(\frac{d}{\alpha^2} + \log k)$, and Theorem 1.1 follows by using the JL Lemma to effectively assume $d = O(\log n)$, or alternatively by an easy adaptation of the proof (essentially via Remark 1.7 below).

Warm Up: the Furthest Point Query Problem. We start with moderate dimension reduction for the furthest point query (FPQ) problem, which may be of independent interest. In this problem, the input is a *data set* $P \subset \mathbb{R}^d$ of size $|P| = n$ and a *query set* $Q \subset \mathbb{R}^d$ of size $|Q| \le k$, and the goal is to report a point from P that is furthest from the set Q.

---

[5] This result is in fact in terms of Δ, but our assumption $\Delta \le \mathrm{poly}(n)$ implies $\Delta^{O(1/\alpha)} = n^{O(1/\alpha)}$.

[6] The k-center value here refers to $C \subset \mathbb{R}^d$, i.e., centers from the ambient space. The theorem extends to preserving solutions, albeit with the restriction $C \subseteq P$, which introduces a factor of 2 in the approximation.

Let $FPQ_k(P,Q)$ denote this optimal value (distance from Q). One can use this problem to achieve 2-approximation for vanilla k-center, by simply employing the famous Gonzalez's algorithm (aka furthest-first traversal) [31], which essentially solves k instances of FPQ. This FPQ problem admits the same dimension reduction as in Theorem 1.1, with a slightly simpler proof than for vanilla k-center, and thus serves as a good warm up. (The theorem and proof are provided rigorously in the full version.)

We consider dimension reduction by a matrix of iid Gaussians, but actually, all our results hold for any randomized map that satisfies (1) the JL Lemma about distortion of distances, and (2) the following sub-Gaussian tail. We say that a random map $f : \mathbb{R}^d \to \mathbb{R}^t$ has *sub-Gaussian tail* if

$$\forall x \in \mathbb{R}^d, r > 0, \qquad \Pr_f \left[\|f(x)\| \geq (1+r)\|x\| \right] \leq e^{-\Omega(r^2 t)}. \tag{1}$$

This tail bound was key to prior work in this context, and it holds for a $t \times d$ matrix of iid Gaussians $N(0, \frac{1}{t})$ [46, 19]. The next technical lemma is key to our proof, and shows that for a $t \times d$ matrix of iid Gaussians $N(0,1)$ (not normalized by $\frac{1}{\sqrt{t}}$), w.h.p. the largest singular value is $O(\sqrt{d})$. It can be derived from [52, Theorem 4.6.1] for a matrix of independent sub-Gaussians, because for a $d \times d$ matrix, the largest singular value is $O(\sqrt{d})$, and removing rows cannot increase the largest singular value. In the full version, we provide a proof using only the JL Lemma and the sub-Gaussian tail property. Our proof probably holds also for random orthogonal projection (scaled appropriately), which was used in [39, 30].

▶ **Lemma 1.6.** *Let $c_0 > 0$ be a suitable universal constant and let $t < d$. Suppose $G \in \mathbb{R}^{t \times d}$ is a matrix of iid Gaussians $N(0,1)$, then*

$$\Pr\Big[\sup_{\|x\| \leq 1} \|Gx\| > c_0\sqrt{d} \Big] \leq 2^{-\Omega(d)}.$$

▶ Remark 1.7. When restricting x to be from a set $P \subset \mathbb{R}^d$, we can replace d with $\mathrm{ddim}(P) = O(\log n)$ by slightly adapting the proof, thus w.h.p. $\|Gx\| \leq c_0 \sqrt{\mathrm{ddim}(P)} \cdot \|x\|$.

Our dimension-reduction map is defined via $\frac{1}{\sqrt{t}} G$ (for G as in Lemma 1.6), which is known to be a JL map [37, 27]. If $t \geq \frac{d}{\alpha^2}$, then by Lemma 1.6, with high probability, our map expands all vectors at most by factor $O(\sqrt{d/t}) = O(\alpha)$, thus the value of $FPQ_k(P,Q)$ increases at most by this factor. At the same time, consider a point $p^* \in P$ that is furthest from Q. If $t \geq c_1 \log k$ for a suitable constant $c_1 > 0$, then we can apply the JL Lemma (say, with $\epsilon = \frac{1}{2}$) on $Q \cup \{p^*\}$ and get that with high probability, $\mathrm{dist}(\frac{1}{\sqrt{t}} Gp^*, \frac{1}{\sqrt{t}} G(Q)) \geq \frac{1}{2} \mathrm{dist}(p^*, Q)$. This concludes the proof for FPQ.

Framework for Problems with Small Witness. An immediate corollary of Lemma 1.6 is that for every k-center variant, if one uses target dimension $t \geq \frac{d}{\alpha^2}$, then the optimal value increases at most by factor $O(\alpha)$. (Obviously, this fact may be useful for many other geometric problems.) We denote by $\mathrm{opt}(X)$ the optimal value of the problem at hand (e.g., vanilla or outliers) for a set X.

It remains to prove that the optimal value does not decrease much, and we devise for it the following approach: prove the existence of a small subset $S \subset P$ (say, of size $O(k)$), that we shall call a *witness*, for which $\mathrm{opt}(S) = \Omega(\mathrm{opt}(P))$, and then apply the JL Lemma on this set (say, with $\epsilon = \frac{1}{2}$). If G decreases all pairwise distances in S by at most factor 2, then we immediately get (by restricting the centers to the dataset, which loses another factor 2),

$$\mathrm{opt}(G(P)) \geq \mathrm{opt}(G(S)) \geq \tfrac{1}{4} \mathrm{opt}(S) = \Omega(\mathrm{opt}(P)),$$

which concludes the proof. We apply this witness-based approach below, viewing it as a framework that may find additional uses in the future. Our notion of "witness" is somewhat analogous to a coreset: both notions preserve the cost in a certain way, and both have a small size. Charikar and Waingarten [16] used an analogous (but technically different) argument, relying on coresets, to prove dimension reduction results for other clustering problems.

Witness for Vanilla k-Center. Consider running Gonzalez's algorithm [31], which is the following iterative algorithm. Maintain a set $S \subseteq P$, initialized to contain one arbitrary point from P, and then while $|S| \leq k$, find a solution for $FPQ_k(P, S)$ (i.e., a point furthest from the current S) and add it to S. It is well known that the distance between the last point added to S and the earlier points is in the range $[\text{opt}_{vanilla}(P), 2\,\text{opt}_{vanilla}(P)]$, and moreover, $\text{opt}_{vanilla}(S) \geq \frac{1}{2}\,\text{opt}_{vanilla}(P)$. This set S of size $k+1$ serves as a witness in Section 2.

Witness for k-Center with z Outliers. For this variant, we provide in Section 3 a witness of size $O(kz)$. The construction is based on executing Gonzalez's algorithm $z+1$ times, where after each execution we delete from P the $k+1$ points found in that execution.

Variant with an Assignment Constraint. We do not present a witness for this variant, but rather bound the decrease in value via a different method. Denote by $\text{opt}_{\mathcal{C}}(\cdot)$ the value of k-center with an assignment constraint $\mathcal{C}$. Our proof in Section 4 compares $\text{opt}_{\mathcal{C}}$ to $\text{opt}_{vanilla}$ (on the same input P) – if these values are similar, then the proof is concluded by the fact that $\text{opt}_{vanilla}$ is preserved. Otherwise, $\text{opt}_{\mathcal{C}}$ is significantly larger than $\text{opt}_{vanilla}$, and for the sake of analysis, we "move" every data point to its nearest "vanilla center", and get a weighted set of only k points, whose total weight is n. The crux of the proof shows that under the random linear map, moving points of P corresponds to moving points of $G(P)$, which in turn does not change $\text{opt}_{\mathcal{C}}(G(P))$ by too much, essentially because by Lemma 1.6, the map keeps all points close to their vanilla center.

Streaming Algorithm. Our streaming algorithm is a corollary of the dimension reduction. First apply the dimension reduction $G : \mathbb{R}^d \to \mathbb{R}^t$ of Theorem 1.1 on every point in the input stream, and then employ a known algorithm whose space is exponential in the reduced dimension t. The success probability can be amplified by standard methods.

The known algorithm returns only an estimation, and not an approximate solution, but this can be resolved, and even extended to return a solution in $\mathbb{R}^d$ using existing techniques, as follows. We refer here to the algorithm of [28] that computes a $(1+\epsilon)$-estimation of vanilla k-center (in $\mathbb{R}^t$) using $\text{poly}(k(1/\epsilon)^t \log n)$ bits of space. It uses an $(\epsilon\,\text{opt}_{vanilla})$-net and sparse recovery, which allow it to find the non-empty net-points (viewed as buckets), but these net-points need not correspond to points in $\mathbb{R}^d$. We adapt this algorithm to report a solution (centers from $\mathbb{R}^d$, actually from P) by using a two-level ℓ_0-sampler [24, Lemma 3.3], which can be viewed as a more sophisticated version of sparse recovery – add to each insertion/deletion of a net-point (in $\mathbb{R}^t$) a "data field" containing the original input point (in $\mathbb{R}^d$); now a two-level sampler will pick a random net-point (bucket) and then a random element from that bucket, with the element's data field revealing an input point. This method extends to the variants listed in Table 2, where throughout (i.e., for all variants) a solution is defined to be just the set of k centers.

1.5 Related Work

JL Maps. There are other constructions of random linear maps that satisfy the JL Lemma besides projection to a random subspace [39, 30] and a matrix of Gaussians [37, 27], like a matrix of iid Rademacher random variables [1] or independent random variables with a sub-Gaussian tail (which includes both Gaussians and Rademacher random variables) [47, 38, 42]. Moreover, there is a long line of work on maps with improved running time, e.g., the two cornerstone results known as fast JL [4] and sparse JL [26], although these are not known to satisfy (1).

Streaming Algorithms in High Dimension. For dynamic streams in high dimension, there is no prior work on k-center (except for $k = 1$, see Section 1.2), although there is work on other problems. Ideal $(1+\epsilon)$-approximations are achieved for a few problems, including clustering problems like k-means [33] and k-median [13], where the space bound is $\text{poly}(\epsilon^{-1} kd \log \Delta)$, and more recently for Max-Cut [19]. For many other problems, existing algorithms provide worse approximations, like $O(d \log \Delta)$ or $O(d)$ [36, 7, 6, 18, 53, 24, 17]. In fact, some of the work mentioned above (and also ours) provides a tradeoff between approximation and space complexity (e.g., the space can vary between polynomial and exponential in d), and it is open whether this tradeoff is necessary for these problems.

2 Dimension Reduction for Vanilla k-Center

In this section, we prove that a linear map via a matrix of iid Gaussians preserves the vanilla k-center value and solution, hence prove the main claim in Theorem 1.1.

▶ **Theorem 2.1.** *Let $d, \alpha > 1$ and $k \le n$. There is a random linear map $G : \mathbb{R}^d \to \mathbb{R}^t$ with $t = O(\log k + \frac{d}{\alpha^2})$, such that for every set $P \subset \mathbb{R}^d$ of size n, with probability at least $2/3$,*

- *$\text{opt}_{vanilla}(G(P))$ is an $O(\alpha)$-estimation for $\text{opt}_{vanilla}(P)$, and*
- *$C \subset P$ is an $O(\alpha)$-approximate vanilla k-center solution of P whenever $G(C)$ is an $O(1)$-approximate vanilla k-center solution of $G(P)$.*

In order to prove Theorem 2.1, we will use the following version of the JL Lemma.

▶ **Fact 2.2** (JL Lemma [37, 27])**.** *Let G be a $t \times d$ matrix of iid Gaussians $N(0,1)$. Then*

$$\forall x \in \mathbb{R}^d \text{ and } \epsilon > 0, \qquad \Pr\left[\tfrac{1}{\sqrt{t}}\|Gx\| \notin (1 \pm \epsilon)\|x\|\right] \le 2^{-\Omega(\epsilon^2 t)}.$$

Proof of Theorem 2.1. Assume without loss of generality that $\text{opt}_{vanilla}(P) = 1$. Let $C^* \subset \mathbb{R}^d$ be a set of optimal centers, and let $S \subset P$ be an output of Gonzalez's algorithm after $k+1$ steps, hence, $|S| = k+1$ and $\min_{s_1,s_2 \in S} \|s_1 - s_2\| \ge 1$. We treat $S \cup C^*$ as a *witness*, as described in Section 1.4. Pick $T = \Theta(\sqrt{d})$ with a hidden constant that satisfies Lemma 1.6, set $\epsilon = \frac{1}{2}$, and pick $t = O(\log k) + 400\frac{T^2}{\alpha^2}$ such that the bound $2^{-\Omega(t)}$ in Fact 2.2 is $\le \frac{1}{10(2k+1)^2}$. Let G be a $t \times d$ matrix of iid Gaussians $N(0,1)$. Then, by Fact 2.2 and a union bound,

$$\Pr\left[\forall p_1, p_2 \in S \cup C^*,\ \|G(p_1 - p_2)\| \in [1 \pm 0.5]\sqrt{t}\|p_1 - p_2\|\right] \ge 1 - (2k+1)^2 2^{-\Omega(t)} \ge \tfrac{9}{10}. \quad (2)$$

By Lemma 1.6,

$$\Pr\left[\forall x \in B(0,1),\ \|Gx\| \le T\right] \ge 1 - 2^{-\Omega(d)}. \quad (3)$$

Assuming the events in Equations (2) and (3) happen, the following holds.

Value. For every set $C_G \subset \mathbb{R}^t$ of size k, by restricting the pointset to the set S, and by that $|S| = k+1$ and $|C_G| = k$,

$$\max_{p\in P} \operatorname{dist}(Gp, C_G) \geq \max_{s\in S} \operatorname{dist}(Gs, C_G) \geq \tfrac{1}{2} \min_{s_1,s_2\in S} \|G(s_1 - s_2)\|$$

and now by the event in Equation (2) and by the choice of S,

$$\geq \tfrac{\sqrt{t}}{4} \min_{s_1,s_2\in S} \|s_1 - s_2\| \geq \tfrac{\sqrt{t}}{4} \geq \tfrac{T}{\alpha}.$$

In the other direction, for all $p \in P$, $\operatorname{dist}(p, C^*) \leq 1$, thus by using $G(C^*)$ as a center set, and by the event in Equation (3),

$$\operatorname{opt}_{vanilla}(G(P)) \leq \max_{p\in P} \operatorname{dist}(Gp, G(C^*)) \leq T.$$

Scaling G by a factor of $\frac{\alpha}{T}$ (i.e., the final map is $G' = \frac{\alpha}{T}G$) proves the first bullet. For ease of presentation, we analyze the second bullet (about preserving solutions) without this scaling.

Solution. We want to show that $C \subset P$ is an $O(\alpha)$-approximate k-center solution of P whenever $G(C)$ is a 2-approximate k-center solution of $G(P)$. We shall actually prove the contrapositive claim, and consider a set $B \subset P$ for which there exists a point $p' \in P$ such that $\operatorname{dist}(p', B) > 5\alpha$. Now for every point $p \in P$, denote by c_p its closest center from C^*. Then the following holds. By the triangle inequality,

$$\operatorname{dist}(Gp', G(B)) = \min_{b\in B} \|G(p' - b)\| \geq \min_{b\in B} \|G(c_{p'} - c_b)\| - \|G(c_{p'} - p')\| - \|G(b - c_b)\|$$

by the events in Equations 3 and 2,

$$\geq \min_{b\in B} \|G(c_{p'} - c_b)\| - 2T \geq \min_{b\in B} \tfrac{1}{2}\sqrt{t}\|c_{p'} - c_b\| - 2T$$

by the triangle inequality,

$$\geq \min_{b\in B} \tfrac{1}{2}\sqrt{t}(\|p' - b\| - \|c_{p'} - p'\| - \|b - c_b\|) - 2T$$

by assumptions,

$$\geq \tfrac{1}{2}\sqrt{t}5\alpha - \sqrt{t} - 2T \geq 20T.$$

Thus, $G(B)$ is not a 2-approximate k-center solution of $G(P)$. This concludes the proof of Theorem 2.1. The constant 2 in the approximation is arbitrary and could be changed to any other constant by adapting the other parameters. ◀

A streaming algorithm now follows as an immediate corollary. Indeed, just apply on the input the dimension reduction of Theorem 2.1, and then run a known streaming algorithm for low dimensions (slightly adapted to report a solution), as explained in Section 1.4.

▶ **Corollary 2.3** (Streaming Vanilla k-Center). *There is a randomized algorithm that, given as input numbers α, k, n, d, Δ and a set $P \subset [\Delta]^d$ of size at most n presented as a stream of* $\operatorname{poly}(n)$ *insertions and deletions of points, returns an $O(\alpha)$-approximation (value and solution) to k-center on P. The algorithm uses* $\operatorname{poly}(k2^{d/\alpha^2}\log(n\Delta))$ *bits of space and fails w.p. at most* $1/\operatorname{poly}(n)$.

3 Dimension Reduction for k-Center with Outliers

In this section, we design a moderate dimension reduction for k-center with z outliers, and demonstrate its application to streaming algorithms. We denote by $\mathrm{opt}_{outliers}(P)$ the optimal value of k-center with z outliers of P, and prove the following.

▶ **Theorem 3.1.** *Let $k, z \leq n$ and $d, \alpha > 1$. There is a random linear map $G : \mathbb{R}^d \to \mathbb{R}^t$ with $t = O(\frac{d}{\alpha^2} + \log(kz))$, such that for every set $P \subset \mathbb{R}^d$ of size n, with probability at least $2/3$,*

- $\mathrm{opt}_{outliers}(G(P)) \in \left[\mathrm{opt}_{outliers}(P), O(\alpha) \cdot \mathrm{opt}_{outliers}(P)\right]$, *and*
- *$C, Z \subset P$ is an $O(\alpha)$-approximate solution of k-center with z outliers for P whenever $G(C), G(Z)$ is an $O(1)$-approximate solution of k-center with z outliers for $G(P)$.*

We first show that every set $P \subset \mathbb{R}^d$ has a witness of size $O(kz)$, i.e., a subset $P' \subset P$ of size $|P'| = O(kz)$, such that $\mathrm{opt}_{outliers}(P') = \Omega(\mathrm{opt}_{outliers}(P))$. The construction is based on executing Gonzalez's algorithm $z + 1$ times, each execution without the previously chosen points. An important property of this construction is that for every choice of z outliers, one of the $z + 1$ executions of Gonzalez's algorithm returns a set of points without outliers. Our analysis uses a specific choice of z outliers that is optimized for P'.

▶ **Lemma 3.2** (Witness for the outliers variant)**.** *For every k, z and $P \subset \mathbb{R}^d$, there is a subset $P' \subset P$ of size $|P'| = (k + 1)(z + 1)$, such that $\mathrm{opt}_{outliers}(P') \in [\frac{1}{3}\mathrm{opt}_{outliers}(P), \mathrm{opt}_{outliers}(P)]$.*

▶ Remark 3.3. The size of this witness is tight up to low-order terms, by the following example. Consider a set X of $(k + 1)z + 1$ points, where there are $k + 1$ locations with pairwise distances 1, such that each location contains z points, and the last remaining point is at distance $\frac{1}{3}$ from one of these locations. Clearly, the cost of k-center with z outliers is $\frac{1}{6}$, by considering the z points from one of the locations as outliers. However, every strict subset of X (i.e., excluding even one point) has cost 0, as there will be a location with $z - 1$ points, which can be taken as outliers, together with that last point.

Proof. For a set X, we denote by $Gonz(X, k+1)$ a set of $k+1$ points computed by executing Gonzalez's algorithm (for k iterations) on X, breaking ties like the starting point arbitrarily. Given $P \subset \mathbb{R}^d$, construct a witness for k-center with z outliers as follows.

1. $X \leftarrow P$
2. for $i = 1, \dots, z + 1$
3. $\quad C_i \leftarrow Gonz(X, k + 1)$
4. $\quad X \leftarrow X \setminus C_i$
5. return $P' = \cup_{i\in[z+1]} C_i$ as a witness

Clearly, $\mathrm{opt}_{outliers}(P') \leq \mathrm{opt}_{outliers}(P)$. It remains to prove that $\mathrm{opt}_{outliers}(P') \geq \frac{1}{3}\mathrm{opt}_{outliers}(P)$.

Let C' and Z' be the optimal centers and outliers for P', respectively. Since $|Z'| \leq z$, there exists $i \in [z + 1]$ such that $C_i \cap Z' = \emptyset$. By the pigeonhole principle, there are $p_1, p_2 \in C_i$ that are clustered to the same cluster by C', and thus by a triangle inequality, $\mathrm{dist}(p_1, p_2) \leq 2\,\mathrm{opt}_{outliers}(P')$. Suppose without loss of generality that p_1 was added to C_i before p_2 in the execution of Gonzalez's algorithm. We can bound $\mathrm{opt}_{outliers}(P)$ by considering centers C' and outliers Z', and thus

$$\mathrm{opt}_{outliers}(P) \leq \max_{p\in P\setminus Z'} \mathrm{dist}(p, C') = \max\{\max_{p\in P\setminus P'} \mathrm{dist}(p, C'), \max_{p\in P'\setminus Z'} \mathrm{dist}(p, C')\}.$$

The second term is by definition $\text{opt}_{outliers}(P')$, so let us bound the first term. For every $p \in P \setminus P'$, by the triangle inequality,

$$\text{dist}(p, C') \leq \min_{p' \in C_i} \{\text{dist}(p, p') + \text{dist}(p', C')\} \leq \text{dist}(p, C_i) + \text{opt}_{outliers}(P'). \quad (4)$$

Let $\hat{C}_i \subseteq C_i$ be the set C_i *at the time that* p_2 *was chosen* (in the i-th execution of Gonzalez's algorithm). Then, because $p \notin P'$ is available at this time, and since $p_1 \in \hat{C}_i$,

$$\text{dist}(p, C_i) \leq \text{dist}(p, \hat{C}_i) \leq \text{dist}(p_2, \hat{C}_i) \leq \text{dist}(p_2, p_1) \leq 2\,\text{opt}_{outliers}(P').$$

Together with (4), we obtain $\text{dist}(p, C') \leq 3\,\text{opt}_{outliers}(P')$, which concludes the proof. ◀

Proof of Theorem 3.1. The proof that the value is preserved within factor $O(\alpha)$ using target dimension $t = O(\frac{d}{\alpha^2} + \log(kz))$ is the same as the proof for the vanilla variant, albeit with the witness given by Lemma 3.2. As for the proof that solutions are preserved, it only requires few minor changes, as follows.

Assume without loss of generality that $\text{opt}_{outliers}(P) = 1$. Let C^* and Z^* be sets of optimal centers and outliers, respectively. We can assume without loss of generality that the points in Z^* are furthest from C^*. Let G be a $t \times d$ matrix of iid Gaussians $N(0, 1)$, and pick $t \geq b_0 \log(kz)$, where $b_0 > 0$ is a fixed constant such that by *Fact* 2.2 and a union bound,

$$\Pr\left[\forall p_1, p_2 \in P' \cup C^* \cup Z^*,\ \|G(p_1 - p_2)\| \in [1 \pm 0.5]\sqrt{t}\|p_1 - p_2\|\right] \geq \tfrac{9}{10}. \quad (5)$$

For every point $p \in P \setminus Z^*$, denote by c_p its closest center from C^*. We slightly abuse notation, and for every $p \in Z^*$ we denote by c_p the point p itself. The proof proceeds by considering sets $B \subset P$ of size k and $Z' \subset P$ of size z for which there exists a point $p' \in P \setminus Z'$ such that $\text{dist}(p', B) > 5\alpha$ and is the same as the proof of Theorem 2.1. ◀

A streaming algorithm now follows as a corollary, using a known streaming algorithm for low dimensions (slightly adapted to report a solution), as explained in Section 1.4.

▶ **Corollary 3.4** (Streaming Algorithm for k-Center with z Outliers)**.** *There is a randomized algorithm that, given as input numbers $\alpha, k, z, n, d, \Delta$ and a set $P \subset [\Delta]^d$ of size at most n presented as a stream of* $\text{poly}(n)$ *insertions and deletions of points, returns an $O(\alpha)$-approximation (value and solution) to k-center with z outliers on P. The algorithm uses* $\text{poly}(kz2^{d/\alpha^2}\log(n\Delta))$ *bits of space and fails with probability at most* $1/\text{poly}(n)$*.*

4 Dimension Reduction for k-Center with an Assignment Constraint

In this section, we consider k-center with an assignment constraint, which captures the capacitated and fair variants of k-center, as described in Section 1.1. We design for this problem a moderate dimension reduction, and demonstrate its application to streaming algorithms. Our definition below of an assignment constraint follows the one used in [51, 34, 8, 12] for other k-clustering problems. The *radius* of a pointset is the optimal value of 1-center clustering for it.

▶ **Definition 4.1.** *An* assignment *is a map $\pi : [n] \to [k]$. An* assignment constraint *is a partition of all possible assignments into feasible and infeasible ones, formalized as $\mathcal{C} : [k]^n \to \{0, 1\}$.*

This definition can model clustering with capacity $L > 0$, by declaring an assignment π to be feasible if $|\pi^{-1}(i)| \leq L$ for all $i \in [k]$. To exemplify how it can model fair clustering, suppose the first $n/3$ points in P are colored blue and the others are red; then declare π to be feasible if in every $\pi^{-1}(i)$, exactly 1/3 of the elements are from the range $\{1, \dots, \frac{n}{3}\}$. See the full version for more details.

▶ **Definition 4.2.** *In k-center with an assignment constraint $\mathcal{C}$, the input is a set $P \subset \mathbb{R}^d$ of n points, and the goal is to partition P into k sets (called clusters) in a manner feasible according to $\mathcal{C}$ when viewed as an assignment,*[7] *so as to minimize the maximum cluster radius. The minimum value attained is denoted by $\mathrm{opt}_{\mathcal{C}}(P)$. A* solution *to this problem is a partition of P into k sets, and it is called α*-approximate *if it is feasible and has value at most $\alpha \cdot \mathrm{opt}_{\mathcal{C}}(P)$.*

The next theorem shows that reducing to dimension $O(\frac{d}{\alpha} + \log k)$ preserves w.h.p. the value of k-center with an assignment constraint up to an $O(\alpha)$ factor, simultaneously for all assignment constraints. Our dimension bound here is worse by factor α compared to the vanilla variant, essentially because our lower bound for the value of $G(P)$ is weaker. More precisely, take as before $t = O(\frac{d}{\alpha^2} + \log k)$ and let G to be a matrix of iid Gaussians $N(0, \frac{1}{t})$. We show that w.h.p., for all constraints $\mathcal{C}$ we have $\frac{1}{\alpha}\,\mathrm{opt}_{\mathcal{C}}(P) \leq \mathrm{opt}_{\mathcal{C}}(G(P)) \leq \alpha\,\mathrm{opt}_{\mathcal{C}}(P)$. The upper bound here is as before, but the lower bound is weaker by factor α, hence we have to scale G appropriately and substitute α^2 with α' to conclude the theorem.

▶ **Theorem 4.3.** *Let $d, \alpha > 1$ and $k \leq n$. There is a random linear map $G : \mathbb{R}^d \to \mathbb{R}^t$ with $t = O(\frac{d}{\alpha} + \log k)$, such that for every set $P \subset \mathbb{R}^d$ of size n, with probability at least $2/3$, the following holds. For all $\mathcal{C} : [k]^n \to \{0, 1\}$,*

- $\mathrm{opt}_{\mathcal{C}}(G(P)) \in \big[\mathrm{opt}_{\mathcal{C}}(P), O(\alpha) \cdot \mathrm{opt}_{\mathcal{C}}(P)\big]$*, and*
- *a feasible partition of P to k clusters is an $O(\alpha)$-approximate solution of k-center with assignment constraint $\mathcal{C}$ whenever the corresponding partition of $G(P)$ is an $O(1)$-approximate solution of k-center with assignment constraint $\mathcal{C}$ for $G(P)$.*

The proof of Theorem 4.3, and its applications to streaming algorithms for capacitated and fair k-center are provided in the full version.

References

1 Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003. `doi:10.1016/S0022-0000(03)00025-4`.
2 Pankaj K. Agarwal and Cecilia Magdalena Procopiuc. Exact and approximation algorithms for clustering. *Algorithmica*, 33(2):201–226, 2002. `doi:10.1007/s00453-001-0110-y`.
3 Pankaj K. Agarwal and R. Sharathkumar. Streaming algorithms for extent problems in high dimensions. *Algorithmica*, 72(1):83–98, 2015. `doi:10.1007/s00453-013-9846-4`.
4 Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, STOC*, pages 557–563, 2006. `doi:10.1145/1132516.1132597`.
5 Evangelos Anagnostopoulos, Ioannis Z. Emiris, and Ioannis Psarros. Randomized embeddings with slack and high-dimensional approximate nearest neighbor. *ACM Trans. Algorithms*, 14(2):18:1–18:21, 2018. `doi:10.1145/3178540`.
6 Alexandr Andoni, Khanh Do Ba, Piotr Indyk, and David P. Woodruff. Efficient sketches for earth-mover distance, with applications. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009*, pages 324–330. IEEE Computer Society, 2009. `doi:10.1109/FOCS.2009.25`.

[7] To view a partition as an assignment, represent P by $[n]$ and the clusters by $[k]$, in an arbitrary manner (not by the geometry of the points).

7 Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Earth mover distance over high-dimensional spaces. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 343–352, 2008. URL: http://dl.acm.org/citation.cfm?id=1347082.1347120.

8 Sayan Bandyapadhyay, Fedor V. Fomin, and Kirill Simonov. On coresets for fair clustering in metric and euclidean spaces and their applications. In *48th International Colloquium on Automata, Languages, and Programming, ICALP*, volume 198 of *LIPIcs*, pages 23:1–23:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPICS.ICALP.2021.23.

9 Judit Bar-Ilan, Guy Kortsarz, and David Peleg. How to allocate network centers. *J. Algorithms*, 15(3):385–415, 1993. doi:10.1006/jagm.1993.1047.

10 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. doi:10.1145/3313276.3316318.

11 Christos Boutsidis, Anastasios Zouzias, and Petros Drineas. Random projections for k-means clustering. In *24th Annual Conference on Neural Information Processing Systems, NeurIPS*, pages 298–306. Curran Associates, Inc., 2010. URL: https://proceedings.neurips.cc/paper/2010/hash/73278a4a86960eeb576a8fd4c9ec6997-Abstract.html.

12 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. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 462–473, 2022. doi:10.1109/FOCS54457.2022.00051.

13 Vladimir Braverman, Gereon Frahling, Harry Lang, Christian Sohler, and Lin F. Yang. Clustering high dimensional dynamic data streams. In *Proceedings of the 34th International Conference on Machine Learning, ICML*, volume 70 of *Proceedings of Machine Learning Research*, pages 576–585. PMLR, 2017. URL: http://proceedings.mlr.press/v70/braverman17a.html.

14 Matteo Ceccarello, Andrea Pietracaprina, and Geppino Pucci. Solving k-center clustering (with outliers) in MapReduce and streaming, almost as accurately as sequentially. *Proc. VLDB Endow.*, 12(7):766–778, 2019. doi:10.14778/3317315.3317319.

15 Moses Charikar, Chandra Chekuri, Tomás Feder, and Rajeev Motwani. Incremental clustering and dynamic information retrieval. *SIAM J. Comput.*, 33(6):1417–1440, 2004. doi:10.1137/S0097539702418498.

16 Moses Charikar and Erik Waingarten. The Johnson-Lindenstrauss Lemma for clustering and subspace approximation: From coresets to dimension reduction. *CoRR*, 2022. arXiv:2205.00371.

17 Xi Chen, Vincent Cohen-Addad, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Streaming Euclidean MST to a Constant Factor. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC*, pages 156–169, 2023. doi:10.1145/3564246.3585168.

18 Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. New streaming algorithms for high dimensional EMD and MST. In *54th Annual Symposium on Theory of Computing, STOC*, pages 222–233. ACM, 2022. doi:10.1145/3519935.3519979.

19 Xiaoyu Chen, Shaofeng H.-C. Jiang, and Robert Krauthgamer. Streaming Euclidean Max-Cut: Dimension vs data reduction. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC*, pages 170–182, 2023. doi:10.1145/3564246.3585170.

20 Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Annual Conference on Neural Information Processing Systems, NeurIPS*, pages 5029–5037, 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/978fce5bcc4eccc88ad48ce3914124a2-Abstract.html.

21 Kenneth L. Clarkson. Nearest neighbor queries in metric spaces. *Discret. Comput. Geom.*, 22(1):63–93, 1999. doi:10.1007/PL00009449.

22 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*, pages 163–172, 2015. doi:10.1145/2746539.2746569.

23 Vincent Cohen-Addad, Chris Schwiegelshohn, and Christian Sohler. Diameter and k-Center in Sliding Windows. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.19.

24 Artur Czumaj, Arnold Filtser, Shaofeng H-C Jiang, Robert Krauthgamer, Pavel Veselỳ, and Mingwei Yang. Streaming facility location in high dimension via geometric hashing. *CoRR*, 2022. The latest version has additional results compared to the preliminary version in [25]. arXiv:2204.02095.

25 Artur Czumaj, Shaofeng H.-C. Jiang, Robert Krauthgamer, Pavel Veselý, and Mingwei Yang. Streaming facility location in high dimension via geometric hashing. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 450–461, 2022. doi:10.1109/FOCS54457.2022.00050.

26 Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson–Lindenstrauss transform. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC*, pages 341–350, 2010. doi:10.1145/1806689.1806737.

27 Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, 2003. doi:10.1002/rsa.10073.

28 Mark de Berg, Leyla Biabani, and Morteza Monemizadeh. k-center clustering with outliers in the MPC and streaming model. In *IEEE International Parallel and Distributed Processing Symposium, IPDPS 2023*, pages 853–863. IEEE, 2023. doi:10.1109/IPDPS54959.2023.00090.

29 Mark de Berg, Morteza Monemizadeh, and Yu Zhong. k-center clustering with outliers in the sliding-window model. In *29th Annual European Symposium on Algorithms, ESA*, volume 204 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ESA.2021.13.

30 Peter Frankl and Hiroshi Maehara. Some geometric applications of the beta distribution. *Ann. Inst. Stat. Math.*, 42(3):463–474, 1990. doi:10.1007/BF00049302.

31 Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, 38:293–306, 1985. doi:10.1016/0304-3975(85)90224-5.

32 Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science, FOCS*, pages 534–543. IEEE Computer Society, 2003. doi:10.1109/SFCS.2003.1238226.

33 Wei Hu, Zhao Song, Lin F. Yang, and Peilin Zhong. Nearly optimal dynamic k-means clustering for high-dimensional data. *CoRR*, 2018. arXiv:1802.00459.

34 Lingxiao Huang, Shaofeng H.-C. Jiang, and Nisheeth K. Vishnoi. Coresets for clustering with fairness constraints. In *Advances in Neural Information Processing Systems 32, NeurIPS*, pages 7587–7598, 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/810dfbbebb17302018ae903e9cb7a483-Abstract.html.

35 Piotr Indyk. Better algorithms for high-dimensional proximity problems via asymmetric embeddings. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 539–545, 2003. URL: http://dl.acm.org/citation.cfm?id=644108.644200.

36 Piotr Indyk. Algorithms for dynamic geometric problems over data streams. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, STOC, pages 373–380, 2004. doi:10.1145/1007352.1007413.

37 Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, STOC*, pages 604–613, 1998. doi:10.1145/276698.276876.

38 Piotr Indyk and Assaf Naor. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31, 2007. doi:10.1145/1273340.1273347.

39 William Johnson and Joram Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, January 1984. doi:10.1090/conm/026/737400.

40 Samir Khuller and Yoram J. Sussmann. The capacitated K-center problem. *SIAM J. Discret. Math.*, 13(3):403–418, 2000. doi:10.1137/S0895480197329776.

41 Sang-Sub Kim and Hee-Kap Ahn. An improved data stream algorithm for clustering. *Comput. Geom.*, 48(9):635–645, 2015. doi:10.1016/j.comgeo.2015.06.003.

42 Boaz Klartag and Shahar Mendelson. Empirical processes and random projections. *Journal of Functional Analysis*, 225(1):229–245, 2005. doi:10.1016/j.jfa.2004.10.009.

43 Christiane Lammersen. *Approximation Techniques for Facility Location and Their Applications in Metric Embeddings*. PhD thesis, Dortmund, Technische Universität, 2010.

44 Christiane Lammersen, Anastasios Sidiropoulos, and Christian Sohler. Streaming embeddings with slack. In *11th International Symposium on Algorithms and Data Structures, WADS*, volume 5664 of *Lecture Notes in Computer Science*, pages 483–494. Springer, 2009. doi:10.1007/978-3-642-03367-4_42.

45 Kasper Green Larsen and Jelani Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 633–638, 2017. doi:10.1109/FOCS.2017.64.

46 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. doi:10.1145/3313276.3316350.

47 Jirí Matousek. On variants of the Johnson-Lindenstrauss Lemma. *Random Struct. Algorithms*, 33(2):142–156, 2008. doi:10.1002/rsa.20218.

48 Richard Matthew McCutchen and Samir Khuller. Streaming algorithms for k-center clustering with outliers and with anonymity. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, volume 5171 of *Lecture Notes in Computer Science*, pages 165–178. Springer, 2008. doi:10.1007/978-3-540-85363-3_14.

49 Shyam Narayanan, Sandeep Silwal, Piotr Indyk, and Or Zamir. Randomized dimensionality reduction for facility location and single-linkage clustering. In *Proceedings of the 38th International Conference on Machine Learning, ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 7948–7957. PMLR, 2021. URL: http://proceedings.mlr.press/v139/narayanan21b.html.

50 Jelani Nelson. Dimensionality reduction in Euclidean space. *Notices of the American Mathematical Society*, 67(10):1498–1507, 2020. doi:10.1090/noti2166.

51 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*, volume 11926 of *Lecture Notes in Computer Science*, pages 232–251. Springer, 2019. doi:10.1007/978-3-030-39479-0_16.

52 Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi:10.1017/9781108231596.

53 David P. Woodruff and Taisuke Yasuda. High-dimensional geometric streaming in polynomial space. In *63rd Annual Symposium on Foundations of Computer Science, FOCS*, pages 732–743. IEEE, 2022. doi:10.1109/FOCS54457.2022.00075.