A $(1+\varepsilon)$-approximation for the minimum enclosing ball problem in $\mathbb{R}^d$ *

Sang-Sub Kim and Barbara Schwarzwald

Institute of Computer Science, University Bonn
schwarzwald@uni-bonn.de

Abstract

Given a set of points P in $\mathbb{R}^d$ for an arbitrary d, the 1-center problem or minimum enclosing ball problem (MEB) asks to find a ball B^* of minimum radius r^* which covers all of P. Kumar et. al. [5] and Bădoiu and Clarkson [1] simultaneously developed core-set based $(1+\varepsilon)$-approximation algorithms. While Kumar et al. achieve a slightly better theoretical runtime of $O(nd/\varepsilon + 1/\varepsilon^{4.5} \log 1/\varepsilon)$, Bădoiu and Clarkson have a stricter bound of $\lceil 2/\varepsilon \rceil$ on the size of their core-set, which strongly affects run-time constants.

We give a gradient-descent based algorithm running in time $O(nd/\varepsilon)$ based on a geometric observation that was used first for a 2-center streaming algorithm by Kim and Ahn [4]. Our approach can be extended to the k-center problem to obtain a $(1+\varepsilon)$-approximation in time $O(nd\,k\,2^{k/\varepsilon})$.

1 Introduction

Given a set of points $P \subset \mathbb{R}^d$, the *minimum enclosing ball problem* (MEB), also known as the 1-center problem, asks to find a ball B^* of minimum radius r^* containing all of P and is an important subproblem in clustering. While it can be solved in worst-case linear time for fixed d [6], the dependence on d is exponential and hence not practical for high dimensional real-world applications. However, Bădoiu et al. [3] presented a $(1+\varepsilon)$-approximation algorithm for arbitrary d running in time $O(nd/\varepsilon^2 + 1/\varepsilon^{10} \log 1/\varepsilon)$ using core-sets of size at most $1/\varepsilon^2$ independent of d. An ε-core-set is a subset $S \subset P$, such that a ball of radius $(1+\varepsilon)r^*$ around the center of a minimum enclosing ball of S covers P. Their algorithm can be extended to approximate the k-center problem, but the running time is then exponential in k and the size of the core-set; so having a tight bound on the size of the core-set is paramount. In fact, no polynomial time approximation scheme for the k-center problem in high dimensions can exist if P $\neq$ NP, see [7].

Kumar et al. [5] improved these results to finding ε-core-sets of size $O(1/\varepsilon)$ in time $O(nd/\varepsilon^2 + 1/\varepsilon^{4.5} \log 1/\varepsilon)$. Indepedently Bădoiu and Clarkson [1] achieved an algorithm with a similar running time of $O(nd/\varepsilon + 1/\varepsilon^5)$ while having a stricter bound of $\lceil 2/\varepsilon \rceil$ on the size of their core-set, which significantly affects run-time especially when extending to the k-center problem. Bădoiu and Clarkson [1] also gave a simple gradient-descent algorithm obtaining a $(1+\varepsilon)$-approximation in time $O(nd/\varepsilon^2)$ and later showed that a tight bound of $\lceil 1/\varepsilon \rceil$ on the size of ε-core-sets exists, see [2]. The gradient-descent algorithm has the advantage of not computing minimum enclosing balls for several subsets of P of size $O(1/\varepsilon)$ which improves the constants involved in the calculation of each step and simplifies implementation. Bădoiu and Clarkson [2] also performed runtime experiments on both the gradient-descent and the different core-set-based algorithms which results showed that the gradient-descent algorithm is competitive in reality, as it converges significantly faster than its theoretical bound suggests.

* This work has been supported by DFG grant Kl 655/19 as part of a DACH project.

36th European Workshop on Computational Geometry, Würzburg, Germany, March 16–18, 2020.
This is an extended abstract of a presentation given at EuroCG'20. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

We combine the analysis of these core-set-based algorithms with the ideas of the gradient-descent algorithm and extend structural observations made by Kim and Ahn [4] for the Euclidean 2-center problem in a streaming model to obtain a new efficient gradient-descent algorithm that converges to a $(1+\varepsilon)$-approximation in time $O(nd/\varepsilon)$. It can be applied to the k-center problem in a similar fashion as the core-set based algorithms. Hence, it gives an alternative $(1+\varepsilon)$-approximation in time $O(nd\,k\,2^{k/\varepsilon})$ for that problem without the use of core-sets and, possibly, a faster algorithm on real-world data.

2 An Algorithm for the Euclidean 1-center problem

In this section we present an algorithm for the Euclidean 1-center problem for high dimensions and show the following theorem:

▶ **Theorem 2.1.** *Given a set $P \in \mathbb{R}^d$, one can compute a $(1+\varepsilon)$-approximation of the minimum enclosing ball in time $O(nd/\varepsilon)$ with the gradient-descent-algorithm* GradientMEB*.*

Let $P \subset \mathbb{R}^d$ for any $d \geq 2$ be a set of n points. Let $B(c,r)$ denote a ball of radius r centered at c and let $r(B)$ and $c(B)$, denote the radius and center of a ball B, respectively. We denote by pq the straight line segment between two points p and q and by $|pq|$ the length of pq. Finally, we denote the boundary of a closed set A by ∂A.

Let $B^* = B(c^*, r^*)$ be the optimal solution of the 1-center problem for a set P. The core idea of the algorithm is a follows. We will start with an arbitrary point p_1 from P as a starting center m_1. For any center m_i constructed, the radius r_i necessary to cover all of P with a ball $B(m_i, r_i)$ is defined by the farthest point in P from m_i. Therefore, in every subsequent step we pick that farthest point as p_{i+1} and construct a new center m_{i+1} on the line segment between p_{i+1} and m_i to reduce r_i. We will use a central structural property proven in Lemma 2.2 to show how to construct $m_i = m(p_{i+1}, m_i)$ in such a way that we can give a bound on its distance $|m_i c^*|$ to the optimal center c^* decreasing with every step i. The exact definition of $m(p_{i+1}, m_i)$ will hence be given after that Lemma.

Algorithm 1 GradientMEB

Input: Set of points $P \subset \mathbb{R}^d$.
Output: A center c such that $B(c, (1+\varepsilon)r^*)$ covers P

```
p_1 ← arbitrary point from P
m_1 ← p_1
bestRadius ← ∞
for i = 1 to ⌊2/ε⌋ do
    p_{i+1} ← farthest point from m_i in P
    if |m_i p_{i+1}| < bestRadius then
        bestCenter ← m_{i−1}
        bestRadius ← |m_{i−1} p_i|
    m_{i+1} ← m(p_{i+1}, m_i)
return bestCenter
```

We will start with the proof of the central structural property and the construction of $m(p_{i+1}, m_i)$ and then show that after at most $k = \lfloor 2/\varepsilon \rfloor$ such steps, $|m_k c^*| < \varepsilon r^*$ and hence $B(m_k, (1+\varepsilon)r^*)$ covers all of P.

Both together will proof the correctness of GradientMEB. As the algorithm runs for $\lfloor 2/\varepsilon \rfloor$ rounds, finding p_i each round takes $O(nd)$ and the computation of m_i takes $O(d)$, this will also proof Theorem 2.1.

▶ **Lemma 2.2.** *Given two d-dimensional balls B and B' with radii r and r' around the same center point c with $r > r'$ with $d \geq 2$. Let $p \in \partial B$ and $p' \in \partial B'$ with $|pp'| = l \geq r$. Let B'' be the d-dimensional ball centered around c that is tangential to pp'. We denote that tangential point with m and the distances $|p'm|$ with l_1 and $|pm|$ with l_2, so $l = l_1 + l_2$. Consider any line segment p_1p_2 with $|p_1p_2| > l$, $p_1 \in B'$ and $p_2 \in B$. Then any point m^* on p_1p_2 with $|p_1m^*| \geq l_1$ and $|p_2m^*| \geq l_2$ lies inside B''.*

Proof. For $d = 2$ we first show that $|p_1p_2|$ intersects B'' at all. It is clear that we can rotate and reflex pp' without changing B'' as long as its length l stays the same. Hence we can assume without loss of generality, that p', c and p_1 are collinear with $p_1 \in cp'$. Then p_2 must lie in $B \setminus B(p', l)$ which means p_1p_2 intersects B'' as illustrated in Figure 1.

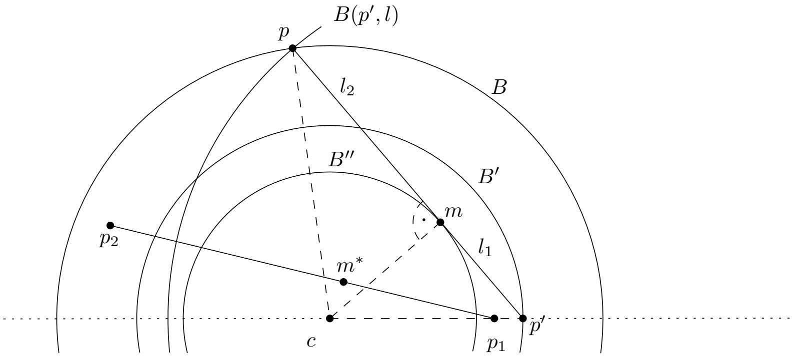

Figure 1 Construction of m and B''. p_1p_2 must intersect B'' if $|p_1p_2| \geq l = |pp'|$.

Now consider a point m^* on p_1p_2 with $|p_1m^*| \geq l_1$ and $|p_2m^*| \geq l_2$. Assume m^* is not contained in B''. Then either $p_1m^* \cap B'' = \emptyset$ and $p_2m^* \cap B'' \neq \emptyset$ or the other way around as p_1p_2 intersects B'' somewhere.

Let's first assume, $p_1m^* \cap B'' = \emptyset$ and $p_2m^* \cap B'' \neq \emptyset$. In this case $p_1 \in B' \setminus B''$. Let p_t be a point on the boundary of B'' such that p_1p_t is tangential to B'' and m^* lies within the triangle p_1p_tc. Clearly $|p_1p_t| > |p_1m^*|$. But by construction of B'', $|p'm| \geq |p_1p_t|$, which contradicts $|p'm| = l_1 \leq |p_1m^*|$. This is illustrated in Figure 2 (assuming without loss of generality that p' is collinear with p_1p_2, as this does not not affect the construction of B'').

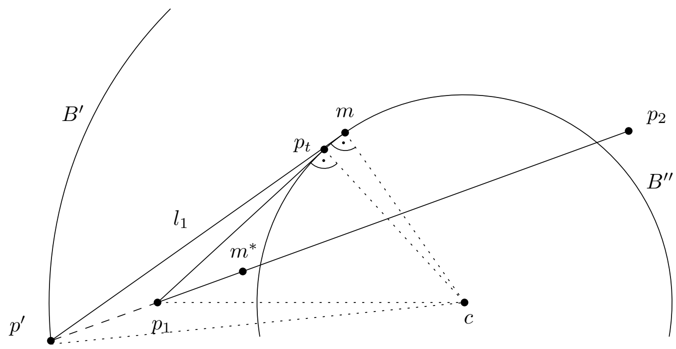

Figure 2 Assume $p_1m^* \cap B'' = \emptyset$. $|p_1m^*| < |p_1p_t| \leq |p'm|$. This contradicts $|p_1m^*| \geq l_1 = |p'm|$.

The other case can be shown equivalently by switching p_1 and p_2 and replacing p' and B' and l_1 with p and B and l_2, respectively.

For $d > 2$ we can show the lemma by choosing the 2-dimensional plane passing through c and the line segment p_1p_2 and then follow the same arguments as for $d = 2$. ◀

Note, that the point m' on p_1p_2 with $\frac{|m'p_1|}{|p_2p_1|} = \frac{l_1}{l} = \frac{|mp'|}{|pp'|}$ fulfils $|m'p_1| \geq l_1$ and $|m'p_2| \geq l_2$. The following corollary follows from that observation, Lemma 2.2 and the Pythagorean theorem and defines a way to calculate that point without actually knowing r^*.

▶ **Corollary 2.3.** *Let B and B' be two balls in $\mathbb{R}^d$ with $c(B) = c(B')$ and radii $r(B) = r^*$ and $r(B') = r' = \delta r^*$ for some $0 < \delta \leq 1$. Then the line segment pp' between any two points $p \in \partial B$ and $p' \in \partial B'$ with distance $|pp'| = l = (1+\varepsilon)r^*$ is tangential to $B'' = B(c, r_m)$ with*

$$r_m \leq r^* \sqrt{1 - \left(\frac{1 + (1+\varepsilon)^2 - \delta^2}{2(1+\varepsilon)}\right)^2} \tag{1}$$

at a point m^* with $l_1 := |m^*p'|$.

Let $p_1 \in B'$ and $p_2 \in B$ with $|p_1p_2| \geq l = (1+\varepsilon)r^*$.

Then

$$m(p_1, p_2) := p_1 + (p_2 - p_1)\frac{l_1}{l} = p_2 + (p_1 - p_2)\frac{\delta^2 + (1+\varepsilon)^2 - 1}{2(1+\varepsilon)^2} \tag{2}$$

lies in the ball $B(c, r_m)$ and can be calculated independent of r^*, only knowing p_1, p_2, δ and ε.

One can extend this definition to a sequence $m_1, m_2, \ldots, m_k$ based on a sequence of points $p_1, \ldots, p_k$ with $p_i \in P$ with $|p_jm_{j-1}| \geq (1+\varepsilon)r^*$ for all $i \geq j > 1$.

Let

$$\delta_i := \begin{cases} 1, & \text{if } i = 1. \\ \sqrt{1 - \left(\frac{1+(1+\varepsilon)^2 - \delta_{i-1}^2}{2(1+\varepsilon)}\right)^2}, & \text{otherwise.} \end{cases} \tag{3}$$

and

$$m_i := \begin{cases} p_1, & \text{if } i = 1. \\ m(p_i, m_{i-1}) = m_{i-1} + (p_i - m_{i-1})\frac{\delta_{i-1}^2 + (1+\varepsilon)^2 - 1}{2(1+\varepsilon)^2}, & \text{otherwise.} \end{cases} \tag{4}$$

As all points in P lie in the ball $B(c^*, r^*)$, it follows by induction from Corollary 2.3 that m_i lies in the ball $B(c^*, \delta_i r^*)$.

GradientMEB starts with an arbitrary point p_1 from P as m_i and always uses the farthest point from m_{i-1} in P as p_i. That way, at each round we either have $|m_ip_{i+1}| > (1+\varepsilon)r^*$ or m_i is already a $(1+\varepsilon)$-approximation. As we do not know, which of both holds at any round, we just return the best m_i out of all rounds.

It remains to prove, that any sequence of points with $|p_jm_{j-1}| \geq (1+\varepsilon)r^*$ for all $i \geq j > 1$ contains at most $k \leq \lfloor 2/\varepsilon \rfloor$ points before $m_i \in B(c^*, \varepsilon r^*)$.

If at step i, $|m_{i-1}p_i| = (1+\varepsilon)r^*$, $p_i \in \partial B^*$ and $m_{i-1} \in \partial B(c^*, \delta_{i-1}r^*)$, then $m_i \in \partial B(c^*, \delta_i r^*)$ by Lemma 2.2. In that case, $m_i \in B(c^*, \varepsilon r^*)$ if and only if $\delta_i < \varepsilon$. As this is the worst-case, we can assume we were given our sequence of points $p_i \in P$ by an adversary, always fulfilling $|m_{i-1}p_i| = (1+\varepsilon)r^*$ and $p_i \in \partial B^*$, which gives $m_{i-1} \in \partial B(c^*, \delta_{i-1})$ by induction.

We use a similar proof as [1] for their core-set based algorithm. For this we consider the line segments $a_i = m_{i-1}m_i$ with $|a_i| = \alpha_i(1+\varepsilon)r^*$ and $b_i = m_ip_i$ with $|b_i| = \beta_i(1+\varepsilon)r^*$ that together form $m_{i-1}p_i$ as illustrated in Figure 3.

As m_i converges towards c^* with increasing i, β_i increases and α_i decreases. However, β_i can be at most $1/(1+\varepsilon)$ by construction as b_i forms a right-angled triangle with c^*p_i as the hypotenuse, so $\beta_i(1+\varepsilon)r^* = |b_i| < |c^*p_i| = r^*$. We will now show a lower bound on β_i and prove that it exceeds $1/(1+\varepsilon)$ for $i \geq 2/\varepsilon - 1$. In that case, there does not exist a point p_i with $|m_{i-1}p_i| = (1+\varepsilon)r^*$ and $p_i \in \partial B^*$ that our adversary could have given us. This

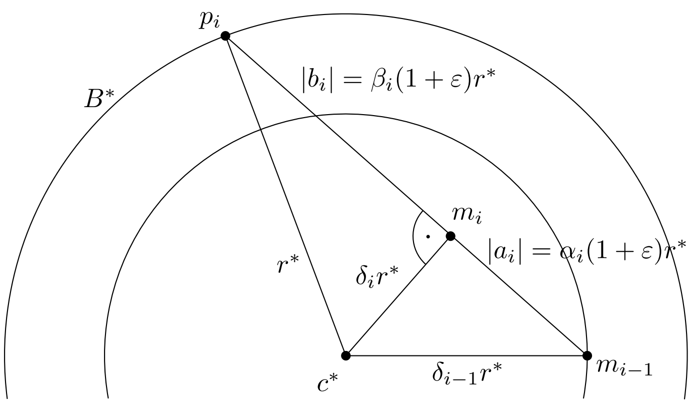

Figure 3 Construction of m_i based on m_{i-1} and p_i.

can only happen if the intersection of ∂B^* and $\partial B(m_{i-1}, (1+\varepsilon)r^*)$ is empty, and therefore, $\partial B(m_{i-1}, (1+\varepsilon)r^*)$ covers B^*.

▶ **Lemma 2.4.** *$\beta_i \geq 1/(1+\varepsilon)$ for $i \geq 2/\varepsilon - 1 > \lfloor 2/\varepsilon \rfloor$.*

Proof. By our definition and our worst-case assumption

$$|b_i| = (1+\varepsilon)r^* - |a_i| \qquad \Rightarrow \qquad \beta_i = 1 - \alpha_i. \tag{5}$$

In addition, by the construction of m_i as illustrated in Figure 3 and the Pythagorean theorem, it holds

$$\begin{aligned} \beta_i^2(1+\varepsilon)^2 &= 1^2 - \delta_i^2 \\ &= 1 - (\delta_{i-1}^2 - \alpha_i^2(1+\varepsilon)^2) \\ &= 1 - ((1-\beta_{i-1}^2(1+\varepsilon)^2) - \alpha_i^2(1+\varepsilon)^2) \\ \Rightarrow \qquad \beta_i^2 &= \alpha_i^2 + \beta_{i-1}^2. \end{aligned} \tag{6}$$

Combining these two equations we get

$$\begin{aligned} 1 - \alpha_i &= \sqrt{\beta_{i-1}^2 + \alpha_i^2} \\ 1 - 2\alpha_i + \alpha_i^2 &= \beta_{i-1}^2 + \alpha_i^2 \\ \Rightarrow \qquad \alpha_i &= \frac{1-\beta_{i-1}^2}{2}. \end{aligned} \tag{7}$$

Applying Equation 5 again we obtain the recurrence

$$\beta_i = \frac{1+\beta_{i-1}^2}{2}. \tag{8}$$

If we substitute $\gamma_i = \frac{1}{1-\beta_i} \Leftrightarrow \beta_i = \frac{\gamma_i - 1}{\gamma_i}$ in Equation 8, we get

$$\gamma_i = \frac{\gamma_{i-1}}{1 - 1/(2\gamma_{i-1})} = \gamma_{i-1}(1 + \frac{1}{2\gamma_{i-1}} + \frac{1}{4\gamma_{i-1}^2} + \cdots) \geq \gamma_{i-1} + \frac{1}{2}. \tag{9}$$

As we have $\beta_1 = 1/2$ and hence $\gamma_1 = 2$, we know $\gamma_i \geq (3+i)/2$ and hence $\beta_i \geq 1 - \frac{2}{3+i}$. To obtain $\beta_i \geq 1/(1+\varepsilon)$ it suffices to have $i \geq 2/\varepsilon - 1$. ◀

This also concludes the proof of Theorem 2.1.

2.1 Extension to the 2-center problem

We employ a strategy quite similar to the approach in [1]. We aim to construct two series of centers $m_{1,j}$ and $m_{2,k}$ based on two series of points from the two optimal balls B_1^* and B_2^*.

We start with an arbitrary point p_1 and set $m_{1,1} = p_1$ as we can assume $p_1 \in B_1^*$ without loss of generality. In every further step, we pick a point p_i farthest from the two current centers $m_{1,j}$ and $m_{2,k}$. As long as we have no center for B_2^*, we pick the point furthest from $m_{1,j}$. We then employ a guessing oracle that tells us whether p_i belongs to B_1^* or B_2^*. Depending on its answer, we add the point to the sequence for the respective ball then calculate a new center $m_{1,j+1}$ or $m_{2,k+1}$ as in our 1-center algorithm.

After at most $2\lfloor 2/\varepsilon \rfloor$ picks, we obtain a $(1+\varepsilon)$-approximation. As we do not have a guessing oracle, we just exhaust all possible guesses and return the best solution encountered, which results in a running time of $O(nd\,2^{1/\varepsilon})$.

2.2 Extension to the k-center problem for $k > 2$

The k-center algorithm is a straight-forwarded extension of the 2-center algorithm. As we need to guess at most $k\lfloor 2/\varepsilon \rfloor$ points to obtain $(1+\varepsilon)$-approximation and have to exhaust k possibilities each, our algorithm runs in time $O(nd\,k\,2^{k/\varepsilon})$.

3 Conclusion

We provided a new efficient gradient-descent $(1+\varepsilon)$ approximation algorithm for MEB in arbitrary dimensions running in time $O(nd/\varepsilon)$, which is strictly better than previous core-set based approaches with running times $O(nd/\varepsilon + 1/\varepsilon^{4.5} \log 1/\varepsilon)$ as long as $nd \in o(1/\varepsilon^{3.5} \log 1/\varepsilon)$. Like the core-set based algorithms it can be extended to the k-center problem with a running time of $O(nd\,k\,2^{k/\varepsilon})$, which makes the gradient-descent based algorithm theoretically equivalent to the core-set based approached with possibly better run-time constants by combining similar analysis with new geometric observations.

References

1. Mihai Bădoiu and Kenneth L. Clarkson. Smaller core-sets for balls. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '03, page 801–802, USA, 2003. Society for Industrial and Applied Mathematics. doi:10.5555/644108.644240.
2. Mihai Bădoiu and Kenneth L. Clarkson. Optimal core-sets for balls. *Computational Geometry*, 40(1):14 – 22, 2008. doi:10.1016/j.comgeo.2007.04.002.
3. Mihai Bădoiu, Sariel Har-Peled, and Piotr Indyk. Approximate clustering via core-sets. In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02, page 250–257, New York, NY, USA, 2002. Association for Computing Machinery. doi:10.1145/509907.509947.
4. Sang-Sub Kim and Hee-Kap Ahn. An improved data stream algorithm for clustering. In Alberto Pardo and Alfredo Viola, editors, *LATIN 2014: Theoretical Informatics*, pages 273–284, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. doi:10.1007/978-3-642-54423-1_24.
5. Piyush Kumar, Joseph S. B. Mitchell, and E. Alper Yildirim. Approximate minimum enclosing balls in high dimensions using core-sets. *J. Exp. Algorithmics*, 8:1.1–es, December 2004. doi:10.1145/996546.996548.

6 Nimrod. Megiddo. "linear-time algorithms for linear programming in r^3 and related problems". *SIAM Journal on Computing*, 12(4):759–776, 1983. doi:10.1137/0212052.

7 Nimrod Megiddo. On the complexity of some geometric problems in unbounded dimension. *Journal of Symbolic Computation*, 10(3):327 – 334, 1990. doi:10.1016/S0747-7171(08)80067-3.