On the Complexity of the Geometric Median Problem with Outliers

Vladimir Shenmaier

Abstract  In the Geometric Median problem with outliers, we are given a finite set of points in $d$-dimensional real space and an integer $m$, the goal is to locate a new point in space (center) and choose $m$ of the input points to minimize the sum of the Euclidean distances from the center to the chosen points. This problem can be solved “almost exactly” in polynomial time if $d$ is fixed and admits an approximation scheme PTAS in high dimensions. However, the complexity of the problem was an open question. We prove that, if the dimension of space is not fixed, Geometric Median with outliers is strongly NP-hard, does not admit approximation schemes FPTAS unless P=NP, and is W[1]-hard with respect to the parameter $m$. The proof is done by a reduction from the Independent Set problem. Based on a similar reduction, we also get the NP-hardness of closely related geometric 2-clustering problems in which it is required to partition a given set of points into two balanced clusters minimizing the cost of median clustering. Finally, we study Geometric Median with outliers in $\ell_\infty$ space and prove the same complexity results as for the Euclidean problem.

Keywords  Geometric median · Outlier detection · Capacitated 1-Median · Single location problem · Balanced clustering · Complexity

Mathematics Subject Classification (2010)  68Q25 · 90C27 · 90B85 · 6207

1 Introduction

We give an answer to the open question on the complexity of the Geometric Median problem with outliers, which is an extension of the classical problem of finding the geometric median (“minisum point”) of given $n$ points in Euclidean space.

Let $\| \cdot \|$ denote values of the Euclidean norm. The problem we consider is formulated as follows:

Supported by the program of fundamental scientific researches of the SB RAS, project 0314-2019-0014.

V.V. Shenmaier
Sobolev Institute of Mathematics, 4 Koptyug avenue, 630090 Novosibirsk, Russia
E-mail: shenmaier@mail.ru
Geometric Median with outliers. Given an n-element set $X$ in space $\mathbb{R}^d$ and an integer $m$. Find a center $c \in \mathbb{R}^d$ and an $m$-element subset $S \subseteq X$ to minimize the value of

$$cost(S, c) = \sum_{x \in S} ||x - c||.$$

This problem may also be referred to as Capacitated Euclidean 1-Median. In fact, we need to find a center $c \in \mathbb{R}^d$ minimizing the total distance from $c$ to $m$ nearest input points (see Fig. 1). In an equivalent version, we need to find a center $c \in \mathbb{R}^d$ maximizing the cardinality of a subset $S \subseteq X$ for which the value of $cost(S, c)$ does not exceed a given upper bound.

As in the usual Geometric Median problem, where $m = n$, the input points are considered as clients or demand points, while the desired center, as a location for placing a facility to serve the clients. The $n - m$ clients which are removed from this service in the solution are called outliers. The problem with outliers arises naturally in the following situations:

- The facility to be placed at the center we are looking for has a limited capacity and may not serve all the demand points.
- There exist an upper limit for the transportation cost, so we need to remove a minimum possible number of clients from the service to satisfy this limit.
- The data contains noise and errors. In this case, a few most distant clients may exert a disproportionally strong influence over the final solution and correspond to the least robust input points.
- The discovered outliers do not fit the rest of the data and are worthy of further investigation. In particular, once identified, they can be used to discover anomalies in the data.

Besides the practical considerations mentioned above, the problem is theoretically interesting. Strictly speaking, no polynomial-time algorithms are known even for the problem without outliers, where we find the geometric median of the whole set $X$. However, one can say that the usual Geometric Median problem is polynomially solvable “almost exactly” since, e.g., the algorithm from [1] computes its $(1 + \varepsilon)$-approximate solution in time $O(nd \log^3 \frac{1}{\varepsilon})$. 

Fig. 1 An instance of Geometric Median with outliers and its solution
The problem with outliers seems to be cardinally harder due to the exponential number of the subsets $S \subseteq X$. The first idea how to solve it is using the property that optimal subsets consist of $m$ input points nearest to some point in space. It allows to find an optimal solution by enumerating the cells of the $m$-order Voronoi diagram for the set $X$ and calculating the geometric medians of the subsets corresponding to these cells. So, by using the known algorithms for constructing high-order Voronoi diagrams [7,8] and for finding geometric medians [4], we compute a $(1 + \varepsilon)$-approximate solution of the Geometric Median problem with outliers in time $O(n^d m \log^3 \frac{n}{\varepsilon})$. Thus, in the case of fixed $d$, we can solve the problem “almost exactly” in polynomial time.

If the dimension of space is not fixed, a more effective idea is to use the approximation algorithm from [17,18] for the following general clustering problem, which contains Geometric Median with outliers:

**General Problem.** Given points $x_1, \ldots, x_n$ in space $\mathbb{R}^d$, integers $k, m \geq 1$, unit distance costs $f_{ij} \geq 0$, and powers $\alpha_{ij} \in [0, \alpha]$, $i = 1, \ldots, k$, $j = 1, \ldots, n$, where $\alpha$ is some constant. Find disjoint subsets $S_1, \ldots, S_k \subseteq \{1, \ldots, n\}$ of total cardinality $m$ and select a tuple $c_1, \ldots, c_k \in \mathbb{R}^d$ to minimize the value of

$$\sum_{i=1}^{k} \sum_{j \in S_i} f_{ij} \|x_j - c_i\|^\alpha_{ij}.$$

The framework suggested in [17,18] allows to find a $(1 + \varepsilon)^\alpha$-approximate solution of General Problem in time $O(n^k \log(2/\varepsilon)\varepsilon^{-\alpha} + kd)$ for any fixed $\varepsilon \in (0, 1]$.

**Related problems.** ElGindy and Keil [9] consider the equivalent version of the Geometric Median problem with outliers where it is required to find a maximum-cardinality subset of input points satisfying a given upper bound for the cost of the geometric median. They suggest an $O(n^{2.5} \log^4 n)$-time exact algorithm for the two-dimensional case with $\ell_1$ or $\ell_\infty$ distances. Two well-known single location problems closest to Geometric Median with outliers are Smallest $m$-Enclosing Ball and $m$-Variance. The first consists of finding $m$ input points minimizing the radius of the ball enclosing these points. In the second, we need to find $m$ input points minimizing the sum of the squared distances from these points to their mean. In high dimensions, both problems are strongly NP-hard [14,12] but admit approximation schemes with running time $O(n^{2/\epsilon} \log d)$ and $O(n^{2/\epsilon} \log^d)$ respectively.

Finally, note that the discrete version of $k$-Median with outliers, in which all the centers must be selected from a given finite set, and also $k$-Means with outliers, Uncapacitated Facility Location with outliers, and other similar problems were considered by Charikar et al. [2], Chen [3], Cohen-Addad et al. [5], and Krishnaswamy et al. [13]. In particular, they suggest constant-factor algorithms for these problems in the case of non-fixed $k$.

**Our contributions.** Surprisingly, the complexity of Geometric Median with outliers was an open question. We prove that, in high dimensions, this problem is
strongly NP-hard, does not admit approximation schemes FPTAS unless P=NP, and is W[1]-hard with respect to the parameter m. The proof is based on a reduction from the Independent Set problem. The main idea of this reduction is constructing instances in which the geometric median of any subset of points corresponding to vertices of a given graph is well approximated by the mean of these points, while the total distance from the points to their mean monotonously depends on the number of edges between the corresponding vertices.

By using a similar reduction from the problem of finding a maximum bisection in a 3-regular graph, we get the NP-hardness and the non-existence of approximation schemes FPTAS for the following closely related problems:

**Equal-Size Geometric 2-Median.** Given an n-element set X in space $\mathbb{R}^d$, where n is even. Find centers $x, y \in \mathbb{R}^d$ and a partition of X into two equal-size subsets $S$ and $X \setminus S$ to minimize the value of $\text{cost}(S, x) + \text{cost}(X \setminus S, y)$.

**Balanced Geometric 2-Median.** Given an n-element set X in space $\mathbb{R}^d$. Find centers $x, y \in \mathbb{R}^d$ and a partition of X into two subsets $S$ and $X \setminus S$ to minimize the value of $|S| \cdot \text{cost}(S, x) + |X \setminus S| \cdot \text{cost}(X \setminus S, y)$.

Additionally, we study the $\ell_\infty$-Median problem with outliers, which consists of finding a center minimizing the sum of the $\ell_\infty$ distances to m input points. For this problem, the same complexity results are proved as for the Euclidean case.

2 Complexity of the Euclidean problem

In this section, we prove the hardness of finding an optimum m-element subset in the case when the distances between input points and the desired center are defined by the Euclidean norm.

2.1 Reduction from Independent Set

We construct a reduction to Geometric Median with outliers from the classic NP-hard problem of determining the existence of an independent set of a given cardinality in a general graph.

Let G be any undirected graph on some n-element set of vertices V. First, construct the following auxiliary graph $G'$. The vertices of $G'$ are the same as in G and the set of edges contains all the edges of G and also $n - \deg_v$ copies of the loop edge $(v, v)$ for each vertex $v \in V$, where $\deg_v$ is the degree of v in G. So each vertex in $G'$ is incident to exactly n edges (see Fig. 2). Then fix an arbitrary orientation on the edges of $G'$: for every non-loop edge, choose an endpoint of this edge which it is “outgoing from” and one which it is “incoming to”. For definiteness, we will assume that every loop edge is “incoming to” its endpoint. Denote by E the set of edges of $G'$. Next, map each vertex $v \in V$ to the point $x_v \in \mathbb{R}^{E \cup V}$ with the following coordinates: $x_v(v) = M$, where $M$ is some big integer which will be specified later; $x_v(e) = 1$ for every edge $e \in E$ outgoing from v and $x_v(e) = -1$ for ones incoming to v; all the other coordinates are zero (see Fig. 2). Finally, define the instance of the Geometric Median problem with outliers corresponding to the graph G as the set $X = \{x_v \mid v \in V\}$. 
Comment 1 Adding the loop edges and selecting an edge orientation in the auxiliary graph \( G' \) provide that all the vectors \( x_v, v \in V \) have the same number of the \( \pm 1 \) coordinates and, for any two vectors \( x_v, x_u \), the sets of their non-zero coordinates are either disjoint or intersecting by the only coordinate \( e \in E \), for which \( x_v(e) = -x_u(e) \). As a corollary, the distance between the vectors \( x_v \) and \( x_u \) for any adjacent different vertices \( u, v \in V \) is greater than that for non-adjacent ones. Indeed, \( \|x_v - x_u\|^2 = 2M^2 + \|x^E_v - x^E_u\|^2 \), where \( x^E \) denotes the projection of \( x \) into space \( \mathbb{R}^E \), while the value of \( \|x^E_v - x^E_u\|^2 \) is \( 2(n - 1) + (1 + 1)^2 = 2n + 2 \) if the vertices \( u \) and \( v \) are adjacent and is \( 2n \) otherwise.

Comment 2 Setting the coordinates \( x_v(v), v \in V \), equal to a big value of \( M \) provides that the geometric median of any subset \( Y \subseteq X \) is close to its mean, while, as it will be proved, the total distance from the points \( x_v \in Y \) to their mean has an almost affine dependence on the sum of pairwise squared distances between the corresponding vectors \( x^E_v \).

Idea of the reduction. We will prove that, given an \( m \)-element subset of vertices \( S \subseteq V \), the minimum value of \( \text{cost}(X_S, c) \), where \( X_S = \{x_v \mid v \in S\} \), over all \( c \in \mathbb{R}^d \) monotonously depends on the number \( \ell_S \) of the edges in the original graph \( G \) connecting the vertices of \( S \) to each other. Therefore, if the set \( X_S \) is an optimal solution of Geometric Median with outliers for the given \( m \), then \( S \) is an \( m \)-element subset of vertices with the minimum number \( \ell_S \).

For any finite set \( Y \subseteq \mathbb{R}^E \cup V \), denote by \( c(Y) \) and \( \mu(Y) \) the mean and the geometric median of this set respectively:

\[
c(Y) = \frac{1}{|Y|} \sum_{y \in Y} y \quad \text{and} \quad \mu(Y) = \arg \min_{c \in \mathbb{R}^E \cup V} \text{cost}(Y, c).
\]
Let $S \subseteq V$ be any subset of vertices with an arbitrary cardinality $m \geq 3$. Put $y = \mu(X_S) - c(X_S)$, $\delta = \|y\|$, $z_v = x_v - c(X_S)$, and $\zeta_v = \sum_{e \in E} z^2(e)$ for each $v \in S$.

Then $\sum_{v \in S} z_v$ is the zero vector and the following property holds:

**Property 1**

(a) $\zeta_v \in \left[ n - \frac{2n}{m} + \frac{2n}{m^2}, n + \frac{n}{m} - \frac{2n}{m^2} \right] \subseteq (0.55m, 1.13n)$;

(b) $\|z_v\| = \sqrt{A^2 + \zeta_v}$, where $A = M\sqrt{1 - 1/m}$.

**Proof**

(a) Given a vector $x \in \mathbb{R}^{E \setminus V}$, let $E(x) = \{ e \in E \mid x(e) \neq 0 \}$. Then it is easy to see that the set $E(c(X_S))$ consists of the loops in $S$ and the edges connecting $S$ and $V \setminus S$. Hence, we have $|E(c(X_S))| = nm - 2\ell_S$. Next, the set $E(\zeta_v)$ consists of all the elements of $E(c(X_S))$ and also the edges connecting $v$ and the other vertices of $S$. So $|E(\zeta_v)| = nm - 2\ell_S + \Delta_S^v$, where $\Delta_S^v$ is the degree of $v$ in the subgraph of $G$ induced by $S$. For $\Delta_S^v$ coordinates $e \in E(\zeta_v)$, the values of $z_v(e)$ are $\pm 1$; for $n - \Delta_S^v$ coordinates, these values are $\pm (1 - 1/m)$; for the other coordinates from $E(\zeta_v)$, these values are $\pm 1/m$. Then

$$
\zeta_v = \Delta_S^v + (n - \Delta_S^v)(1 - \frac{1}{m}) + \frac{n(m - 1) - 2\ell_S + \Delta_S^v}{m^2} = n \left(1 - \frac{1}{m}\right)^2 + \frac{2\Delta_S^v}{m} - \frac{\Delta_S^v}{m^2} + \frac{n}{m} - \frac{2\ell_S}{m^2} + \frac{\Delta_S^v}{m^2} = n - \frac{n - 2\ell_S + \Delta_S^v}{m^2}.
$$

But $\ell_S \geq \Delta_S^v$ and $\Delta_S^v < n$, so $\zeta_v \leq n - \frac{n - 2\ell_S}{m^2} + \frac{2\Delta_S^v}{m} < n + \frac{n}{m} - \frac{2n}{m^2}$. The latter is maximized when $m = 4$, therefore, we have $\zeta_v < 9n/8 < 1.13n$.

On the other hand, we have $\ell_S = \Delta_S^v + \ell_{S \setminus \{v\}}$, $\ell_{S \setminus \{v\}} \leq \frac{(m - 1)(m - 2)}{2}$, and $\Delta_S^v \geq 0$, so

$$
\zeta_v = n - \frac{n - 2\ell_S + \Delta_S^v}{m^2} \geq n - \frac{n - 2\ell_{S \setminus \{v\}} - 2\Delta_S^v}{m^2} = n - \frac{n - 3}{m^2} \geq 0.
$$

But it can be easily proved that $n - 1 - \frac{n - 3}{m^2} - \frac{2}{m^2} \geq n - \frac{2n}{m^2} + \frac{2n}{m^2}$ for all $n \geq m \geq 3$. It follows that $\zeta_v \geq 5n/9 > 0.55m$.

(b) Obviously, we have $\|z_v\|^2 = M^2 \left(1 - \frac{1}{m}\right)^2 + (m - 1)\left(\frac{M}{m}\right)^2 + \zeta_v = A^2 + \zeta_v$, which implies the required equation. 

The main geometric statement underlying the proposed reduction is that the distance between $c(X_S)$ and $\mu(X_S)$ is very close to zero for big $M$; $\delta < \frac{4.5n}{mA}$ if $A \geq 100nm$. This statement will be proved in Section 2.2 (Lemma 5).

**Lemma 1** Suppose that $A \geq 100nm$. Then, for some $\gamma \in [-1, 1]$, we have

$$
cost(X_S, \mu(X_S)) = mA + \frac{n(m - 1)}{2A} + \frac{\ell_S}{mA} + \gamma \frac{23n^2m}{A^2}.
$$
Proof By the cosine theorem, we have
\[
\text{cost}(X_S, \mu(X_S)) = \sum_{v \in S} \sqrt{\|z_v\|^2 + \|y\|^2 - 2\langle y, z_v \rangle} = \sum_{v \in S} \sqrt{A^2 + \zeta_v + \delta^2 - 2\langle y, z_v \rangle},
\]
where \(\langle \cdot, \cdot \rangle\) is the dot product. Property 1, Lemma 5, and the condition for \(A\) yield that
\[
\zeta_v + \delta^2 - 2\langle y, z_v \rangle \leq 1.13n + \frac{(4.5n)^2}{(mA)^2} + 2 \frac{4.5n}{mA} \sqrt{A^2 + 1.13n} < 4.2n,
\]
so
\[
\left| \frac{\zeta_v + \delta^2 - 2\langle y, z_v \rangle}{A^2} \right| < \frac{4.2n}{A^2} < 0.001.
\]
On the other hand, Taylor’s theorem (in the Lagrange remainder form) implies the equation
\[
\sqrt{1 + \varepsilon} = 1 + \frac{\varepsilon}{2} - \theta \frac{\varepsilon^2}{7.99^2} \text{ for some } \theta \in [0, 1] \text{ if } |\varepsilon| \leq 0.001.
\]
Therefore,
\[
\text{cost}(X_S, \mu(X_S)) = \sum_{v \in S} A \left( 1 + \frac{\zeta_v + \delta^2 - 2\langle y, z_v \rangle}{2A^2} - \theta \frac{(\zeta_v + \delta^2 - 2\langle y, z_v \rangle)^2}{7.99^2 A^4} \right),
\]
where \(\theta \in [0, 1]\). Since \(\sum_{v \in S} z_v\) is the zero vector, the sum of the terms \(\langle y, z_v \rangle\) is zero.

Then, taking into account the inequality \(\delta < \frac{4.5n}{mA}\) and the above observations, we obtain that
\[
\text{cost}(X_S, \mu(X_S)) = \sum_{v \in S} \left( A + \frac{\zeta_v}{2A} + \theta \frac{(4.5n)^2}{2mA^3} - \theta_1 \frac{(4.2n)^2}{7.99mA^3} \right.
\]
\[
\left. + \frac{2m^2 - m + 4f_S}{2m} \right) + \gamma \frac{2.3n^2m}{A^3},
\]
where \(\theta_1, \theta_2 \in [0, 1]\) and \(\gamma \in [-1, 1]\).

Next, calculate the value of \(\sum_{v \in S} \zeta_v\). Given a vertex \(v \in S\), consider the projection \(x_v^E \in \mathbb{R}^E\) of the vector \(x_v\) into space \(\mathbb{R}^E\). It is easy to see that the value of \(\zeta_v\) equals the squared distance from the vector \(x_v^E\) to the mean of these vectors. But it is well-known (e.g., see [11,14]) that the sum of such squared distances equals the sum of all the pairwise squared distances divided by 2m:
\[
\sum_{v \in S} \zeta_v = \frac{1}{2m} \sum_{v \in S} \sum_{u \in S} \|x_v^E - x_u^E\|^2.
\]
At the same time, by the construction of the vectors \(x_v^E\), each pairwise squared distance \(\|x_v^E - x_u^E\|^2\) equals \(2n + 2\) if the vertices \(v\) and \(u\) are adjacent and equals \(2n\) otherwise (see Comment 1). So we have
\[
\sum_{v \in S} \zeta_v = \frac{2n (m^2 - m) + 4f_S}{2m} = n(m - 1) + \frac{2f_S}{m}.
\]
It follows the required equation. The lemma is proved.
Theorem 1 Geometric Median with outliers is strongly NP-hard, does not admit approximation schemes FPTAS unless P=NP, and is \(W[1]\)-hard with respect to the parameter \(m\).

Proof Suppose that \(G\) is any \(n\)-vertex graph, \(S\) is any \(m\)-element subset of its vertices, where \(m \geq 3\), and put \(M = 123nm\). Then \(A \geq M\sqrt{\frac{2}{3}} > 100nm\) and, by Lemma 1, we have the equation \(\text{cost}(X_S, \mu(X_S)) = f(n, m, \ell, \gamma)\), where

\[
f(n, m, \ell, \gamma) = mA + \frac{n(m-1)}{2A} + \frac{\ell}{mA} + \gamma \frac{2m^2}{A^2}\]

and \(\gamma \in [-1, 1]\). By the choice of \(M\), the absolute value of the latest term in this expression is at most 0.00023 times of \(mA\). Therefore, \(S\) is an independent set in the graph \(G\) if and only if \(\text{cost}(X_S, \mu(X_S)) < f(n, m, 1, -1)\). Thus, the Independent Set problem is reduced to Geometric Median with outliers. Taking into account that \(M\) is an integer bounded by a polynomial in the length of the input, it gives the strong NP-hardness of the latter problem.

Moreover, by the above, if \(\ell \geq 1\), then

\[
f(n, m, \ell, -1) - f(n, m, 0, 1) \geq \frac{1 - 2 \cdot 0.00023}{mA} > \frac{0.999}{123n^3} > \frac{0.008}{n^6}.
\]

On the other hand, both values \(f(n, m, \ell, -1)\) and \(f(n, m, 0, 1)\) are less than \(mM \leq 123n^3\), which implies that Geometric Median with outliers is NP-hard to approximate within a factor of \(1 + \frac{1}{p(n)}\). But, for an arbitrary polynomial \(p(n)\), any approximation scheme FPTAS allows to get a \((1 + \frac{1}{p(n)})\)-approximation in polynomial time. Therefore, the existence of such schemes is impossible unless P=NP.

Finally, note that the constructed reduction can be represented as a parameterized reduction of the \(W[1]\)-hard Independent Set problem parameterized by \(m\) [6] to Geometric Median with outliers parameterized by \(m\). It follows the statement on the \(W[1]\)-hardness. The theorem is proved. \(\square\)

Balanced Geometric 2-Median clustering

Based on the same construction of the set \(X\), we can give similar reductions from the well-known NP-hard problem of finding a maximum bisection in a 3-regular graph [10] to the Equal-Size and Balanced Geometric 2-Median problems.

Indeed, in a 3-regular graph, the number of the edges connecting an \(m\)-element subset of vertices \(S\) with its complement equals \(3m - 2\ell_S\). Therefore, by Lemma [7] the minimum value of \(\text{cost}(X_S, \mu(X_S)) + \text{cost}(X_{V\setminus S}, \mu(X_{V\setminus S}))\) in the case when \(m = n - m\) corresponds to a bisection with the maximum number of crossing edges. It gives the NP-hardness and the non-existence of approximation schemes FPTAS for Equal-Size Geometric 2-Median.

To get the similar result for Balanced Geometric 2-Median, we need to note that the value of the objective function of this problem on the bipartition \(X_S, X \setminus X_S\) equals the expression

\[
f(n, m, \ell_1, \gamma_1, \ell_2, \gamma_2) = mf(n, m, \ell_1, \gamma_1) + (n - m)f(n, n - m, \ell_2, \gamma_2),
\]
where $\ell_{1,2} \leq 3n/2$ and $\gamma_{1,2} \in [-1, 1]$. But this expression can be approximated by the value of $m^2 M \sqrt{1 - \frac{1}{m} + (n - m)^2 M \sqrt{1 - \frac{1}{n - m}}}$ for big $M$. It follows that the minimum of $f(n, m, \ell_{1,2}, \gamma_{1,2})$ is attained when $m = \lfloor n/2 \rfloor$ and, by the above observations, corresponds to a maximum bisection.

### 2.2 Location of the geometric median

In this section, we prove the upper bound for the distance between the geometric median and the mean of the set $X_S$ underlying Lemma 1.

Let $S$ be an arbitrary $m$-element subset of $V$. Since the mean of any single- and two-element set is also its geometric median, we will assume that $m \geq 3$.

#### Rough estimations

First, we prove some rough estimation for the value $\|\mu(X_S) - c(X_S)\|$ which will be needed for the proof of the stronger upper bound.

We start by introducing some notations. Given a set $Y \subseteq \mathbb{R}^{E \cup V}$, denote by $\text{Aff}(Y)$ the affine hull of this set. Denote by $V$ the subspace in space $\mathbb{R}^{E \cup V}$ consisting of all the vectors $y$ with the zero edge coordinates: $y(e) = 0$, $e \in E$. Given a vertex $v$ of the graph $G$, define the point $y_v \in \mathbb{R}^{E \cup V}$ for which $y_v(v) = M$ and all the other coordinates are zero. In other words, the vector $y_v$ is the orthogonal projection of $x_v$ into the space $V$. Note that the distance from the mean $c(Y_S)$ of the set $Y_S = \{y_v \mid v \in S\}$ to each of its elements is $A = M \sqrt{1 - 1/m}$.

**Lemma 2** The geometric median of the set $Y_S$ coincides with its mean. Moreover, if a vector $z \in \mathbb{R}^{E \cup V}$ is orthogonal to the affine space $\text{Aff}(Y_S)$, then the minimum of the value $\text{cost}(Y_S, z + x)$ subject to $x \in \text{Aff}(Y_S)$ is attained at $x = c(Y_S)$.

**Proof** The derivative of the function $\text{cost}(Y_S, x)$ along any coordinate $u$ is

$$\left( \sum_{v \in S} \|x - y_v\| \right)'_u = \sum_{v \in S} \frac{(x - y_v)(u)}{\|x - y_v\|}. $$

So the value of this derivative at the point $x = c(Y_S)$ is

$$\frac{1}{A} \left( m \cdot c(Y_S) - \sum_{v \in S} y_v \right)(u) = 0. $$

It follows that the point $c(Y_S)$ is the geometric median of the set $Y_S$.

Then the derivative of the function $f(x) = \text{cost}(Y_S, x)$ along any direction $\phi$ in the affine space $\text{Aff}(Y_S)$ is equal to zero at the point $x = c(Y_S)$: $f'_\phi(c(Y_S)) = 0$. On the other hand, the derivative of the function $g(x) = \text{cost}(Y_S, z + x)$ along any direction $\phi$ in $\text{Aff}(Y_S)$ is

$$g'_\phi(x) = \left( \sum_{v \in S} \sqrt{\|z\|^2 + \|x - y_v\|^2} \right)'_\phi = \sum_{v \in S} \frac{\|x - y_v\| \left( \|x - y_v\| \right)'_\phi}{\sqrt{\|z\|^2 + \|x - y_v\|^2}}. $$

This completes the proof of Lemma 2.
It follows that the point $x = c(Y_S)$ is 

$$
\frac{A}{\sqrt{\|z\|^2 + A^2}} \sum_{v \in S} (\|x - y_v\|)'(c(Y_S)) = \frac{A}{\sqrt{\|z\|^2 + A^2}} f'_\varphi(c(Y_S)) = 0.
$$

It follows that the point $x = c(Y_S)$ is the minimum of the function $\text{cost}(Y_S, z + x)$ subject to $x \in \text{Aff}(Y_S)$. The lemma is proved.

**Lemma 3** Suppose that $A \geq 300n$ and $\text{dir}_v = \frac{y_v - c(Y_S)}{\|y_v - c(Y_S)\|}$, $v \in S$. Then, for each $v \in S$, the following holds: $|\langle \mu(X_S) - c(Y_S), \text{dir}_v \rangle| < 1.45\sqrt{n}$.

**Proof** Let $\mu'$ be the orthogonal projection of the point $\mu(X_S)$ into the space $\Gamma$. Then $\mu' \in \text{Aff}(Y_S)$ and $\text{cost}(Y_S, \mu') \leq \text{cost}(X_S, \mu(X_S))$ since the set $Y_S$ is the projection of the set $X_S$ into $\text{Aff}(Y_S)$. Next, consider the affine subspace $H_v$ in $\text{Aff}(Y_S)$ orthogonal to the vector $y_v - c(Y_S)$ and passing through the point $\mu'$. Denote by $\mu_v$, the intersection point of this subspace with the line passing through the points $c(Y_S)$ and $y_v$ (see Fig. 3). Then $\mu_v$ is the point nearest to $y_v$ in $H_v$.

On the other hand, it is easy to see that $H_v$ can be obtained by shifting the affine subspace $\text{Aff}(Y_S \setminus \{y_v\})$ by the vector $z = \mu_v - c(Y_S \setminus \{y_v\})$ and $z$ is orthogonal to $\text{Aff}(Y_S \setminus \{y_v\})$. Hence, applying Lemma 2 to the set $Y_S \setminus \{y_v\}$, we get that the point $\mu_v$ is the minimum of the function $\text{cost}(Y_S \setminus \{y_v\}, x)$ subject to $x \in H_v$. It follows that $\text{cost}(Y_S, \mu_v) \leq \text{cost}(Y_S, \mu') \leq \text{cost}(X_S, \mu(X_S)) \leq \text{cost}(X_S, c(Y_S))$.

By the cosine theorem, we have

$$
\text{cost}(Y_S, \mu_v) = A - \varrho + (m - 1) \sqrt{A^2 + \varrho^2 + \frac{2Ag}{m - 1}},
$$

where $|\varrho| = \|\mu_v - c(Y_S)\|$. At the same time,

$$
\text{cost}(X_S, c(Y_S)) = m \sqrt{A^2 + n} \leq m \left( A + \frac{n}{2A} \right).
$$

Then, since $\text{cost}(Y_S, \mu_v) \leq \text{cost}(X_S, c(Y_S))$, we get the inequality

$$
\sqrt{A^2 + \varrho^2 + \frac{2Ag}{m - 1}} \leq \frac{m(A + \frac{n}{2A}) + \varrho - A}{m - 1} = \frac{m \frac{A + \frac{n}{2A}}{m - 1} + \varrho}{m - 1}.
$$

\[\text{Fig. 3 Proof of Lemma 3} \]
which implies that
\[ A^2 + \varepsilon^2 + \frac{2A\varepsilon}{m-1} \leq A^2 + 2A \frac{nm}{2A} \frac{\varepsilon}{m-1} + \left( \frac{nm}{2A} + \varepsilon \right)^2. \]
Taking into account that \( m \geq 3 \) and, therefore, \( m^{m-1} \leq 1.5 \), we have
\[ \varepsilon^2 \leq 1.5n^2 + \frac{9n^2}{16A^2} + \frac{3nA + \varepsilon^2}{4}. \]
So \(|\varepsilon| < 1.45\sqrt{n}\) if \( A \geq 300n \).

It remains to note that the vectors \( \mu(X_S) - \mu' \) and \( c(X_S) - c(Y_S) \) are orthogonal to the space \( \text{Aff}(Y_S) \) and, therefore, to the vector \( \text{dir}_v \). But the vector \( \mu' - \mu_v \) is also orthogonal to \( \text{dir}_v \) by the choice of \( \mu_v \). Thus, we obtain that
\[ \langle \mu(X_S) - c(X_S), \text{dir}_v \rangle = \langle \mu_v - c(Y_S), \text{dir}_v \rangle, \]
which equals \(|\varepsilon|\) in absolute value. The lemma is proved. \( \square \)

**Lemma 4** Suppose that \( A \geq 300 \). Then \( ||\mu(X_S) - c(X_S)|| < 1.8\sqrt{mn} \).

**Proof** For every \( v \in S \), consider the plane \( P_v \) containing the points \( y_v, c(Y_S) \), and the origin \( 0 \). Note that this plane is orthogonal to the affine space \( \text{Aff}(Y_S) \) since the line \( c(Y_S)0 \) is orthogonal to \( \text{Aff}(Y_S) \) and is contained in \( P_v \). At the same time, the projection \( \mu' \) of the point \( \mu(X_S) \) into the space \( \Gamma \) lies in \( \text{Aff}(Y_S) \). So the orthogonal projection \( \mu'' \) of the point \( \mu' \) into the plane \( P_v \) belongs to the intersection of \( P_v \) with \( \text{Aff}(Y_S) \), i.e., lies on the line \( y_v c(Y_S) \) (see Fig. 4). Then
\[
(\mu(X_S) - c(X_S))(v) = (\mu' - c(Y_S))(v) = (\mu'' - c(Y_S))(v) =
\]
\[
||\mu'' - c(Y_S)|| \cos \alpha = ||\mu(X_S) - c(X_S), \text{dir}_v|| \cos \alpha,
\]
where \( \alpha \) is the angle between the lines \( y_v 0 \) and \( y_v c(Y_S) \). Taking into account the equation \( \cos \alpha = \sqrt{1 - \frac{1}{m}} \) and Lemma 3, it follows that the absolute value of the \( v \)-th coordinate of the vector \( \mu(X_S) - c(X_S) \) is less than \( 1.45\sqrt{m}\sqrt{1 - \frac{1}{m}} \).

On the other hand, for each edge \( e \in E \) incident to \( S \), the absolute value of the
\( e \)-th coordinate of this vector is at most 1 since the geometric median is a convex combination of the points \( x_v, v \in S \). For every other coordinate \( i \in E \cup V \) and every \( v \in S \), we have \( x_v(i) = 0 \), which implies that \( (\mu(X_S) - c(X_S))(i) = 0 \). Thus, we obtain the inequality \( \|\mu(X_S) - c(X_S)\|^2 < 1.45^2n^2 + m < 3.2nm \), so \( \|\mu(X_S) - c(X_S)\| < 1.8\sqrt{nm} \). The lemma is proved. \( \square \)

A tighter estimation

Now, we get the tighter estimation for the geometric median location which is used in Section 2.1. We start by proving the following geometric properties of the vectors \( z_v = x_v - c(X_S) \), \( v \in S \).

**Property 2** If \( M \geq 300n \), then the angle between the vectors \( z_u \) and \( z_v \) for different \( u, v \in S \) lies in the interval \( \left( \arccos \frac{-0.99}{m - 1}, \arccos \frac{-1.01}{m - 1} \right) \).

**Proof** Indeed, \( \langle z_u, z_v \rangle = \sum_{a \in V} z_u(a)z_v(a) + \sum_{e \in E} z_u(e)z_v(e) \). The first term in this expression is \( 2M(1 - 1/m)(-M/m) + (m - 2)M^2/m^2 = -M^2/m \); the second is at least \(-1 - (2n - 2)(1 - 1/m)/m > -1 - 2n/m \) and is at most \((nm - 2n)/m^2 \).

Then, taking into account Property 1(a), we obtain that the cosine of the angle between the vectors \( z_u \) and \( z_v \) is between the values \( a = \frac{-M^2/m - 1 - 2n/m}{A^2 + 0.55m} \) and \( b = \frac{-M^2/m + (nm - 2n)/m^2}{A^2 + 1.33n} \). It remains to note that \( A^2 = M^2(1 - 1/m) \) and that \( M \) is sufficiently big, so we have \( a > \frac{-1.01}{m - 1} \) and \( b < \frac{-0.99}{m - 1} \). \( \square \)

**Property 3** If \( M \geq 300n \), then \( \sum_{v \in S} \langle y, z_v \rangle^2 \leq \frac{m + 1.515}{2} \max_{v \in S} \|z_v\|^2 \).

**Proof** By Property 2, the angles between the vectors \( z_u \) and \( \pm z_v \) for \( u \neq v \) are at least \( \alpha = \pi - \arccos \frac{1.01}{m - 1} = \arccos \frac{1.01}{m - 1} \). So there exists at most one vector \( z_u, u \in S \), such that the angles between \( y \) and \( \pm z_v \) are at most some \( \beta < \alpha/2 \).

Then \( \sum_{v \in S} \langle y, z_v \rangle^2 \leq \|y\|^2 \max_{v \in S} \|z_v\|^2 t \), where \( t = \cos^2 \beta + (m - 1)\cos^2(\alpha - \beta) \).

Note that \( t = \frac{m}{2} + \frac{\cos(2\beta) + (m - 1)\cos(2\alpha - 2\beta)}{2} \) and that this expression has the only local extremum on \( [0, \alpha/2] \); at the point \( \beta = 0 \). Therefore, we have \( t = \max(t(0), t(\alpha/2)) = t(\alpha/2) = m\cos(\alpha/2) = \frac{m}{2}(1 + \cos(\alpha)) \), which equals \( \frac{m}{2} \left( 1 + \frac{1.01}{m - 1} \right) = \frac{m}{2} + \frac{1.01}{2} + \frac{1.01}{2(m - 1)} \leq \frac{m + 1.515}{2} \) if \( m \geq 3 \). \( \square \)

**Property 4** If \( M \geq 300n \), then \( \sum_{v \in S} \langle y, z_v \rangle \zeta_v \leq 1.061n \|y\| \max_{v \in S} \|z_v\| \).

**Proof** By Property 1(a), we have \( \zeta_v = n - \frac{0.5n}{m} + t_v n \), where \( |t_v| \) is bounded by the value \( c = \frac{1.5}{m} - \frac{2}{m^2} \). Since \( \sum_{v \in S} z_v \) is the zero vector, it follows the inequality
Geometric Median with Outliers

\[ \sum_{v \in S} \langle y, z_v \rangle \leq cn \sum_{v \in S} |\langle y, z_v \rangle|. \]
Next, by using the observations from the proof of Property 3, we obtain that \( \sum_{v \in S} |\langle y, z_v \rangle| \leq \|y\| \max_{v \in S} \|z_v\| t \), where

\[ t = m \cos(\alpha/2) = m \sqrt{\cos \frac{\alpha}{2}} = \frac{m \sqrt{1 + \frac{1}{m} - 1}}{m - 1}. \]

Therefore, the multiplier \( ct \) equals \( \frac{1}{\sqrt{2}} \left( 1.5 - \frac{2}{m} \right) \sqrt{1 + \frac{1.01}{m - 1}}, \) which is less than \( \frac{1.5}{\sqrt{2}} < 1.061. \)

**Lemma 5** Suppose that \( A \geq 100nm \). Then \( \|\mu(X_S) - c(X_S)\| < \frac{4.5n}{mA}. \)

**Proof** We will compare the values of \( \text{cost}(X_S, c(X_S)) \) and \( \text{cost}(X_S, \mu(X_S)) \). The first is

\[ \sum_{v \in S} \|z_v\| = \sum_{v \in S} \sqrt{A^2 + \zeta_v}, \]

while the second is

\[ \sum_{v \in S} \sqrt{\|z_v\|^2 + \|y\|^2 - 2\langle y, z_v \rangle} = \sum_{v \in S} \sqrt{A^2 + \zeta_v + \delta^2 - 2\langle y, z_v \rangle}. \]

Note that \( \left| \frac{\zeta_v}{A^2} \right| < 0.001 \) and

\[ \left| \frac{\zeta_v + \delta^2 - 2\langle y, z_v \rangle}{A^2} \right| < \left| 1.13n + 3.24nm + 2 \cdot 1.8\sqrt{\frac{mn}{m\sqrt{A^2 + 1.13n}}} \right| < 0.02 \]

by Property 1, Lemma 4, and the condition for \( A \). On the other hand, Taylor’s theorem (in the Lagrange remainder form) implies the equation

\[ \sqrt{1 + \varepsilon} = 1 + \frac{\varepsilon}{2} - \frac{\varepsilon^2}{8} + \theta \frac{\varepsilon^3}{15.7} \]

for some \( \theta \in [0, 1] \) if \( |\varepsilon| \leq 0.02 \).

Therefore, \( \text{cost}(X_S, c(X_S)) \leq \sum_{v \in S} A \left( 1 + \frac{\zeta_v}{2A^2} - \frac{\zeta_v^2}{8A^4} + \frac{\zeta_v^3}{15.7A^6} \right) \) and

\[ \text{cost}(X_S, \mu(X_S)) \geq \sum_{v \in S} A \left( 1 + \frac{\zeta_v + \delta^2 - 2\langle y, z_v \rangle}{2A^2} - \frac{\zeta_v^2 + 2\delta^2\zeta_v + \delta^4 - 4\langle y, z_v \rangle^2\delta^2 + 2\langle y, z_v \rangle^2 + 8\langle y, z_v \rangle^3}{8A^4} \right) \]

Since \( \sum_{v \in S} z_v \) is the zero vector, the sums of the terms \( \langle y, z_v \rangle \zeta_v \) are zero. On the other hand, by Properties 4 and 1, the sum of the terms \( \langle y, z_v \rangle \zeta_v \) is greater than \(-1.061n\delta\sqrt{A^2} + 1.13n \). At the same time, by Properties 3 and 1, the sum of the
terms $-\langle y, z \rangle^2$ is greater than $-\frac{m + 1.515}{2} \delta^2(A^2 + 1.13n)$. Therefore, the inequality $\text{cost}(X_S, c(X_S)) \geq \text{cost}(X_S, \mu(X_S))$ implies that

$$\frac{mA^2}{2A^2} = \frac{1.13nmA^2}{4A^4} = \frac{m\delta^4}{8A^2} \geq \frac{1.061mA^2 + 1.13n}{2A^2} \cdot \frac{1}{A^2} \cdot \frac{8mA^3(A^2 + 1.13n)^{3/2}}{15.7A^6} \leq \frac{1.13n^3m}{15.7A^6}.$$

Recall that $\delta < 1.8\sqrt{nm}$ by Lemma 4. Then, taking into account the obvious inequality $(A^2 + 1.13n)^{\alpha} < 1.001A^{2\alpha}$ for $\alpha \leq 3/2$ and $A \geq 100n$, we have

$$\delta \left( \frac{1}{4} - \frac{1.13n}{4A^2} - \frac{1.8^2nm}{8A^2} - \left( \frac{1.001}{4} + \frac{1.517}{4m} \right) - \frac{8 \cdot 1.8\sqrt{nm} \cdot 1.001}{15.7A} \right) < \frac{(1.13n)^3 + 1.061 \cdot 1.001n}{15.7A^4}.$$

But $m \geq 3$, so

$$\delta \left( \frac{0.123n}{4A^2} - \frac{1.13n}{4A^2} - \frac{1.8^2nm}{8A^2} - \frac{1.8\sqrt{nm} \cdot 1.001}{15.7A} \right) < \frac{(1.13n)^3}{15.7A^4} + \frac{1.061 \cdot 1.001n}{2mA}.$$

Since $A$ is sufficiently big, it follows that $\delta < \frac{4.5n}{mA}$. The lemma is proved. \(\square\)

3 The case of $\ell_\infty$ distances

To prove the statements of Theorem 4 for the case of $\ell_\infty$ distances, we construct a similar but much simpler reduction from the Independent Set problem.

Let $G = (V, E)$ be any $n$-vertex undirected graph. Without loss of generality, we will assume that $G$ has no isolated vertices. Fix an arbitrary orientation on its edges: for every edge, choose an endpoint of this edge which it is "outgoing from" and one which it is "incoming to". Next, map each vertex $v \in V$ to the point $x_v \in \mathbb{R}^E$ with the following coordinates: $x_v(e) = 1$ for every edge $e \in E$ outgoing from $v$ and $x_v(e) = -1$ for ones incoming to $v$; all the other coordinates are zero (see Fig. 5). Note that, since $G$ has no isolated vertices, all the vectors $x_v$, $v \in V$, are distinct. Define the instance of the $\ell_\infty$-Median problem with outliers corresponding to the graph $G$ as the set $X = \{x_v \mid v \in V\}$.

Comment 3 Selecting an edge orientation provides that the $\ell_\infty$-distance between the vectors $x_v, x_u$ for adjacent different vertices $u, v \in V$ is 2, while that for non-adjacent ones is 1. It will be used below to justify the proposed reduction.
Estimate the value of the $\ell_\infty$-median cost of the set $X_S = \{x_v \mid v \in S\}$ for an arbitrary $m$-element subset of vertices $S \subseteq V$, where $m \geq 2$. Consider the point $c_S \in \mathbb{R}^E$ for which $c_S(e) = 1/2$ (or $-1/2$) if the edge $e$ is outgoing from $S$ and incoming to $V \setminus S$ (or vice versa), otherwise, $c_S(e) = 0$. Note that, if the set $S$ is independent, then any edge $e \in E$ which is incident to vertices from $S$ is either outgoing from $S$ or incoming to $S$. In this case, the $\ell_\infty$-distance from $c_S$ to each point $x_v$, $v \in S$, is exactly $1/2$, so $\sum_{v \in S} \|x_v - c\|_\infty = m/2$.

On the other hand, for any subset $S \subseteq V$ and any point $c \in \mathbb{R}^E$, we have

$$\sum_{v \in S} \|x_v - c\|_\infty = \frac{1}{2(m-1)} \sum_{v \in S} \sum_{u \in S \setminus \{v\}} \left(\|x_v - c\|_\infty + \|x_u - c\|_\infty\right) \geq \frac{1}{2(m-1)} \sum_{v \in S} \sum_{u \in S \setminus \{v\}} \|x_v - x_u\|_\infty \geq \frac{m^2 - m}{2(m-1)} = m/2.$$

Moreover, by Comment 3, if the set $S$ contains at least two adjacent vertices $a$ and $b$, then

$$\sum_{v \in S} \|x_v - c\|_\infty = \|x_a - c\|_\infty + \|x_b - c\|_\infty + \sum_{v \in S \setminus \{a, b\}} \|x_v - c\|_\infty \geq \|x_a - x_b\|_\infty + \sum_{v \in S \setminus \{a, b\}} \|x_v - c\|_\infty \geq 2 + (m - 2)/2 = m/2 + 1.$$

It follows that the $\ell_\infty$-median cost of the set $X_S$ is $m/2$ if the set $S$ is independent and is at least $m/2 + 1$ otherwise. Thus, the Independent Set problem is reduced to $\ell_\infty$-Median with outliers and we obtain the following theorem:

**Theorem 2** The $\ell_\infty$-Median problem with outliers is strongly NP-hard, does not admit approximation schemes FPTAS unless $P=NP$, and is $W[1]$-hard with respect to the parameter $m$.

**References**

1. Agarwal P.K., Har-Peled S., Varadarajan K.R. Geometric approximation via coresets // Combinatorial and Computational Geometry, MSRI Publications 52, 1–30 (2005) [http://library.msri.org/books/Book52/files/01agar.pdf]
2. Charikar M., Khuller S., Mount D.M., Narasimhan G. Algorithms for facility location problems with outliers // Proc. 12th ACM-SIAM Symposium on Discrete Algorithms (SODA 2001), 642–651 (2001) [https://doi.acm.org/doi/10.5555/365411.365555]
3. Chen K. A constant factor approximation algorithm for k-median clustering with outliers // Proc. 19th ACM-SIAM Symposium on Discrete Algorithms (SODA 2008), 826–835 (2008) [https://doi.acm.org/doi/10.5555/1347032.1347173]
4. Cohen M.B., Lee Y.T., Miller G., Pachocki J., Sidford A. Geometric median in nearly linear time // [arXiv:1606.05225] [cs.DS] (2016) [https://arxiv.org/abs/1606.05225]
5. Cohen-Addad V., Feldmann A.E., Saulpic D. Near-linear time approximations schemes for clustering in doubling metrics // Proc. 60th Symposium on Foundations of Computer Science (FOCS 2019), 540–559 (2019) [https://doi.org/10.1109/FOCS.2019.00041]
6. Downey R.G., Fellows M.R. Fundamentals of Parameterized Complexity. Texts in Computer Science. Springer-Verlag, London (2013) [https://doi.org/10.1007/978-1-4471-5559-1]
7. Edelsbrunner H., O’Rourke J., Seidel R. Constructing arrangements of lines and hyperplanes with applications // SIAM J. Computing 15(2), 341–363 (1986) https://doi.org/10.1137/0215024
8. Edelsbrunner H., Seidel R., Sharir M. On the zone theorem for hyperplane arrangements // SIAM J. Computing 22(2), 418–429 (1993) https://doi.org/10.1137/0222031
9. ElGindy H., Keil J.M. Efficient algorithms for the capacitated 1-median problem // ORSA J. Computing 4(4), 418–425 (1992) https://doi.org/10.1287/ijoc.4.4.418
10. Feige U., Karpinski M., Langberg M. A note on approximating max-bisection on regular graphs // Inf. Proc. Letters 79(4), 181–188 (2001) https://doi.org/10.1016/S0020-0190(00)00189-7
11. Inaba M., Katoh N., Imai H. Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering // Proc. 10th ACM Symposium on Computational Geometry, 332–339 (1994) https://doi.org/10.1145/177424.178042
12. Kel’manov A.V., Pyatkin A.V. NP-completeness of some problems of choosing a vector subset // J. Appl. Industr. Math. 5(3), 352–357 (2011) https://doi.org/10.1134/S1990478911030089
13. Krishnawamy R., Li S., Sandeep S. Constant approximation for k-median and k-means with outliers via iterative rounding // Proc. 50th ACM Symposium on Theory of Computing (STOC 2018), 646–659 (2018) https://doi.org/10.1145/3188745.3188852
14. Shenmaier V.V. An approximation scheme for a problem of search for a vector subset // J. Appl. Industr. Math. 6(3), 381–386 (2012) https://doi.org/10.1134/S1990478912030131
15. Shenmaier V.V. The problem of a minimal ball enclosing k points // J. Appl. Industr. Math. 7(3), 444–448 (2013) https://doi.org/10.1134/S1990478913030156
16. Shenmaier V.V. Complexity and approximation of the smallest k−enclosing ball problem // European J. Comb. 48, 81–87 (2015) https://doi.org/10.1016/j.ejc.2015.02.011
17. Shenmaier V.V. A structural theorem for center-based clustering in high-dimensional Euclidean space // Proc. 5th Conference on Machine Learning, Optimization, and Data Science (LOD 2019), LNCS 11943, 284–295 (2019) https://doi.org/10.1007/978-3-030-37599-7_24
18. Shenmaier V.V. Some estimates on the discretization of geometric center-based problems in high dimensions // Proc. 19th Conf. Mathematical Optimization Theory and Operations Research (MOTOR 2020), CCIS 1275, 88–101 (2020) https://doi.org/10.1007/978-3-030-58657-7_10
19. Shenmaier V.V. Polynomial approximate discretization of geometric centers in high-dimensional Euclidean space // Adv. Data Anal. Classif. https://doi.org/10.1007/s11634-021-00481-4