Coresets for 1-Center in ℓ_1 Metrics

Amir Carmel
Weizmann Institute of Science, Rehovot, Israel

Chengzhi Guo
Peking University, China

Shaofeng H.-C. Jiang
Peking University, China

Robert Krauthgamer
Weizmann Institute of Science, Rehovot, Israel

Abstract

We explore the applicability of coresets – a small subset of the input dataset that approximates a predefined set of queries – to the 1-center problem in ℓ_1 spaces. This approach could potentially extend to solving the 1-center problem in related metric spaces, and has implications for streaming and dynamic algorithms.

We show that in ℓ_1, unlike in Euclidean space, even weak coresets exhibit exponential dependency on the underlying dimension. Moreover, while inputs with a unique optimal center admit better bounds, they are not dimension independent. We then relax the guarantee of the coreset further, to merely approximate the value (optimal cost of 1-center), and obtain a dimension-independent coreset for every desired accuracy $\epsilon > 0$. Finally, we discuss the broader implications of our findings to related metric spaces, and show explicit implications to Jaccard and Kendall's tau distances.

2012 ACM Subject Classification Theory of computation → Sketching and sampling; Theory of computation → Streaming models

Keywords and phrases clustering, k-center, minimum enclosing balls, coresets, ℓ_1 norm, Kendall's tau, Jaccard metric

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.28

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, by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center, and by a research grant from the Estate of Harry Schutzman. Part of this work was done while visiting the Simons Institute for the Theory of Computing.

1 Introduction

Clustering is a fundamental task in unsupervised learning and data analysis in general, with wide-ranging applications. Typically, the input is a dataset of n points in $\mathbb{R}^d$ or some other feature space of dimension d, and the objective is to partition the dataset into clusters, each characterized by a high degree of similarity. In the current era of big data, both the number of points and their dimension are often excessive, making the computational demands significant. In center-based clustering, the goal is to further assign to each cluster a representative "center" point. The number of centers (and ergo clusters) is often specified in advance, and denoted by k, giving rise to the fundamental k-center, k-median and k-mean problems, which are famous for their simplicity and broad applicability. The case $k = 1$ is particularly important as the most basic setting, and focuses not on partitioning the points but rather on aggregating them, by identifying a representative that best captures the entire dataset.

16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 28; pp. 28:1–28:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

We study the 1-center problem in ℓ_1 metrics (i.e., Manhattan distance). The motivation for ℓ_1 metrics is twofold. First, ℓ_1 metrics arise in many practical settings, especially when the dimension is large. Second, common discrete metric spaces, such as the Hamming, Kendall's tau, and Jaccard metrics, are closely related to ℓ_1, as they can be embedded into $\mathbb{R}^d$ endowed with the ℓ_1 metric with all distances preserved isometrically (i.e., with no distortion).

The 1-center problem has been investigated extensively in many metric spaces. In the context of strings, 1-center under the Hamming distance and Edit Distance is a classical problem known as the Closest String Problem [26, 27, 29]. In the context of permutations (a common model for rankings), 1-center under Kendall's tau distance is known as the maximum rank aggregation problem [5, 32]. The 1-center problem has also been studied in other discrete metrics, such as the Ulam distance between permutations and the Jaccard distance between sets [9, 13]. Additionally, recent work studied the 1-center problem under various metrics, including ℓ_p, edit distance, and Ulam, specifically to examine how the time complexity depends on the dimension d [1]. In fact, the 1-center problem is known to be NP-hard in many metric spaces, and it seems that each one requires distinct algorithmic techniques. For some metrics, a PTAS is known, while for others, the existence of a PTAS or even a non-trivial approximation algorithm remains an open question and an active area of research.

A coreset is a data-summarization concept, where the input dataset is replaced by a small subset that approximates its clustering properties. It can lead to significant reductions in computational resources, particularly storage and communication, and plays a pivotal role in modern algorithmic settings, such as streaming, distributed, and dynamic. It is thus crucial to understand the tradeoff between the size of a coreset and its accuracy (approximation factor). Coresets have been studied intensely since their introduction by [2], see the surveys [3, 21, 28, 31], and they have been successfully applied in a broad range of settings, for example different clustering objectives and metric spaces. Surprisingly, little is known about coresets in ℓ_1, particularly in high dimension, leaving significant gaps in our understanding. Small coresets in ℓ_1, if they exist and can be computed efficiently, could potentially pave the way for a unified approach for 1-center also in related metrics, like Kendall's tau and Ulam. Moreover, studying coresets for these metrics may provide insights into the geometric structure of these metric spaces.

1.1 Problem setup

In the 1-*center problem*, also known as *Minimum Enclosing Ball (MEB)*, the input is a set P of points in a metric space (X, dist), and the goal is to find a center point $c \in X$ that minimizes the objective function $\mathrm{cost}(c, P) := \max_{p \in P} \mathrm{dist}(p, c)$. Denote the optimal value by $\mathrm{opt}(P) := \min_{c \in X} \mathrm{cost}(c, P)$, and observe that it is monotone, i.e., $Q \subseteq P$ implies $\mathrm{opt}(Q) \leqslant \mathrm{opt}(P)$. While we mostly consider the space $(\mathbb{R}^d, \ell_1)$, these definitions capture also other metrics, like Kendall's tau distance where X is the set of permutations over $[d]$.

There are different (and sometimes inconsistent) definitions for coresets, which were usually developed when the desirable guarantees were deemed impossible or difficult to prove, leading to more relaxed guarantees. We face this same issue (especially for high dimension), and thus consider a sequence of definitions. For brevity, we present only the definitions most relevant to our work, and do not discuss possible generalizations (e.g., to k-center) or variants that appear in the literature. A common and very useful type of coreset is a *strong coreset*, which approximates the cost of every center, as follows.

▶ **Definition 1.1** (Strong Coreset). *A subset $Q \subseteq P$ is a* strong ϵ-coreset *for a* 1*-center instance $P \subseteq X$ if*

$$\forall c \in X, \qquad \operatorname{cost}(c, P) \leqslant (1+\epsilon) \operatorname{cost}(c, Q). \tag{1}$$

The guarantee in (1) is extremely effective in reducing (the task of solving) P to Q, with a small loss in the objective value. In Euclidean spaces, strong coresets for 1-center are known, however their size is exponential in the dimension, namely, $|Q| \leqslant (1/\epsilon)^{O(d)}$ [3], and unfortunately this upper bound is existentially tight.[1] The upper bound actually extends to every metric space with doubling dimension d [4, 10, 20], and therefore applies to $\mathbb{R}^d$ endowed with the ℓ_1 norm (or every other norm).

It is thus natural to ask: Can the coreset size be improved specifically for ℓ_1 norm? If not, can weaker notions lead to polynomial (in d) or even dimension-independent size?

1.2 Our results

Weak coresets. We point our attention to a more relaxed notion, called *weak coreset*, which only approximates the cost of certain centers, namely, centers that may be found by solving the coreset instance. Trivially, every strong ϵ-coreset is also a weak ϵ-coreset.

▶ **Definition 1.2** (Weak Coreset). *A subset $Q \subseteq P$ is a* weak ϵ-coreset *for a* 1*-center instance $P \subseteq X$ if*

$$\forall c^Q \in \operatorname*{argmin}_{c \in X} \operatorname{cost}(c, Q), \qquad \operatorname{cost}(c^Q, P) \leqslant (1+\epsilon) \operatorname{cost}(c^Q, Q). \tag{2}$$

Weak coresets were introduced in [8] specifically for 1-center in Euclidean space. It is well-known that in Euclidean space, there is a unique optimal solution,[2] which is why previous work often refers to c^Q as *the* optimal center. The challenge then becomes how to extend this definition to other metric spaces, like ℓ_1 norm, that might have multiple optimal solutions. The typical use of a coreset Q is to apply an algorithm on Q and take the resulting center c^Q as an approximated solution for the original instance P. Since a generic algorithm might pick any of the multiple optimal centers for Q, it is essential to have a guarantee that applies to every possible optimal solution c^Q for Q, hence the universal quantifier ("for all") in (2).

Remarkably, in Euclidean space, every instance P admits a weak ϵ-coreset of size $\lceil 1/\epsilon \rceil$, which is independent of the input size $n = |P|$ and of the Euclidean dimension d, and is in fact tight [7]. This result improved an earlier $O(1/\epsilon^2)$ bound from [8], which was also dimension independent. Can similar bounds be proved for ℓ_1 metrics?

We demonstrate that, in sharp contrast to the ℓ_2 setting, weak coresets in ℓ_1 are dimension dependent and moreover grow exponentially with d, even for a fixed $\epsilon > 0$. At the same time, we show how to construct a strong 0-coreset of size 2^d, which again contrasts with the ℓ_2 setting (and doubling metric spaces), where coresets grow with $1/\epsilon$ and are thus not applicable for $\epsilon = 0$. It is well-known that a strong coreset is *composable*, i.e., the union of coresets is itself a coreset for the union of the original datasets, which is very useful for designing algorithms in the streaming and dynamic settings. Thus, our coreset construction can be used in these settings, particularly in low dimension.

[1] There is a folklore lower bound, which follows by considering an instance P formed by an ϵ-net of the unit sphere S^{d-1}; every strong coreset must contain all of P and thus have size $(1/\epsilon)^{\Omega(d)}$.

[2] The uniqueness follows from the fact that ℓ_2 is strictly convex, whereas ℓ_1 is convex but not strictly convex.

▶ **Theorem 1.3.** *Consider the* 1*-center problem in* $\mathbb{R}^d$ *under* ℓ_1 *norm.*

(a) *Every instance* $P \subset \mathbb{R}^d$ *admits a strong* 0*-coreset of size* 2^d*.*

(b) *There exists* $P \subset \mathbb{R}^d$*, such that every weak* ϵ*-coreset for* $\epsilon < \frac{1}{3}$ *must have size* $2^{\Omega(d)}$*.*

We prove this theorem in Section 2. The upper bound leverages the unique geometry of ℓ_1, where only 2^d directions are in effect important. The lower bound builds a hard instance and employs classical techniques from coding theory, particularly the Hamming bound. We remark that an unpublished result from [30, Theorem 5], for a related problem about a convex shape in ℓ_2 metric, seems to imply a weak ϵ-coreset of size $\mathrm{poly}(d/\epsilon)$ for our problem of 1-center in ℓ_1 metrics. This would contradict our lower bound in Theorem 1.3, indicating that its statement is inaccurate or its proof is flawed; see also a similar discussion in [6].

Due to the exponential dependency on d in the general case, it is important to identify cases with a smaller coreset size, say $\mathrm{poly}(d)$ or even dimension independent. Additionally, one may suspect that the huge disparity between ℓ_1 and Euclidean space, which does admit a dimension-independent weak coreset [7, 8], is the fact that in ℓ_1 the optimal center need not be unique. We thus study the special case of ℓ_1 where the optimal center is unique, which may occur naturally in certain contexts, or be "enforced" by adding at most two points to the input dataset. We provide bounds that are exponentially smaller than in the general case, but still depend on d, which establishes that ℓ_1 is inherently more complex than ℓ_2.

▶ **Theorem 1.4.** *Consider the* 1*-center problem in* $\mathbb{R}^d$ *under* ℓ_1 *norm, on instances that have a unique optimal center.*

(a) *Every instance* $P \subset \mathbb{R}^d$ *with unique optimum admits a weak* 0*-coreset of size* $2d$*.*

(b) *There exists* $P \subset \mathbb{R}^d$ *with unique optimum, such that every weak* ϵ*-coreset for fixed* $\epsilon < 1$ *must have size* $\Omega(\log d)$*.*

We prove this theorem in Section 3. We further conjecture that the lower bound can be improved to $\Omega(d)$, and prove it in the special case $\epsilon = 0$ using Hadamard code. The upper bound in our theorem employs tools from convex geometry, specifically the Steinitz Theorem, which is relatively obscure, to prove the existence non-constructively (i.e., without an efficient algorithm). For the lower bound, we consider the instance $P = \{\pm 1\}^d$, which has a unique optimal center at the origin, and show that a coreset Q that is too small, must have an optimal center c^Q in which most coordinates are ± 1, and this center is clearly a poor solution for P.

Value-preserving coreset. To further reduce the coreset size, we consider an even less restrictive variant that preserves solely the objective function $\mathrm{opt}(P)$, without imposing any conditions on the optimal centers of the coreset. This concept, which we shall refer to as a *value-preserving coreset*, has been studied previously under the general term "coreset" [6, 14, 33], and it is most useful in applications that focus on measuring the similarity among points in the cluster, rather than identifying a center. It is easy to verify that every weak coreset is also a value-preserving coreset.

▶ **Definition 1.5** (Value-Preserving Coreset). *A subset* $Q \subseteq P$ *is a* value-preserving ϵ-coreset *for a* 1*-center instance* $P \subseteq X$ *if*

$$\mathrm{opt}(P) \leqslant (1+\epsilon)\,\mathrm{opt}(Q). \tag{3}$$

We show that there exist value-preserving ϵ-coresets of size $\tilde{O}(1/\epsilon^2)$, which is dimension independent, in contrast with Theorems 1.3 and 1.4. We thus establish a sharp separation between preserving an optimal solution (center point) or only preserving its value. This result is stated in the next theorem and proved in Section 4. It also answers a question posed in [6], which studies a generalization of 1-center, called the Minimum Enclosing Polytope problem, and asks for polytopes that admit a dimension-independent value-preserving coreset.[3]

▶ **Theorem 1.6.** *Consider the* 1*-center problem in* $\mathbb{R}^d$ *under* ℓ_1 *norm. Every instance* $P \subset \mathbb{R}^d$ *admits a value-preserving* ϵ*-coreset of size* $O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$.

We prove this theorem in Section 4. Our algorithm is based on random sampling, a technique that was rarely used (if at all) for 1-center coresets, but widely used for k-median and k-means coresets [15, 18, 22]. This is very natural because sampling is an excellent fit for objectives like k-median that that are formed by summation, but not for objectives like k-center that are formed by maximization, which are sensitive to missing even one term. It is thus not surprising that our sampling is conducted not on the 1-center problem but rather on its dual. More precisely, we formulate 1-center as a linear programming (LP) problem and write its dual problem, then solve this dual problem to obtain sampling probabilities for the input points. Crucially, the dual objective is formed by summation, and is thus conducive to sampling. Compared with the primal-dual framework, our algorithm is dual-only and it solves the dual LP explicitly, which is then related in the analysis to the primal via strong duality. Our sampling is similar to the randomized rounding commonly used in approximation algorithms, however its purpose here is to generate a coreset instead of a solution, and also the analysis is somewhat different.

Related metrics and applications. We demonstrate that our results for weak coresets in ℓ_1 (in Sections 2 and 3) apply also to other discrete metric spaces. Specifically, we provide in Section 5 examples that use Kendall's tau and Jaccard metrics, drawing inspiration from the results obtained in the ℓ_1 setting.

1.3 Related work

Another closely related aggregation problem is 1-median. This problem, unlike 1-center, can be solved in ℓ_1 metrics easily, by simply taking in each coordinate the median value of the input points. Utilizing this structure, 1-median in ℓ_1 metrics has a weak ϵ-coreset of size $O(\epsilon^{-2}\log(1/\epsilon))$, which is independent of the dimension [19], in sharp contrast to 1-center. The 1-median problem also admits a PTAS (approximation arbitrarily close to 1 in polynomial time) in many other metrics, including Kendall's tau distance [25], Jaccard distance [16], and a near-linear time PTAS in ℓ_2 [17]. However, it is unclear if a PTAS exists for edit distance, or even for a special case called Ulam distance, and the best upper bound is $(2-\rho)$-approximation for some fixed $\rho > 0$ [11, 12].

2 Weak ϵ-coresets in ℓ_1

2.1 Construction of strong 0-coreset of size 2^d

In this section we will prove the following theorem:

▶ **Theorem 2.1.** *Every instance* $P \subset \mathbb{R}^d$ *admits a strong* 0*-coreset of size* 2^d.

[3] Prior to our result showing that the ℓ_1 ball (also known as the cross-polytope) admits dimension-independent value-preserving coreset, the only known example was the parallelotope, which trivially has a value-preserving 0-coreset with only 2 points.

Proof. Let $\Sigma = \{-1,1\}^d$ denote the set of all sign vectors in $\mathbb{R}^d$. For each sign vector $\sigma \in \Sigma$, let p_σ be a point in P such that the inner product $p_\sigma \cdot \sigma$ is maximized. Define $Q_\Sigma = \{p_\sigma : \sigma \in \Sigma\}$, clearly $|Q_\Sigma| \leqslant |\Sigma| = 2^d$.

Let $c \in \mathbb{R}^d$. Since Q_Σ is a subset of P, we have that $\operatorname{cost}(c, Q_\Sigma) \leqslant \operatorname{cost}(c, P)$. Let p be some point in P, then there exists a sign vector $\sigma \in \Sigma$ for which we can write the distance between p and c as:

$$\|p - c\|_1 = \sigma \cdot (p - c) = \sigma \cdot p - \sigma \cdot c$$

By definition of Q_Σ, there exists a point $p_\sigma \in Q_\Sigma$ satisfying $\sigma \cdot p \leqslant \sigma \cdot p_\sigma$. Thus,

$$\|p - c\|_1 = \sigma \cdot p - \sigma \cdot c \leqslant \sigma \cdot p_\sigma - \sigma \cdot c \leqslant \sigma \cdot (p_\sigma - c) \leqslant |\sigma \cdot (p_\sigma - c)| \leqslant \|p_\sigma - c\|_1 \leqslant \operatorname{cost}(c, Q_\Sigma).$$

That is,

$$\operatorname{cost}(c, P) = \max_{p \in P} \|p - c\|_1 \leqslant \operatorname{cost}(c, Q_\Sigma)$$ ◀

2.2 Lower bound of $2^{\Omega(d)}$ for weak ϵ-coresets

In this section, we present a set $P \subseteq \mathbb{R}^d$ of size $2^{d(1-o(1))}$ that has only a trivial 0-coreset, namely, only itself. Furthermore, weak ϵ-coresets of P, for $\epsilon \in (0, \frac{1}{3})$, must have size $2^{\Omega(d)}$.

▶ **Theorem 2.2.** *There exists a set $P \subseteq \mathbb{R}^d$ of size $2^{d(1-o(1))}$, such that*

- *if $Q \subseteq P$ is a weak 0-coreset of P, then $Q = P$; and*
- *for all $\epsilon \in [0, \frac{1}{3})$, if $Q \subseteq P$ is a weak ϵ-coreset of P, then $|Q| \geqslant 2^{\Omega(d)}$.*

Proof. We may assume without loss of generality that d is even, as otherwise we can take the construction for $d - 1$ and append 0 to all the points (vectors).

Denote by $\vec{1}$ the vector $(1, 1, \ldots, 1)$, and let $B \subseteq \{\pm 1\}^d$ be the set of points that are balanced, in the sense that they have an equal number of 1 and -1 coordinates, i.e., $B = \{b \in \{\pm 1\}^d : \sum_{i=1}^d b_i = 0\}$. Fix $0 < \delta \leqslant \frac{1}{3}$, and let $P = \{-\vec{1}, \vec{1}\} \bigcup (1-\delta)B$, where $(1-\delta)B$ denotes the set of points in B scaled by factor $1 - \delta$. Thus, $|P| = 2 + \binom{d}{d/2} = \Theta(2^d/\sqrt{d})$. One can verify that the origin $\vec{0} \in \mathbb{R}^d$ is an optimal center, with value d. However it is not unique; to see this, observe that the antipodal pair $\{-\vec{1}, \vec{1}\}$ establishes the optimal value $\operatorname{opt}(P) = d$, thus an optimal center must have the same distance d to each of them, which holds for points $x \in [-1, 1]^d$ that lie on the hyperplane $\sum_{i=1}^d x_i = 0$. The set $(1 - \delta)B$ restricts the optimal solutions (inside that hyperplane) in a delicate manner, as we shall see.

We first show that P is the only weak 0-coreset of P. Let $Q \subseteq P$ be a weak 0-coreset, and notice that $\operatorname{opt}(Q) = \operatorname{opt}(P) = d$, implying that Q must contain the antipodal pair $-\vec{1}$ and $\vec{1}$. Assume towards contradiction that Q is a proper subset of P. Then there exists $(1 - \delta)\tilde{b} \in P \backslash Q$ for some $\tilde{b} \in B$. Let $\eta = \frac{1}{d}$ and consider $c^* = -(1 + \eta)\delta\tilde{b}$; it cannot be an optimal center of P, because

$$\left\|-(1 + \eta)\delta\tilde{b} - (1 - \delta)\tilde{b}\right\|_1 = \left\|(-1 - \eta\delta)\tilde{b}\right\|_1 = (1 + \eta\delta) \left\|\tilde{b}\right\|_1 > d.$$

We next show that this point $c^* = -(1 + \eta)\delta\tilde{b}$ is an optimal center of Q. Indeed, observe that $\delta \leqslant \frac{1}{3}$ implies $(1 + \eta)\delta \leqslant 1 - \delta$, and thus the sign of $1 - \delta$ will not change if we add/subtract to it $(1 + \eta)\delta$. Let $\operatorname{sgn}(x) \in \{-1, 0, 1\}$ denote the sign function. Now consider $a \in Q$; its distance to $-(1 + \eta)\delta\tilde{b}$ is:

$$\begin{aligned}\left\|a+(1+\eta)\delta\tilde{b}\right\|_1 &= \sum_{i=1}^{d}\left|a_i+(1+\eta)\delta\tilde{b}_i\right| = \sum_{i=1}^{d}\operatorname{sgn}(a_i)\left(a_i+(1+\eta)\delta\tilde{b}_i\right)\\ &= \|a\|_1+(1+\eta)\delta\sum_{i=1}^{d}\operatorname{sgn}(a_i)\tilde{b}_i. \end{aligned} \tag{4}$$

In the case $a \in \{-\vec{1},\vec{1}\}$, the above equals d. In the remaining case $a \in Q\backslash\{-\vec{1},\vec{1}\}$, the choice of $\tilde{b}$ implies that $\sum_{i=1}^{d}\operatorname{sgn}(a_i)\tilde{b}_i \leqslant d-2$, and thus

$$\left\|a+(1+\eta)\delta\tilde{b}\right\|_1 \leqslant \|a\|_1+(1+\eta)\delta(d-2) \leqslant (1-\delta)d+(1+\eta)\delta(d-2) < d.$$

We have thus confirmed that $-(1+\eta)\delta\tilde{b}$ is an optimal center of Q, but not for P. This contradicts our assumption that Q is a weak 0-coreset, and completes the proof of the first item.

To prove the second item, let $P \subset \mathbb{R}^d$ be as before but now for $\delta = \frac{1}{3}$, and consider a weak ϵ-coreset Q of P, for some $0 < \epsilon < \delta = \frac{1}{3}$. Notice that Q must contain both $-\vec{1}$ and $\vec{1}$, as otherwise, if $-\vec{1},\vec{1} \notin Q$, consider the origin $\vec{0} \in \mathbb{R}^d$ as a center. If Q contains just one of $\{-\vec{1},\vec{1}\}$ say $\vec{1}$, consider $\delta\vec{1}$ as a center. In both cases, we obtain that $\operatorname{opt}(Q) \leqslant (1-\delta)d < (1-\epsilon)d < \frac{1}{1+\epsilon}\operatorname{opt}(P)$, in contradiction to the weak ϵ-coreset.

Denote by $\psi(Q)$ the minimum Hamming distance between any point in B and its furthest point in Q, that is, $\psi(Q) = \min_{b\in B}\max_{a\in Q}\operatorname{Hamm}(a,b)$. The next proposition follows by a standard counting-bound.

▶ **Proposition 2.3.** *If $|Q| \leqslant \frac{1}{\sqrt{2d}}2^{0.18d}$, then $\psi(Q) < \frac{3}{4}d$.*

Proof. Denote by $C_{\frac{3d}{4}}(a)$ the set of all points in $\{\pm1\}^d$ whose Hamming distance from $a \in \{\pm1\}^d$ is at least $\frac{3d}{4}$. Then

$$\left|C_{\frac{3d}{4}}(a)\right| \leqslant \sum_{0\leqslant j\leqslant\frac{d}{4}}\binom{d}{j} \leqslant 2^{dH(\frac{1}{4})},$$

where the last inequality is the known entropy bound for binomial coefficients and $H(p) = -p\log p-(1-p)\log(1-p)$, hence $H(\frac{1}{4}) \approx 0.811$. Taking a union over all $a \in Q$,

$$\left|\bigcup_{a\in Q}C_{\frac{3d}{4}}(a)\right| \leqslant \sum_{a\in Q}\left|C_{\frac{3d}{4}}(a)\right| \leqslant 2^{dH(\frac{1}{4})}|Q| < \binom{d}{d/2} = |B|,$$

where in the last inequality we used the known bound $\frac{1}{\sqrt{2d}}2^d \leqslant \binom{d}{d/2}$. Hence, there exists some point in B for which the maximum Hamming distance to points in Q is smaller than $\frac{3d}{4}$. ◀

Assume towards contradiction that $|Q| \leqslant \frac{1}{\sqrt{2d}}2^{0.18d}$. Then by Proposition 2.3, there is a point $\tilde{b} \in B$ for which the maximum Hamming distance to points in Q is smaller than $\frac{3d}{4}$. We next show that the point $2\delta\tilde{b}$ is an optimal center for Q. Similarly to Equation (4), since $2\delta \leqslant 1-\delta$, for every $a \in Q$ we can write:

$$\left\|a-2\delta\tilde{b}\right\|_1 = \sum_{i=1}^{d}\left|a_i-2\delta\tilde{b}_i\right| = \|a\|_1-2\delta\sum_{i=1}^{d}\operatorname{sgn}(a_i)\tilde{b}_i,$$

where:

$$\sum_{i=1}^{d} \operatorname{sgn}(a_i)\tilde{b}_i = \sum_{i:\operatorname{sgn}(a_i)=\operatorname{sgn}(\tilde{b}_i)} 1 - \sum_{i:\operatorname{sgn}(a_i)\neq\operatorname{sgn}(\tilde{b}_i)} 1 = d - 2\operatorname{Hamm}(a,\tilde{b}),$$

and thus:

$$\left\|a - 2\delta\tilde{b}\right\|_1 = \|a\|_1 - 2\delta d + 4\delta \operatorname{Hamm}(a,\tilde{b}).$$

In the case $a \in \{-\vec{1},\vec{1}\}$, since $\tilde{b} \in B$ then we have $\left\|a - 2\delta\tilde{b}\right\|_1 = \|a\|_1 = d$. Otherwise $a \in (1-\delta)B$, and $\left\|a - 2\delta\tilde{b}\right\|_1 < \|a\|_1 - 2\delta d + 4\delta \cdot \frac{3}{4}d \leqslant (1-\delta)d + \delta d = d$. In both cases, the distance from $a \in Q$ to $2\delta\tilde{b}$ is at most $d = \operatorname{opt}(Q)$, with equality for some $a \in Q$. Thus, $2\delta\tilde{b}$ is an optimal center for Q. However, $\operatorname{cost}(2\delta\tilde{b}, P) \geqslant \left\|2\delta\tilde{b} + (1-\delta)\tilde{b}\right\|_1 = (1+\delta)d > (1+\epsilon)\operatorname{opt}(P)$. This contradicts that Q is a weak ϵ-coreset of P, and completes the proof of Theorem 2.2. ◀

3 Weak ϵ-coresets in ℓ_1 for inputs with a unique solution

In this section we consider the special case, where the input P has a unique optimal center. We show that even in this special case, the size of the coreset depends on the dimension and provide a lower bound of $\Omega(\log d)$ on the size of every ϵ-coreset for any fixed $\epsilon < 1$. This restriction over ϵ is necessary, as for larger values of ϵ, a naive coreset construction (based on Gonzalez algorithm for k-center [23]) with only 2 points achieves a 3-approximation.[4] We also prove that there always exists a weak 0-coreset of size at most $2d$. This is in stark contrast to Section 2, where we show that in the general case of multiple solutions, 0-coresets might require size exponential in d. Lastly, we present a set such that every weak 0-coreset of this set is of size $2d$, hence the upper bound of $2d$ is tight for the case $\epsilon = 0$.

3.1 Lower bound of $\Omega(\log d)$ for weak ϵ-coresets

▶ **Theorem 3.1.** *Fix $\epsilon < 1$ and consider the set of points $P = \{\pm 1\}^d \subset \mathbb{R}^d$. Then P has a unique optimal center and every weak ϵ-coreset of P must have $\Omega(\log d)$ points.*

Proof. It is easy to note that the origin $\vec{0} \in \mathbb{R}^d$ is a unique optimal center with value d as the input contains a point from every face of the ℓ_1 ball centered at $\vec{0}$. Consider a weak ϵ-coreset $Q \subset P$ of size $|Q| \leqslant \frac{1}{2}\log d$. Every point $p \in Q$ induces a partition Π_p of $[d]$ into at most two parts, by splitting the coordinates of p into the positive and negative ones. Let Π be the common refinement of all these partitions (i.e. placing $i, j \in [d]$ in the same part of Π if and only if for every $p \in P$ they are in the same part of Π_p). Since each Π_p has at most two parts, Π has at most $2^{|Q|} \leqslant \sqrt{d}$ parts. Observe that for every point $p \in Q$, coordinates in the same part of Π have the same sign.

Consider an optimal center c^* for Q. The idea is to modify it iteratively, without increasing its cost for Q, so as to have more coordinates take values in $\{-1, 1\}$. Each iteration takes two indices i, j in the same part of Π, whose coordinates in c^* are *both* not in $\{-1, 1\}$, say, $-1 < c^*_i \leqslant c^*_j < 1$. The iterations stop when no such indices exist. Now add δ to c^*_i and $-\delta$ to c^*_j, where $\delta > 0$ is as large as possible while keeping both new values in the range

[4] Pick arbitrary $p \in P$, then a point $p' \in P$ that is furthest from p, and take $Q = \{p, p'\}$ as the coreset. Note that $\operatorname{opt}(Q) = \frac{1}{2}\left\|p - p'\right\|_1$, and $\frac{1}{2}(p + p')$ is a possible optimal center. For every optimal center c^* for Q, we can write $\operatorname{cost}(c^*, P) \leqslant \left\|c^* - p\right\|_1 + \operatorname{cost}(p, P) \leqslant 1.5\left\|p - p'\right\|_1 = 3\operatorname{opt}(Q) \leqslant 3\operatorname{opt}(P)$.

$[-1,1]$, that is, $\delta = \min\{1 - c_j^*, c_i^* - 1\}$. It is easy to see that after this modification, at least one of c_i^* and c_j^* will be in $\{\pm 1\}$, and at the same time the distance to every $p \in Q$ does not change. The iteration stop at a center $\bar{c}^*$, where in each part of Π at most one coordinate is not in $\{\pm 1\}$, and in total at most $|\Pi| \leqslant 2^{|Q|} \leqslant \sqrt{d}$ coordinates are not in $\{\pm 1\}$.

Finally, we show that $\bar{c}^*$, which is an optimal center for Q, has a too large cost for P. Consider the point in P that is closest to $-\bar{c}^*$, i.e., where each coordinate of $-\bar{c}^*$ is rounded to the nearest value among $\{\pm 1\}$. This point disagrees with $-\bar{c}^*$ (i.e., rounding was "needed") on at most $|\Pi|$ coordinates, hence its distance from $\bar{c}^*$ is at least $2d - |\Pi|$, hence $\mathrm{cost}(\bar{c}^*, P) \geqslant 2d - |\Pi| \geqslant 2d - \sqrt{d} > (1+\epsilon)\,\mathrm{opt}(P)$. ◀

We remark that the proof works also for the discrete space $\{-1, 0, +1\}^d$ (endowed with ℓ_1 norm), which has more limited choices for the center point, e.g., for c^* above. Moreover, note that the distance of the modified optimal center $\bar{c}^*$ from $\vec{0}$ is at least $d - 2^{|Q|}$.

3.2 Upper bound of $2d$ for weak 0-coresets

We will show that if P has a unique solution, then P has a weak 0-coreset of size at most $2d$.

Two standard notations from the domain of convex geometry are the convex-hull and the interior. The convex hull of P, denoted $\mathrm{conv}(P)$, consists of all convex combinations of points in P, that is every point in $\mathrm{conv}(P)$ can be expressed as a weighted sum of points from P with non-negative weights summing to 1. The interior of P, denoted $\mathrm{int}(P)$, refers to the set of points that lie strictly inside the convex hull. In other words, it consists of points that can be surrounded by a small ball completely contained within $\mathrm{conv}(P)$.

An important tool we use is Steinitz's theorem [24].

▶ **Theorem 3.2** (Steinitz's theorem). *Let $\vec{0} \in \mathrm{int}(\mathrm{conv}(S))$ for some $S \subseteq \mathbb{R}^d$. Then there exists $R \subseteq S$ of size at most $2d$, such that $\vec{0} \in \mathrm{int}(\mathrm{conv}(Q))$.*

Next we introduce the notion of complete set, we will then provide an alternative formulation of Steinitz's theorem that will be useful in our application.

▶ **Definition 3.3.** *A set $S \subseteq \mathbb{R}^d$ is complete if for every $v \in \mathbb{R}^d \backslash \{\vec{0}\}$ there exists $s \in S$ with $v \cdot s > 0$.*

▶ **Lemma 3.4.** *$\vec{0} \in \mathrm{int}(\mathrm{conv}(S))$ if and only if S is complete.*

Proof. $(\Rightarrow)$: Write $\vec{0}$ as a convex combination, $\vec{0} = \sum_i \lambda_i s_i$ where $\lambda_i \geqslant 0$ and $\sum_i \lambda_i = 1$. Then,

$$0 = v \cdot \vec{0} = v \cdot (\sum_i \lambda_i s_i) = \sum_i \lambda_i (v \cdot s_i)$$

If for all s_i, $v \cdot s_i \leqslant 0$, then it must hold that for all s_i, $v \cdot s_i = 0$, but then $\vec{0}$ is not in the interior of $\mathrm{conv}(S)$.

$(\Leftarrow)$: If $\vec{0} \notin \mathrm{int}(\mathrm{conv}(S))$ then there exists a hyperplane $\mathcal{H}$ containing $\vec{0}$ such that $S \subseteq \mathcal{H}^+$. Since $\vec{0}$ is a point in the hyperplane we can write the hyperplane equation as $vx = 0$ for some $v \in \mathbb{R}^d \backslash \{\vec{0}\}$. It follows that $v \cdot s \leqslant 0$ for every $s \in S$, hence S is not complete. ◀

This leads to the following alternative statement of Steinitz's theorem.

▶ **Corollary 3.5.** *If $S \subseteq \mathbb{R}^d$ is complete, then there exists $R \subseteq S$ of size at most $2d$, such that R is complete.*

W.l.o.g., assume that $\vec{0}$ is the unique solution of P and let $\text{opt}(P) = \text{cost}(\vec{0}, P) = r$. Since $\vec{0}$ is unique, we can further assume that $\|\vec{0} - p\|_1 = r$ for every $p \in P$; otherwise, any points that do not satisfy this can be removed without affecting the optimal solution or its cost. Given a pair $a = (\sigma, p) \in \{-1,1\}^d \times P$, we denote by $\mathcal{H}_a$ the hyperplane $\sigma \cdot (x - p) = r$ and by $\mathcal{H}_a^+$ the halfspace $\sigma \cdot (x - p) \leqslant r$. In general, $x \in \mathbb{R}^d$ is a feasible 1-center of P if for every $p \in P$, $\|x - p\|_1 \leqslant r$. Assuming r is known, we can write the LP formulation of the 1-center problem with d variables $x = (x_1, ..., x_d)$ and the $2^d|P|$ constraints $\sigma \cdot (x - p) \leqslant r$ for every $\sigma \in \{-1,1\}^d$ and $p \in P$. That is, every constraint defines a hyperplane $\mathcal{H}_{(\sigma,p)}$. We further remove from the set of constraints $\{-1,1\}^d \times P$, pairs (σ, p) for which $\vec{0} \notin \mathcal{H}_{(\sigma,p)}$ and denote the updated set of constraint by $A = S \times P$. Note that no points from P are removed in this process, as each has distance r from $\vec{0}$. Since $\vec{0}$ is the unique optimal solution of P, this also does not change the set of feasible solutions, and $\vec{0}$ is the unique intersection point of all hyperplanes in A.

▶ **Proposition 3.6.** *Let $A = S \times P$ be non-empty set such that $\vec{0} \in \mathcal{H}_a$ for every $a \in A$. Then, $\bigcap_{a \in A} \mathcal{H}_a^+ = \{\vec{0}\}$ if and only if S is complete.*

Proof. $(\Rightarrow)$: Let $v \in \mathbb{R}^d \backslash \{\vec{0}\}$, then there exists $a \in A$ such that $v \notin \mathcal{H}_a^+$. That is, $\sigma \cdot (v - p) > r$ and since $\vec{0} \in \mathcal{H}_a$, $\sigma \cdot (\vec{0} - p) = r$, and it follows $\sigma \cdot v > 0$.

$(\Leftarrow)$: A is non-empty so clearly $\vec{0} \in \bigcap_{a \in A} \mathcal{H}_a^+$. Let $v \neq \vec{0}$, then there exists $(\sigma, p) \in A$ such that $\sigma \cdot v > 0$. Since $\vec{0} \in \mathcal{H}_a$ we have $\sigma(\vec{0} - p) = r$, it follows that $\sigma(v - p) > r$, that is, $v \notin \mathcal{H}_a^+$. ◀

We are now ready to prove our main Theorem:

▶ **Theorem 3.7.** *Every instance $P \subset \mathbb{R}^d$ with unique optimum admits a weak 0-coreset of size $2d$.*

Proof. By Proposition 3.6 the set S is complete and by Corollary 3.5 there exists $R \subseteq S$ such that R is complete and of size at most $2d$. For each $\sigma \in R$ we select a single point $p \in P$ such that the pair (σ, p) is a constraint in A. That is we obtain a set of at most $2d$ different points from P we denote by Q and denote by A' this refined set of constraints. Since R is complete, using the second direction of Proposition 3.6 we have that $\bigcap_{a \in A'} \mathcal{H}_a^+ = \{\vec{0}\}$, meaning that Q is a weak 0-coreset of size at most $2d$. ◀

We conclude this section by providing an example of a set P of size $2d$, where P is the sole weak 0-coreset of itself, demonstrating that the upper bound of $2d$ cannot be further improved. The set P is composed of the row vectors of the Hadamard matrix and their negations. The proof is provided in Appendix A.1.

4 Value-preserving coresets

In this section we provide an algorithm for constructing a dimension-independent value-preserving ϵ-coreset for 1-center. In addition, we provide a complete characterization for the special case $\epsilon = 0$, by showing a 0-coreset of size $d + 1$, and moreover that this bound is tight, see in Section 4.2.

▶ **Theorem 4.1.** *For every instance $P \subset \mathbb{R}^d$ and $0 < \epsilon < 1$, there exists a subset $S \subseteq P$ of $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$ points such that*

$$\text{opt}(S) \geqslant (1 - \varepsilon)\, \text{opt}(P).$$

Our proof of Theorem 4.1 is based on the following LP formulation of 1-center and its dual. Unlike many primal-dual algorithms, our algorithm needs to explicitly solve and use this optimal dual solution to guide the coreset construction.

LP formulation of 1-center. Observe that 1-center on input $P \subset \mathbb{R}^d$ is equivalent to the following mathematical program with variables r and $x = (x_1, \dots, x_d)$.

$$\begin{aligned} &\text{minimize} && r \\ &\text{subject to} && \sum_{i \in [d]} |p_i - x_i| \leqslant r && \forall p \in P. \end{aligned}$$

This is formally not an LP due to the absolute values in the constraints, but each such constraint can obviously be expanded into 2^d linear constraints, which leads to the following equivalent LP formulation.

$$\begin{aligned} &\text{minimize} && r \\ &\text{subject to} && \sum_{i \in [d]} \sigma_i (p_i - x_i) \leqslant r && \forall \sigma \in \{\pm 1\}^d, p \in P \\ &&& r \in \mathbb{R}, x \in \mathbb{R}^d. \end{aligned}$$

We derive the dual of this LP, with $2^d n$ variables $\{u_{\sigma,p}\}_{\sigma,p}$, as follows. One can think of each pair (σ, p) as generating a "new" point $(\sigma_1 p_i, \dots, \sigma_d p_d) \in \mathbb{R}^d$, that is obtained from $p \in P$ by flipping signs.

$$\begin{aligned} &\text{maximize} && \sum_{\sigma \in \{\pm 1\}^d} \sum_{p \in P} \sum_{i \in [d]} \sigma_i p_i u_{\sigma,p} && && \text{(DLP)} \\ &\text{subject to} && \sum_{\sigma \in \{\pm 1\}^d} \sum_{p \in P} \sigma_i u_{\sigma,p} = 0 && \forall i \in [d] \\ &&& \sum_{\sigma \in \{\pm 1\}^d} \sum_{p \in P} u_{\sigma,p} = 1 \\ &&& u_{\sigma,p} \geqslant 0 && \forall \sigma \in \{\pm 1\}^d, p \in P. \end{aligned}$$

Coreset Algorithm. Our coreset construction, presented in Algorithm 1, builds the coreset by sampling, where the probabilities come from an optimal solution to (DLP). Technically, our sampling step is similar to randomized rounding that is often used in approximation algorithms, however we use it here to find a coreset rather than an approximate solution. The analysis of this algorithm appears in Section 4.1. We slightly abuse terminology and refer to $\widehat{W}$ as a multiset, but formally it is always a sequence of m pairs, namely, $\widehat{W} := ((\widehat{\sigma}^1, p^1), \dots, (\widehat{\sigma}^m, p^m))$, hence, summing over all $(\widehat{\sigma}, p) \in \widehat{W}$ actually means summation over $j \in [m]$.

Algorithm 1 Value-preserving ϵ-coreset.

```
shift $P$ such that $\sum_{p \in P} p = \vec{0}$ and solve (DLP) on $P$ to obtain an optimal solution $u^*$
let $\widehat{W}$ be a multiset of $m := O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$ i.i.d. samples from the distribution $u^*$
    // $m$ is an even number, $u^*$ viewed as distribution over $\{\pm 1\}^d \times P$
return $S := \{p : (\sigma, p) \in \widehat{W}\}$
```

▶ Remark 4.2. Algorithm 1 can be implemented in poly(nd)-time. The most expensive step is clearly to solve the dual LP. Since it has only $d+1$ constraints, one can solve it while using only poly(d) non-zero variables. Alternatively, one can first solve the primal LP using an ellipsoid algorithm, to find a subset of poly(n) constraints that has the same optimal value, and then solve the dual of this smaller primal LP. Moreover, inspecting the analysis, particularly (6), one can verify that it suffices to $(1+\varepsilon)$-approximate the dual LP.

4.1 Proof of Theorem 4.1: Analysis of Algorithm 1

We shall prove that the algorithm's output S satisfies $\Pr[\mathrm{opt}(S) \geqslant (1-\epsilon)\,\mathrm{opt}(P)] \geqslant 0.8$. The plan is to build a dual solution for $S \subset \mathbb{R}^d$, and show that with high probability, this dual solution is feasible for (DLP) and its objective is at least $(1-\epsilon)$ times the optimal value (of the same dual LP) for P. This would imply that $\mathrm{DLP}(S) \geqslant (1-\epsilon)\cdot\mathrm{DLP}(P)$, where $\mathrm{DLP}(P)$ denotes the optimal value of (DLP) on input P, and by strong LP duality this is equivalent to $\mathrm{opt}(S) \geqslant (1-\epsilon)\cdot\mathrm{opt}(P)$, proving the theorem.

A lower bound on opt($\boldsymbol{P}$). Recall that Algorithm 1 shifts P so that $\sum_{p\in P} p = \vec{0}$. This does not change the optimal value, hence our analysis simply assumes that the input P already satisfies $\sum_{p\in P} p = \vec{0}$, avoiding cumbersome notations for before and after the shift. This property of P yields the following immediate lower bound on $\mathrm{opt}(P)$.

▶ **Lemma 4.3.** *If $P \subset \mathbb{R}^d$ satisfies $\sum_{p\in P} p = \vec{0}$, then* $\mathrm{opt}(P) \geqslant \max\{\|p\|_1/2 : \ p \in P\}$.

Proof. For every point $p \in P$,

$$\|p\|_1 = \Big\|\tfrac{1}{|P|}\sum_{q\in P}(p-q)\Big\|_1 \leqslant \tfrac{1}{|P|}\sum_{q\in P}\|p-q\|_1 \leqslant 2\,\mathrm{opt}(P).$$ ◀

Dual solution induced by $\widehat{\boldsymbol{W}}$. We now wish to use $\widehat{W}$ from Algorithm 1 to construct a solution for the dual LP. Informally, the idea is to construct a solution $u^{\widehat{W}}$ (can also think of it as a vector with $2^d n$ coordinates), that is the average of m sparse solutions, namely, each has a single non-zero variable, whose value is 1 and it is drawn at random from the probability distribution defined by u^* (think of u^* as a vector with $2^d n$ coordinates that sum to 1), using precisely the random draws made in Algorithm 1. Next, we formalize this construction in a slightly more general form.

Given a multiset $Z \subseteq \{\pm1\}^d \times P$ of size m, such as $\widehat{W}$, we define the *dual solution induced* by Z, denoted u^Z, as follows. Take each pair $(\sigma,p)\in Z$ and build a dual solution u, where the variable corresponding to this pair is set as $u_{\sigma,p}=1$, and the other $2^d n - 1$ variables are set to zero, and then we average all these m solutions. Formally, u^Z is given by

$$u^Z_{\sigma',p'} := \tfrac{1}{m}\sum_{(\sigma,p)\in Z} \mathbb{1}_{\{\sigma'=\sigma,p'=p\}}.$$

The following fact rewrites certain linear forms over u^Z, that appear in (DLP), in terms of pairs $(\sigma,p)\in Z$.

▶ **Fact 4.4.** *Let u^Z be induced by some $Z \subseteq \{\pm1\}^d \times P$ as above. Then for every $i\in[d]$,*

$$\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in P}\sigma_i u^Z_{\sigma,p} = \tfrac{1}{m}\sum_{(\sigma,p)\in Z}\sigma_i,$$

$$\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in P}\sigma_i p_i u^Z_{\sigma,p} = \tfrac{1}{m}\sum_{(\sigma,p)\in Z}\sigma_i p_i.$$

Initial dual solution. Before building a *feasible* dual solution for (DLP) on S, we first consider an initial dual solution $\widehat{u} := u^{\widehat{W}}$, which is just the solution induced by $\widehat{W}$ from Algorithm 1. We shall see below that $\widehat{u}$ is feasible in an expected sense, and that turning the expectation into a high-probability guarantee introduces an additive error in the constraints (and in the objective). Nonetheless, this is still useful as we will later "fix" this solution into one that does satisfy the constraints (without additive error).

Let us examine this initial random solution $\widehat{u}$. It always satisfies the second constraint of (DLP) by construction (recall that $\widehat{u}$ is the average of m sparse solutions), i.e.,

$$\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in S}\widehat{u}_{\sigma,p} = m\cdot\tfrac{1}{m} = 1.$$

By linearity of expectation, $\widehat{u}$ satisfies the first constraint in expectation (the middle term uses Theorem 4.4 to provide an alternate formulation),

$$\forall i\in[d],\quad \mathbb{E}\Big[\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in P}\sigma_i\widehat{u}_{\sigma,p}\Big] = \mathbb{E}\Big[\tfrac{1}{m}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i\Big] = \tfrac{1}{m}\cdot m\cdot\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in P}\sigma_i u^*_{\sigma,p} = 0. \tag{5}$$

And again by linearity of expectation, the expected objective of $\widehat{u}$ is the same as the optimal solution u^*, because:

$$\mathbb{E}\Big[\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in P}\sum_{i\in[d]}\sigma_i p_i\widehat{u}_{\sigma,p}\Big] = \mathbb{E}\Big[\frac{1}{m}\sum_{i\in[d]}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i p_i\Big] = \mathrm{opt}(P). \tag{6}$$

Almost feasibility. We bound the deviation of $\widehat{u}$ from satisfying the first constraint, as follows. Consider $i\in[d]$ and let $\mathcal{F}_i$ be the event that $|\sum_{\sigma\in\{\pm1\}^d}\sum_{p\in S}\sigma_i\widehat{u}_{\sigma,p}| > \epsilon$, which by Theorem 4.4 can be also written as $|\frac{1}{m}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i| > \epsilon$. Informally, we want to bound the probability of this event, because when it does not occur, the constraint is "almost satisfied". By applying Hoeffding's inequality, where we use (5) for the expectation and the fact that $|\sigma_i|\leqslant 1$, we get (by our choice of m),

$$\Pr[\mathcal{F}_i] = \Pr\Big[\Big|\tfrac{1}{m}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i\Big| > \epsilon\Big] \leqslant 2e^{-\Omega(\epsilon^2 m)} \leqslant \epsilon. \tag{7}$$

A feasible solution u. We next use $\widehat{W}$ to construct a new multiset W whose induced dual u^W is feasible, and in particular satisfies the first constraint. We will then take our final dual solution to be $u := u^W$, and it will remain to bound its objective value. At a high level, W is obtained by "flipping" some of the signs appearing in $\widehat{W}$. More precisely, given $\widehat{W} = ((\widehat{\sigma}^1,p^1),\ldots,(\widehat{\sigma}^m,p^m))$, we shall define new sign vectors $\sigma^1,\ldots,\sigma^m$ but keep the exact same points $p^1,\ldots,p^m\in P$, and construct $W = ((\sigma^1,p^1),\ldots,(\sigma^m,p^m))$. It will be convenient to arrange the sign vectors $\widehat{\sigma}^1,\ldots,\widehat{\sigma}^m$ as the rows of a matrix $\widehat{M}\in\{\pm1\}^{m\times d}$, and use this $\widehat{M}$ to define a new matrix M, whose rows define the new sign vectors $\sigma^1,\ldots,\sigma^m$.

We construct M (from $\widehat{M}$) using the following procedure, which operates separately on each column $i\in[d]$. First copy column i of $\widehat{M}$ to be also column i of M, and let n_i^+ and n_i^- be the number of positive and negative entries in this column, respectively. If $n_i^+ = n_i^-$, leave this column of M as is, noticing that it sums to zero. Otherwise, flip $\frac{|n_i^+-n_i^-|}{2}$ signs chosen carefully from among the majority sign in column i, and this will ensure that column i of M sums to zero. (This is possible because m is even.) The careful choice (which signs to flip) will favor row indices $j\in[m]$ with small value $|p_i^j|$. Formally, let $R_i\subseteq[m]$ be the set of rows j with the majority sign in column i of $\widehat{M}$, then sort the indices $j \in R_i$ by their value $|p_i^j|$, pick the $|n_i^+ - n_i^-|/2$ indices with smallest value (breaking ties arbitrarily), and flip their entries. As explained above, the rows of the resulting matrix M define sign vectors $\sigma^1, \dots, \sigma^m$, which in turn define $W = ((\sigma^1, p^1), \dots, (\sigma^m, p^m))$, and our final dual solution is $u = u^W$.

▶ **Lemma 4.5.** *The solution $u = u^W$ is feasible for* (DLP) *on S.*

Proof. The first constraint is satisfied because the signs in every column of M sum to zero. The second constraint is satisfied because flipping a sign σ_i^j "moves" some amount (say $1/m$) from variable $u_{\sigma,p}$ to $u_{\sigma',p}$, where σ, σ' differ in coordinate j, but the total remains the same. Finally, it uses only variables $u_{\sigma,p}$ for $p \in S$, i.e., all other variables are 0, and thus it is feasible not only for (DLP) on P but also on S. ◀

The next lemma bounds the decrease in objective when constructing w from $\widehat{W}$, i.e., comparing that of u with that of $\hat{u}$. Its proof is based on an averaging argument.

▶ **Lemma 4.6.** *For all $i \in [d]$,*

$$\Big| \sum_{(\sigma,p)\in\widehat{W}} \sigma_i p_i - \sum_{(\sigma,p)\in W} \sigma_i p_i \Big| \leqslant \frac{2|n_i^+ - n_i^-|}{m} \sum_{(\sigma,p)\in\widehat{W}} |p_i|.$$

Proof. We consider only the case $n_i^+ \geqslant n_i^-$, as the other case is symmetric. Since $n_i^+ + n_i^- = m$, we have $n_i^+ \geqslant m/2$. Recall that the construction of M flips in column i exactly $\frac{n_i^+ - n_i^-}{2}$ signs, picking the indices $j \in R_i$ with smallest value $|p_i^j|$. Every such flip changes the objective by $2|p_i^j|$, and the number of choices $|R_i| = n_i^+ \geqslant m/2$, hence by an averaging argument,

$$\Big| \sum_{(\sigma,p)\in\widehat{W}} \sigma_i p_i - \sum_{(\sigma,p)\in W} \sigma_i p_i \Big| \leqslant \frac{(n_i^+ - n_i^-)/2}{|R_i|} \sum_{j\in R_i} 2|p_i^j| \leqslant \frac{n_i^+ - n_i^-}{m/2} \sum_{j\in[m]} |p_i^j|.$$ ◀

The objective value of u. Using Theorem 4.4, we can write the objective value of u as

$$L := \sum_{i\in[d]} \sum_{\sigma\in\{\pm1\}^d} \sum_{p\in S} \sigma_i p_i u_{\sigma,p} = \tfrac{1}{m} \sum_{i\in[d]} \sum_{(\sigma,p)\in W} \sigma_i p_i. \tag{8}$$

and that of $\hat{u}$ as

$$\widehat{L} := \sum_{i\in[d]} \sum_{\sigma\in\{\pm1\}^d} \sum_{p\in S} \sigma_i p_i \hat{u}_{\sigma,p} = \tfrac{1}{m} \sum_{i\in[d]} \sum_{(\sigma,p)\in\widehat{W}} \sigma_i p_i, \tag{9}$$

Our plan is to bound L (with high probability) by relating it to $\widehat{L}$, and then bound the latter using Chebyshev's inequality. These steps rely on the next two lemmas, and before proving them, we show how they imply Theorem 4.1.

▶ **Lemma 4.7.** $\mathbb{E}[|L - \widehat{L}|] \leqslant O(\epsilon)\,\mathrm{opt}(P)$.

▶ **Lemma 4.8.** $\mathrm{Var}(\widehat{L}) \leqslant \frac{4}{m}(\mathrm{opt}(P))^2$.

Proof of Theorem 4.1. We already established in Theorem 4.5 that u is a feasible solution for (DLP) on S, hence $\mathrm{opt}(S) \geqslant L$. It thus suffices to show that $\Pr[L \geqslant (1 - O(\epsilon))\,\mathrm{opt}(P)] \geqslant 0.8$. By Theorem 4.7 and Markov's inequality, $\Pr[|L - \widehat{L}| \leqslant O(\epsilon)\,\mathrm{opt}(P)] \geqslant 0.9$. Next, we apply Chebyshev's inequality to $\widehat{L}$, using the variance bound Theorem 4.8 and our choice of m, to obtain $\Pr[|\widehat{L} - \mathbb{E}[\widehat{L}]| \geqslant \epsilon\,\mathrm{opt}(P)] \leqslant \frac{4/m}{\epsilon^2} \leqslant 0.1$. Recall also from (6) that $\mathbb{E}[\widehat{L}] = \mathrm{opt}(P)$. It follows by a union bound that with probability at least 0.8,

$$L \geqslant \widehat{L} - O(\epsilon)\,\mathrm{opt}(P) \geqslant (1 - O(\epsilon))\,\mathrm{opt}(P).$$ ◀

Proof of Theorem 4.7. Observe that $\mathcal{F}_i$ is the event that $|n_i^+ - n_i^-| > \epsilon m$, in which case we can still bound $|n_i^+ - n_i^-| \leqslant m$ (which holds always). Thus, $|n_i^+ - n_i^-| \leqslant \epsilon m + \mathbb{1}_{\mathcal{F}_i} \cdot m$. Now applying Theorem 4.6 for every $i \in [d]$ and taking a sum, we get (after dividing by m)

$$\begin{aligned}\tfrac{1}{m}\Big|\sum_{i\in[d]}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i p_i - \sum_{i\in[d]}\sum_{(\sigma,p)\in W}\sigma_i p_i\Big| &\leqslant \tfrac{1}{m}\sum_{i\in[d]}\frac{2|n_i^+ - n_i^-|}{m}\sum_{(\sigma,p)\in\widehat{W}}|p_i| \\ &\leqslant \tfrac{2}{m}\sum_{i\in[d]}\Big[(\epsilon + \mathbb{1}_{\mathcal{F}_i})\sum_{(\sigma,p)\in\widehat{W}}|p_i|\Big]. \qquad (10)\end{aligned}$$

To bound (10) further, we expand its final expression into two parts. We then bound one part in (11) using Theorem 4.3, and the other part in Theorem 4.9.

$$\mathbb{E}\Big[\tfrac{1}{m}\sum_{i\in[d]}\epsilon\sum_{(\sigma,p)\in\widehat{W}}|p_i|\Big] = \epsilon\cdot\mathbb{E}\Big[\tfrac{1}{m}\sum_{(\sigma,p)\in\widehat{W}}\|p_i\|_1\Big] \leqslant 2\epsilon\cdot\mathrm{opt}(P). \qquad (11)$$

▷ Claim 4.9. $\mathbb{E}\Big[\tfrac{1}{m}\sum_{i\in[d]}\mathbb{1}_{\mathcal{F}_i}\sum_{(\sigma,p)\in\widehat{W}}|p_i|\Big] \leqslant 2\epsilon\cdot\mathrm{opt}(P)$.

Proof. The main obstacle here is that we have a product of two random variables, $\mathbb{1}_{\mathcal{F}_i}$ and $|p_i|$, that are not independent. Intuitively, their dependence should be relatively small, and our actual proof bounds their product with another product, of two independent random variables. Recall that $\widehat{W} = \{(\widehat{\sigma}^1, p^1), \dots, (\widehat{\sigma}^m, p^m)\}$ consists of m i.i.d. samples, hence we can focus on analyzing (say) the last one, formally

$$\mathbb{E}\Big[\tfrac{1}{m}\sum_{i\in[d]}\mathbb{1}_{\mathcal{F}_i}\cdot\sum_{(\sigma,p)\in\widehat{W}}|p_i|\Big] = \mathbb{E}\Big[\sum_{i\in[d]}\mathbb{1}_{\mathcal{F}_i}\cdot|p_i^m|\Big].$$

Consider $i \in [d]$, and define $\mathcal{F}_i'$ to be the event that $|\frac{1}{m}\sum_{j=1}^{m-1}\widehat{\sigma}_i^j| > \epsilon - \frac{1}{m}$. Observe that $\mathbb{1}_{\mathcal{F}_i} \leqslant \mathbb{1}_{\mathcal{F}_i'}$ because if $\mathcal{F}_i$ occurs then also $\mathcal{F}_i'$ must occur. The crux is that $\mathcal{F}_i'$ is independent of $(\widehat{\sigma}^m, p^m)$, and thus $\mathbb{E}[\mathbb{1}_{\mathcal{F}_i'}\cdot|p_i^m|] = \mathbb{E}[\mathbb{1}_{\mathcal{F}_i'}]\cdot\mathbb{E}[|p_i^m|]$. We can still bound $\mathbb{E}[\mathbb{1}_{\mathcal{F}_i'}] = \Pr[\mathcal{F}_i'] \leqslant \epsilon$ by the argument we had for $\mathcal{F}_i$ in (7), except that we use $m-1$ instead of m. Altogether,

$$\mathbb{E}\Big[\sum_{i\in[d]}\mathbb{1}_{\mathcal{F}_i}\cdot|p_i^m|\Big] \leqslant \mathbb{E}\Big[\sum_{i\in[d]}\mathbb{1}_{\mathcal{F}_i'}\cdot|p_i^m|\Big] = \sum_{i\in[d]}\mathbb{E}[\mathbb{1}_{\mathcal{F}_i'}]\cdot\mathbb{E}[|p_i^m|] \leqslant \epsilon\cdot\mathbb{E}[\|p^m\|_1],$$

and the claim follows by using Theorem 4.3 to bound $\|p^m\|_1 \leqslant 2\,\mathrm{opt}(P)$. ◁

We now proceed with the proof of Theorem 4.7. Plugging the bounds from (11) and Theorem 4.9 into (10), we obtain

$$\mathbb{E}\Big[\tfrac{1}{m}\Big|\sum_{i\in[d]}\sum_{(\sigma,p)\in W}\sigma_i p_i - \sum_{i\in[d]}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i p_i\Big|\Big] \leqslant O(\epsilon)\,\mathrm{opt}(P).$$

which completes the proof of Theorem 4.7. ◀

Proof of Theorem 4.8. For $\sigma \in \{\pm 1\}^d$ and $p \in P$, define $X_{\sigma,p} := \sum_{i\in[d]}\sigma_i p_i$. Then we can write

$$\widehat{L} = \tfrac{1}{m}\sum_{i\in[d]}\sum_{(\sigma,p)\in\widehat{W}}\sigma_i p_i = \tfrac{1}{m}\sum_{(\sigma,p)\in\widehat{W}}X_{\sigma,p}.$$

Observe that for every (σ, p) we have $|X_{\sigma,p}| \leqslant \sum_{i\in[d]} |p_i| \leqslant 2\,\mathrm{opt}(P)$ by Theorem 4.3. Now since the m samples in $\widehat{W}$ are independent, the variance of their sum is the sum of their variances, and we obtain

$$\mathrm{Var}(\widehat{L}) = \tfrac{1}{m^2} \cdot \mathrm{Var}\Big(\sum_{(\sigma,p)\in\widehat{W}} X_{\sigma,p} \Big) \leqslant \tfrac{1}{m^2} \cdot m \cdot (2\,\mathrm{opt}(P))^2 = \tfrac{4}{m}(\mathrm{opt}(P))^2. \tag{12}$$

This completes the proof of Theorem 4.8. ◀

4.2 Lower bound for value-preserving 0-coreset

The Helly number of a body S, denoted $\mathcal{H}(S)$, is the smallest positive integer h, such that if every subset of size h from a family of translations of S has a non-empty intersection, then the entire family also has a non-empty intersection. The Helly theorem shows that the Helly number of every convex body is smaller than or equal to $d + 1$.

The Helly number is equivalent to a worst-case result of value-preserving 0-coreset (cf. [6]). In the next proposition we show that the $d + 1$ upper bound is tight, that is, the Helly number of the ℓ_1 ball is $d + 1$.

▶ **Proposition 4.10.** *Assume $d > 2$. There exists $P \subset \mathbb{R}^d$, such that every value-preserving 0-coreset must have size $d + 1$.*

Proof. Consider the set of $d + 1$ points $P = \{p_0, \dots, p_d\}$, where $p_0 = (-1, -1, \dots, -1)$ and p_i is the vector with all entries equal to 1 except for a -1 in the i-th position. We will show that:

- The point $\vec{0} = (0, 0, \dots, 0)$ is the only center of P with a cost of d.
- Each subset of P containing d points has a cost strictly less than d.

To prove the first claim, let c be an optimal center of P, clearly $c \in [-1, 1]^d$. We know that for every i:

$$\mathrm{cost}(c, P) \geqslant \frac{1}{2}(\|c - p_0\|_1 + \|c - p_i\|_1) = \frac{1}{2}(2(d-1) + 2\,|c_i + 1|) = d - 1 + |c_i + 1|$$

On the other hand, $\mathrm{cost}(c, P) \leqslant d$, since the zero vector $\vec{0}$ has a cost d, consequently $1 \geqslant |c_i + 1|$ for all i. This implies $c_i \leqslant 0$ for all i, leading to:

$$\mathrm{cost}(c, P) \geqslant \|c - p_i\|_1 \geqslant (d-1) + (1 + c_i) - \sum_{j\neq i} c_j = d + c_i - \sum_{j\neq i} c_j$$

Since $d \geqslant \mathrm{cost}(c, P)$, it follows that $\sum_{j\neq i} c_j \geqslant c_i$ for all i. Given that $d > 2$, this implies $c = \vec{0}$.

For the second claim, consider any subset Q of d points. If $p_0 \notin Q$, then clearly $\mathrm{opt}(Q) < d$. Assuming $p_0 \in Q$ and, without loss of generality, $p_d \notin Q$, we set $c_i = -\frac{1}{d-2}$ for $i < d$, and $c_d = 1$. The cost for this choice of c is:

$$\begin{aligned}\mathrm{cost}(c, P) &= \max\Big\{2 + (d-1)|-1-c_i|, (d-2)|1-c_i| + |-1-c_i|\Big\}\\ &= \max\Big\{2 + (d-1)(1+c_i), (d-2)(1-c_i) + (1+c_i)\Big\} < d\end{aligned}$$ ◀

5 Generalization to discrete metric spaces

In this section we show how to extend our lower bounds results to other discrete metric spaces by providing two examples for Jaccard and Kendall's tau metric spaces.

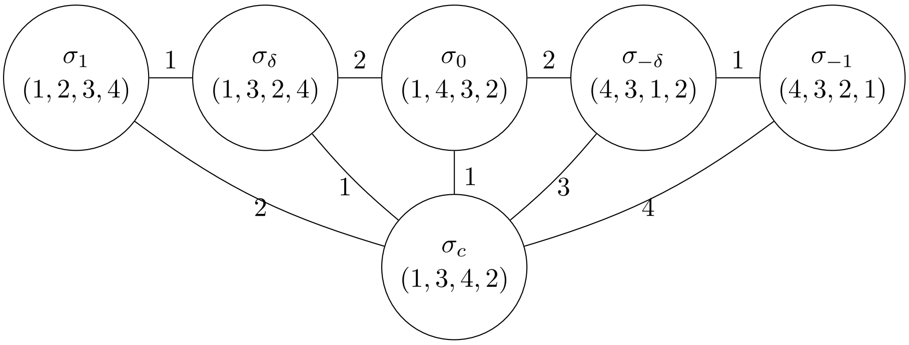

Figure 1 The figure shows the distances between the different base permutations used in the construction of the set P in Proposition 5.1. The distances between $\sigma_1, \sigma_\delta, \sigma_0, \sigma_{-\delta}, \sigma_{-1}$ are additive over the straight lines. The curved lines indicate the distances from $(1,3,4,2)$ to the others.

5.1 Theorem 2.2 in Kendall's tau metric space

Denote by $\mathcal{S}_d$ the set of permutations over the set of elements $[d]$. For two permutations $\sigma_1, \sigma_2 \in \mathcal{S}_d$, the Kendall's tau distance is the number of pairs (i,j) such that the order of i and j is reversed between σ_1 and σ_2. Formally,

$$\tau(\sigma_1, \sigma_2) = \left|\left\{\{i,j\} : (\sigma_1(i) - \sigma_1(j))(\sigma_2(i) - \sigma_2(j)) < 0\right\}\right|$$

Following a similar approach as in Section 2 we will prove a lower bound of size $2^{\Omega(d)}$ on the size of weak coreset.

▶ **Proposition 5.1.** *There exists a set $P \subseteq \mathcal{S}_{4d}$, such that for all $\epsilon \in [0, \frac{1}{3})$, if $Q \subseteq P$ is a weak ϵ-coreset of P, then $|Q| \geqslant 2^{\Omega(d)}$.*

Proof. We will design a gadget that will allow us to replicate the structure of P that is given in the proof of Theorem 2.2. We denote $\sigma_1 = (1,2,3,4)$, $\sigma_\delta = (1,3,2,4)$, $\sigma_0 = (1,4,3,2)$, $\sigma_{-\delta} = (4,3,1,2)$, $\sigma_{-1} = (4,3,2,1)$ and $\sigma_c = (1,3,4,2)$. The distances between these base permutations are detailed in Figure 1.

We will now introduce some notations that will allow us to describe permutations in $\mathcal{S}_{4d}$ using the base permutations. For $\sigma = (\pi_1, \dots, \pi_k)$, with a slight abuse of notation, we denote by $i + \sigma$ the permutation $(i + \pi_1, \dots, i + \pi_k)$ and also introduce an operator $\odot$ to denote the product $A \odot B = \{\sigma \cdot \hat{\sigma} : \sigma \in A, \hat{\sigma} \in B\}$ where the product of two permutations is their composition. We will extend a base permutation over 4 elements to a permutation over $4d$ elements using the notation $\sigma_{\vec{x}} = (0 + \sigma_x) \cdot (4 + \sigma_x) \cdot \dots \cdot (4(d-1) + \sigma_x)$, for $x \in \{-1, -\delta, 0, \delta, 1, c\}$. Now we are ready to describe P. Let $B = \odot_{i=0}^{d-1}\{4i + \sigma_\delta, 4i + \sigma_{-\delta}\}$, then $P = B \cup \{\sigma_{\vec{1}}, \sigma_{-\vec{1}}\}$. The permutation $\sigma_{\vec{0}}$, has cost $3d$ (see also Figure 1) and this cost is optimal the base permutations composing $\sigma_{\vec{1}}, \sigma_{-\vec{1}}$ are antipodal (for every $i \leqslant d-1$, every pair that agrees with $(4i + \sigma_1)$ disagrees with $(4i + \sigma_{-1})$, and vice versa).

Let Q be a weak ϵ-coreset of P. Notice that Q must contain both $\sigma_{-\vec{1}}$ and $\sigma_{\vec{1}}$, as otherwise, if $\sigma_{-\vec{1}}, \sigma_{\vec{1}} \notin Q$, consider $\sigma_{\vec{0}}$ as a center, and if Q contains just one of $\{\sigma_{-\vec{1}}, \sigma_{\vec{1}}\}$ say $\sigma_{\vec{1}}$, consider the permutation $\sigma_{\vec{c}}$ as a center. In both cases, we obtain that $2d = \mathrm{opt}(Q) < \frac{1}{1+\epsilon}\,\mathrm{opt}(P) = \frac{1}{1+\epsilon}3d$, in contradiction that Q is a weak ϵ-coreset.

Similarly to Proposition 2.3 it follows that if $|Q| \leqslant \frac{1}{\sqrt{2d}}2^{0.18d}$, then there exists $b \in B$ such that the Kendall tau distance $\tau(b,q) \leqslant 3d$ for every $q \in Q$. That is, b is an optimal center of Q. However, $\mathrm{cost}(b,P) = 4d$, by considering the opposite permutation of b in P, thus $|Q| \geqslant 2^{\Omega(d)}$. ◀

5.2 Theorem 3.1 in the Jaccard metric space

For two sets A and B, the Jaccard distance is defined as: $J(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|}$. We will use the symbol $\circledast$ to denote the product $A \circledast B = \{a \cup b : a \in A, b \in B\}$.

Following a similar approach as in Section 3.1 we will prove a lower bound of size $\Omega(\log d)$ on the size of weak coreset.

▶ **Proposition 5.2.** *Let $P = \circledast_{i=0}^{d-1} \{\{3i\}, \{3i, 3i+1, 3i+2\}\}$ be a collection of sets, i.e. points in the Jaccard metric space. For every $\epsilon < \frac{1}{3}$, a weak ϵ-coreset has $\Omega(\log d)$ points.*

Proof. Let $\mathcal{U} = \{0, 1, \ldots, 3d-1\}$ denote the universe. We also denote elements in $\mathcal{U}$ as *type 0* if their remainder modulo 3 is 0 and *type 1* otherwise. Note that all type 0 elements appear in every subset of P and thus appear in every optimal center of every coreset of P. It is then easy to verify that the set of optimal centers of P is the $\circledast$ product of $\{3i+1, 3i+2\}$ for $i \in [d]$, along with all type 0 elements, and that $\operatorname{opt}(P) = \frac{1}{2}$.

Using the same arguments as in the proof of Theorem 3.1, assume $Q \subset P$ is a weak ϵ-coreset of size $t \leqslant \frac{1}{2} \log d$. We denote by $A_i = \{3i\}$ and $B_i = \{3i, 3i+1, 3i+2\}$ the two building blocks of the set P, corresponding to -1 and 1 in the proof of Theorem 3.1. Every set in $p \in Q$ induces a partition of $\mathcal{U}$ into at most two parts, depending on their assignment to A_i or B_i. Again, we use Π to denote the common refinement of all these partitions.

Let c^* be any optimal 1-center of Q, and consider a partition $A \in \Pi$. Let k denote the number of type 1 elements from A in c^*. By selecting k different type 1 elements from A, say the first k type 1 indices of A, we preserve the cost of c^* over Q, as the number of differences c^* has with each $p \in Q$ remains unchanged. We apply this modifications for all partitions in Π and denote the new center by $\bar{c}^*$. However, by the construction of $\bar{c}^*$ there exists $p \in P$ such that p intersects at most one type 1 element of $\bar{c}^*$ in every partition and the union of p and c^* is $\mathcal{U}$. Consequently, the distance between c^* and p is at least $1 - \frac{d+|\Pi|}{3d} \geqslant \frac{2}{3} - \frac{2^t}{3d} = \frac{2}{3} - \frac{\sqrt{d}}{3d} > (1+\epsilon)\operatorname{opt}(P)$. ◀

References

1 Amir Abboud, MohammadHossein Bateni, Vincent Cohen-Addad, Karthik C. S., and Saeed Seddighin. On complexity of 1-center in various metrics. In *APPROX/RANDOM*, volume 275 of *LIPIcs*, pages 1:1–1:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.APPROX/RANDOM.2023.1.

2 Pankaj K. Agarwal, Sariel Har-Peled, and Kasturi R. Varadarajan. Approximating extent measures of points. *J. ACM*, 51(4):606–635, 2004. doi:10.1145/1008731.1008736.

3 Pankaj K. Agarwal, Sariel Har-Peled, and Kasturi R. Varadarajan. Geometric approximation via coresets. In *Combinatorial and computational geometry*, volume 52 of *MSRI Publications*, chapter 1, pages 1–30. Cambridge University Press, 2005.

4 Sepideh Aghamolaei and Mohammad Ghodsi. A composable coreset for k-center in doubling metrics. In *CCCG*, pages 165–171, 2018. URL: http://www.cs.umanitoba.ca/%7Ecccg2018/papers/session4A-p2.pdf.

5 Christian Bachmaier, Franz J. Brandenburg, Andreas Gleißner, and Andreas Hofmeier. On the hardness of maximum rank aggregation problems. *J. Discrete Algorithms*, 31:2–13, 2015. doi:10.1016/j.jda.2014.10.002.

6 René Brandenberg and Stefan König. No dimension-independent core-sets for containment under homothetics. *Discrete & Computational Geometry*, 49(1):3–21, 2013. doi:10.1007/s00454-012-9462-0.

7 Mihai Bădoiu and Kenneth L. Clarkson. Optimal core-sets for balls. *Comput. Geom.*, 40(1):14–22, 2008. doi:10.1016/j.comgeo.2007.04.002.

8 Mihai Bădoiu, Sariel Har-Peled, and Piotr Indyk. Approximate clustering via core-sets. In *STOC*, pages 250–257. ACM, 2002. doi:10.1145/509907.509947.

9 Marc Bury, Michele Gentili, Chris Schwiegelshohn, and Mara Sorella. Polynomial time approximation schemes for all 1-center problems on metric rational set similarities. *Algorithmica*, 83(5):1371–1392, 2021. doi:10.1007/s00453-020-00787-3.

10 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.

11 Diptarka Chakraborty, Debarati Das, and Robert Krauthgamer. Approximating the median under the ulam metric. In *SODA*, pages 761–775. SIAM, 2021. doi:10.1137/1.9781611976465.48.

12 Diptarka Chakraborty, Debarati Das, and Robert Krauthgamer. Clustering permutations: New techniques with streaming applications. In *ITCS*, volume 251 of *LIPIcs*, pages 31:1–31:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ITCS.2023.31.

13 Diptarka Chakraborty, Kshitij Gajjar, and Agastya Vibhuti Jha. Approximating the center ranking under ulam. In *FSTTCS*, volume 213 of *LIPIcs*, pages 12:1–12:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.FSTTCS.2021.12.

14 Timothy M. Chan. Faster core-set constructions and data-stream algorithms in fixed dimensions. *Computational Geometry*, 35(1):20–35, 2006. doi:10.1016/j.comgeo.2005.10.002.

15 Ke Chen. On coresets for k-median and k-means clustering in metric and euclidean spaces and their applications. *SIAM Journal on Computing*, 39(3):923–947, 2009. doi:10.1137/070699007.

16 Flavio Chierichetti, Ravi Kumar, Sandeep Pandey, and Sergei Vassilvitskii. Finding the jaccard median. In *SODA*, pages 293–311. SIAM, 2010. doi:10.1137/1.9781611973075.25.

17 Michael B. Cohen, Yin Tat Lee, Gary L. Miller, Jakub Pachocki, and Aaron Sidford. Geometric median in nearly linear time. In *STOC*, pages 9–21. ACM, 2016. doi:10.1145/2897518.2897647.

18 Vincent Cohen-Addad, David Saulpic, and Chris Schwiegelshohn. A new coreset framework for clustering. In *STOC*, pages 169–182. ACM, 2021. doi:10.1145/3406325.3451022.

19 Matan Danos. Coresets for clustering by uniform sampling and generalized rank aggregation. Master's thesis, Weizmann Institute of Science, 2021.

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

21 Dan Feldman. Core-sets: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(1):e1335, 2020. doi:10.1002/widm.1335.

22 Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In *STOC*, pages 569–578. ACM, 2011. doi:10.1145/1993636.1993712.

23 Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. doi:10.1016/0304-3975(85)90224-5.

24 Peter M Gruber and Jörg M Wills. *Handbook of convex geometry, Volume B*. North-Holland, 1993.

25 Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In *STOC*, pages 95–103. ACM, 2007. doi:10.1145/1250790.1250806.

26 Ming Li, Bin Ma, and Lusheng Wang. On the closest string and substring problems. *J. ACM*, 49(2):157–171, 2002. doi:10.1145/506147.506150.

27 Bin Ma and Xiaoming Sun. More efficient algorithms for closest string and substring problems. *SIAM Journal on Computing*, 39(4):1432–1443, 2010. doi:10.1137/080739069.

28 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. doi:10.1007/s13218-017-0519-3.

29 François Nicolas and Eric Rivals. Hardness results for the center and median string problems under the weighted and unweighted edit distances. *J. Discrete Algorithms*, 3(2-4):390–415, 2005. doi:10.1016/j.jda.2004.08.015.

30 Rina Panigrahy. Minimum enclosing polytope in high dimensions. *CoRR*, cs.CG/0407020, 2004. doi:10.48550/arXiv.CS/0407020.

31 Jeff M. Phillips. Coresets and sketches. In *Handbook of discrete and computational geometry*, chapter 48, pages 1269–1288. Chapman and Hall/CRC, 3rd edition, 2017. doi:10.1201/9781315119601.

32 Alvin Yan Hong Yao and Diptarka Chakraborty. Approximate maximum rank aggregation: Beyond the worst-case. In *FSTTCS*, volume 284 of *LIPIcs*, pages 12:1–12:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.FSTTCS.2023.12.

33 Hai Yu, Pankaj K. Agarwal, Raghunath Poreddy, and Kasturi R. Varadarajan. Practical methods for shape fitting and kinetic data structures using coresets. *Algorithmica*, 52(3):378–402, 2008. doi:10.1007/s00453-007-9067-9.

A Appendix

A.1 Lower bound for weak 0-coreset for inputs with a unique solution

We present an example of a set P of size $2d$, where P is the sole weak 0-coreset of itself. This demonstrates that the upper bound of $2d$ is indeed tight.

We start with a simple proposition regarding the distances from a given point within the unit hypercube to two antipodal vertices of the cube.

▶ **Proposition A.1.** *If $p \in \{-1,1\}^d$ and $x \in [-1,1]^d$ then $\|p-x\|_1 + \|(-p)-x\|_1 = 2d$.*

Proof. $\|p-x\|_1 + \|(-p)-x\|_1 = \sum_{i=1}^d |1-x_i| + |-1-x_i| = \sum_{i=1}^d ((1-x_i)+(1+x_i)) = \sum_{i=1}^d 2 = 2d$ ◀

An immediate consequence of the above proposition is that the cost of every optimal center of P is at least d for every set $P \subseteq \{-1,1\}^d$ containing an antipodal pair. Since $\vec{0}$ has cost at most d to P we get that $\mathrm{opt}(P) = d$ and that $\vec{0}$ is an optimal center. This is summarized in the following corollary.

▶ **Corollary A.2.** *If $P \subseteq [-1,1]^d$ contains antipodal pair, then $\mathrm{opt}(P) = d$.*

Let H denote the set of row vectors of the Hadamard matrix, and let $P = H \cup -H$.

▶ **Proposition A.3.** *There exists $P \subset \mathbb{R}^d$ with unique optimum, such that every weak 0-coreset must have size $2d$.*

Proof. Following Corollary A.2, $\mathrm{opt}(P) = d$. We will further show that if $Q \subseteq P$ does not contain an antipodal pair, then $\mathrm{opt}(Q) < d$. Note that for $a, b \in P$, $a \neq \pm b$ we have $a \cdot b = 0$. Also, since $a, b \in \{-1,1\}^d$, $a \cdot b = \sum_{i:a_i=b_i} 1 - \sum_{i:a_i \neq b_i} 1 = d - 2\sum_{i:a_i\neq b_i} 1$. That is:

$$\|a-b\|_1 = 2 \sum_{i:a_i \neq b_i} 1 = d - a \cdot b = d \tag{13}$$

That is, the distance from every $a \in P$ to all points $P \backslash \{a, -a\}$ is d.

Let $c = \frac{1}{|Q|}\sum_{q\in Q} q$ be some center. Since Q does not contain an antipodal pair it follows that $|Q| \leqslant d$. Now, using Equation 13, for every $\tilde{q} \in Q$:

$$\|c-\tilde{q}\|_1 = \frac{1}{|Q|}\left\|\sum_{q\in Q} q - \tilde{q}\right\|_1 = \frac{1}{|Q|}\left\|\sum_{q\in Q, q\neq\tilde{q}} d\right\|_1 = \frac{|Q|-1}{|Q|} d < d$$

That is $\mathrm{opt}(Q) < d$. It follows that every weak 0-coreset of P contains an antipodal pair. If a weak 0-coreset Q is a proper subset of P then there exists $q \in Q$ with $-q \notin Q$. But $\mathrm{cost}(q, Q) = d = \mathrm{opt}(Q)$, whereas $\mathrm{cost}(q, P) = 2d$. Hence P is the sole weak 0-coreset of P. ◀