Kinetic Euclidean 2-centers in the black-box model

Kinetic Euclidean 2-centers in the Black-Box Model*

Mark de Berg† Marcel Roeloffzen† Bettina Speckmann†

Abstract

We study the 2-center problem for moving points in the plane. Given a set P of n points, the Euclidean 2-center problem asks for two congruent disks of minimum size that together cover P. Our methods work in the black-box KDS model, where we receive the locations of the points at regular time steps and we know an upper bound $d_{\max}$ on the maximum displacement of any point within one time step.

We show how to maintain a $(1+\varepsilon)$-approximation of the Euclidean 2-center in amortized sub-linear time per time step, under certain assumptions on the distribution of the point set P. In many cases—namely when the distance between the centers of the disks is relatively large or relatively small—the solution we maintain is actually optimal.

1 Introduction

The clustering problem is to partition a given set of objects into clusters, that is, into subsets consisting of similar objects. These objects are often (represented by) points in some 2- or higher dimensional space, and the similarity between points corresponds to the distance between them. We are interested in a setting with two clusters of points in the plane. Given a set P of n points, the Euclidean 2-center problem—or 2-center problem for short—asks for two congruent disks of minimum size that together cover P. The 2-center problem can also be interpreted as a facility-location problem, where the goal is to place two facilities such that the distance from any client in P to its nearest facility is minimized.

The 2-center problem and the more general k-center problem—which asks for k disks to cover P—have been studied extensively since their introduction by Sylvester [20] in 1857. Closely related is the rectilinear k-center problem which asks for k congruent squares to cover the point set. Both the Euclidean and the rectilinear k-center problem are NP-hard [14] when k is part of of the input, but polynomial-time solutions are possible when k is a constant. The rectilinear k-center problem can be solved quite efficiently for small k. For $k = 2, 3$ the optimal rectilinear k-center can be computed in $O(n)$ time [8, 13, 19] and for $k = 4, 5$ in $O(n \log n)$ time [15, 16]. In contrast, no sub-quadratic algorithm was known for the Euclidean 2-center for many years, until Sharir [18] developed an $O(n \log^9 n)$ time algorithm. The currently best solution takes $O(n \log^2 n (\log \log n)^2)$ time [6]. Other results include an $\Omega(n \log n)$ lower bound for $k = 2$ [17] and algorithms that compute a $(1+\varepsilon)$-approximation of the k-center [1, 2]. For the 2-center problem in $\mathbb{R}^2$ such an $(1+\varepsilon)$-approximation can be computed in $O(n) + (2/\varepsilon)^{O(1)}$ time.

The kinetic 2-center problem. The results mentioned so far are for static point sets, but also *kinetic* versions of the 2-center problem have been studied. Here we want to maintain the optimal 2-centers as the points in P move. Unfortunately, even under the restriction that the speed of the points in P is bounded by a given value $v_{\max}$, the speed of the centers cannot be bounded if one maintains the optimal 2-center. For mobile facility location this is undesirable as the centers represent moving facilities which often have a bounded speed as well. Hence, Durocher and Kirkpatrick [9] describe a general strategy for maintaining an approximate 2-center in such a way that the speed of the disk centers is bounded. One variant of their strategy achieves an approximation ratio of $8/\pi \approx 2.55$ while the maximum speed of the disk centers is bounded by $(8/\pi + 1)\, v_{\max} \approx 3.55\, v_{\max}$. Maintaining this approximation is done using the kinetic data structures (KDS) framework by Basch *et al.* [3]. When viewed as a clustering problem the centers have no explicit meaning and no bound on their speed is necessary. For this case efficient KDSs for the discrete version of the k-center problem, where the disk centers must be chosen from the input set P, have been given [7, 11].

Problem statement. The previous results on the kinetic 2-center problem [7, 9, 11] use the standard KDS model, where the trajectories of the points are known in advance. However, in many applications the trajectories are not known and the standard KDS framework cannot be used. Hence, we study the kinetic 2-center problem in the so-called *black-box model* [4, 5, 12]. In the black-box model the locations of the points are reported at regular time steps $t_0, t_1, \cdots$, and we are given a value $d_{\max}$ such that any point can move at most distance $d_{\max}$ from one time step to the next. Thus, when $p(t)$ denotes the location

*M.R. and B.S. were supported by the Netherlands' Organisation for Scientific Research (NWO) under project no. 600.065.120 and 639.022.707, respectively.

†Department of Computing Science, TU Eindhoven, P.O. Box 513, 5600 MB Eindhoven, the Netherlands. {*mdberg,mroeloff,speckman*}@*win.tue.nl*.

of point p at time t, we have $\operatorname{dist}(p(t_i), p(t_{i+1})) \leqslant d_{\max}$ for all $p \in P$ and every time step t_i. We want to maintain the 2-center of the set $P(t) := \{p(t) : p \in P\}$ at every time step while using coherence to speed up the computations. This is not possible without restricting the relation of the maximum displacement $d_{\max}$ and the distribution of the point set P: if all points lie within distance $d_{\max}$ from each other then the distribution at time $t+1$ need not have any relation to the distribution at time t, and so there is no coherence that can be used. Following our previous papers [4, 5] we assume the following.

> ***Displacement Assumption**: There is a maximum displacement $d_{\max}$ and constant k such that for each point $p \in P$ and any time step t_i we have*
>
> - $\operatorname{dist}(p(t_i), p(t_{i+1})) \leqslant d_{\max}$, *and*
> - *there are at most k other points from P within distance $d_{\max}$ from $p(t_i)$.*

Under this assumption, we formulate our bounds in terms of the so-called *k-spread* [10] of P, which is defined as follows. Let $\operatorname{mindist}_k(P) := \min_{p \in P} \operatorname{dist}(p, \operatorname{NN}_k(p, P))$ denote the smallest distance from any point $p \in P$ to its k-nearest neighbor $\operatorname{NN}_k(p, P)$. Then the k-spread Δ_k of P is defined as $\Delta_k(P) := \operatorname{diam}(P)/\operatorname{mindist}_k(P)$. The 1-spread is simply the regular spread of a point set. We use the k-spread instead of regular spread, since two points may pass by each other at a very close distance, causing a small value for $\operatorname{mindist}_1(P)$ and, consequently, a high spread. It is much less likely that k points are very close simultaneously, and so $\operatorname{mindist}_k(P)$ tends to not be very small. For a good k-spread we also need the diameter not to be too large. This is somewhat unnatural for the 2-center problem: when the two clusters are far apart, the k-spread may become very large even though within each cluster, the points are very evenly distributed. Hence we introduce the so-called *$(2,k)$-spread* $\Delta_{2,k}(P)$:

$$\Delta_{2,k}(P) := \min_{P_1, P_2} \max(\Delta_k(P_1), \Delta_k(P_2)),$$

where the minimum is taken over all possible partitions of P into two subsets P_1, P_2. (The partition defining $\Delta_{2,k}(P)$ does not need to be the same as the partition defining the optimal clustering in the 2-center problem, but since $\Delta_{2,k}(P)$ is the minimum over all partitions this can lead only to a better $(2,k)$-spread.) We express our results using the $(2,k)$-spread of P, that is, using the maximum value of $\Delta_{2,k}(P(t_i))$ over all time steps t_i, which we abbreviate as $\Delta_{2,k} := \max_{t_i} \Delta_{2,k}(P(t_i))$.

Results and Organization. We study the kinetic Euclidean 2-center problem from a clustering point of view: without restrictions on the speed of the centers.

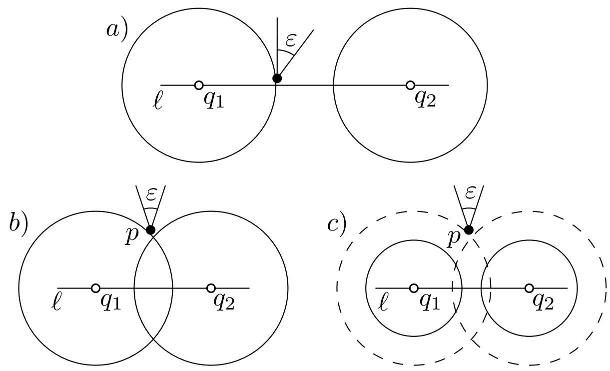

Figure 1: When *a*) $\operatorname{dist}(q_1, q_2) \geqslant 2r + \varepsilon r$ or *b*) $\operatorname{dist}(q_1, q_2) \leqslant 2r - \varepsilon r/2$ the 2-center disks that cover P_ε also cover P, and *c*) otherwise blowing up the disks by a factor $1+\varepsilon$ ensures they cover P.

We investigate the Euclidean 2-center problem and show how to maintain a $(1+\varepsilon)$-approximation, for any $0 < \varepsilon \leqslant \pi/4$ in $O((k/\varepsilon^3)\Delta_{2,k} \log^3 n (\log\log n)^2)$ amortized time. In many cases—when the distance between the centers of the disks is relatively large or small—the solution we maintain is optimal.

2 The Euclidean 2-center

The Euclidean 2-center problem asks for two congruent disks of minimum size that together cover P. Our global strategy to maintain the Euclidean 2-center kinetically is to find a subset $Q \subseteq P$ containing points that are in some sense on the outside of P. We then compute the optimal 2-center for Q and show that it is an approximation of the 2-center of P. Maintaining the approximate 2-center can then be done by maintaining Q. First we define what exactly it means for a point to be on the outside of the point set.

Define a point $p \in P$ to be *ε-interesting* if there is a wedge $W_\varepsilon(p)$ with apex p and opening angle ε such that $W_\varepsilon(p)$ does not contain any other points of P. Let P_ε denote the set of ε-interesting points in P. We show that it suffices to consider the points in P_ε to get (an approximation of) an optimal solution to the 2-center problem on P. In the following we use $\operatorname{disk}(q, r)$ to denote the disk of radius r centered at q.

Lemma 1 *Let $\operatorname{disk}(q_1, r)$ and $\operatorname{disk}(q_2, r)$ be the two disks of an optimal solution for the Euclidean 2-center problem on P_ε, for some $\varepsilon < \pi/4$. If $\operatorname{dist}(q_1, q_2) \leqslant 2r - \varepsilon r/2$ or $\operatorname{dist}(q_1, q_2) \geqslant 2r + \varepsilon r$ then $\operatorname{disk}(q_1, r)$ and $\operatorname{disk}(q_2, r)$ are an optimal solution for the 2-center problem on P; otherwise $\operatorname{disk}(q_1, r + \varepsilon r)$ and $\operatorname{disk}(q_2, r + \varepsilon r)$ are a $(1+\varepsilon)$-approximation for the 2-center problem on P.*

Proof (sketch). First consider the case $\operatorname{dist}(q_1, q_2) \leqslant 2r - \varepsilon r/2$. Since $P_\varepsilon \subseteq P$, the disks in an optimal solution for P cannot have radius smaller than r. Hence, it suffices to prove that $\mathrm{disk}(q_1, r) \cup \mathrm{disk}(q_2, r)$ covers P. Suppose for a contradiction there is an uncovered point in P. Assume without loss of generality that the line ℓ through q_1 and q_2 is horizontal, and let p be the highest uncovered point above ℓ, see Figure 1. (If all uncovered points lie below ℓ we can apply a similar argument to the lowest uncovered point.) The condition $\mathrm{dist}(q_1, q_2) \leqslant 2r - \varepsilon r/2$ implies that the "vertical" wedge $W_\varepsilon(p)$ does not intersect $\mathrm{disk}(q_1, r) \cup \mathrm{disk}(q_2, r)$, and the fact that p is the highest uncovered point implies that $W_\varepsilon(p)$ does not contain any other uncovered point. Hence, $p \in P_\varepsilon$, contradicting that $\mathrm{disk}(q_1, r)$ and $\mathrm{disk}(q_2, r)$ form a solution for P_ε.

When $\mathrm{dist}(q_1, q_2) \geqslant 2r + \varepsilon r$ we can show in a similar way that $\mathrm{disk}(q_1, r) \cup \mathrm{disk}(q_2, r)$ covers P. When $2r - \varepsilon r/2 \leqslant \mathrm{dist}(q_1, q_2) \leqslant 2r + \varepsilon r$ we cannot guarantee this, but if we blow up the disks by a factor $(1+\varepsilon)$ we are essentially back in the first case and we can apply the same reasoning. □

By Lemma 1 we obtain a $(1 + \varepsilon)$-approximation for the Euclidean 2-center problem if we can maintain the set P_ε. This seems difficult, so instead we maintain a superset $P^*_\varepsilon \supseteq P_\varepsilon$ defined as follows. Let $W_{\varepsilon/2}$ be a wedge of opening angle $\varepsilon/2$. We say that $W_{\varepsilon/2}$ is a *canonical* $(\varepsilon/2)$*-wedge* if the counter-clockwise angle that its angular bisector makes with the positive x-axis is $i\varepsilon/2$, for some integer $0 \leqslant i < \lceil 4\pi/\varepsilon \rceil$. We now define P^*_ε as the set of points p in P that have an empty canonical $(\varepsilon/2)$-wedge (that is, a wedge not containing points from P) with apex p. The following observation implies that Lemma 1 is still true if we replace P_ε by P^*_ε.

Observation 1 *Any point $p \in P$ that is the apex of an empty ε-wedge is also the apex of an empty canonical $(\varepsilon/2)$-wedge, so $P_\varepsilon \subseteq P^*_\varepsilon$.*

We are left with the problem of maintaining P^*_ε. This is done in a similar fashion as we maintained the convex hull vertices in a previous paper [4]: Each point $p \in P$ gets a time stamp that indicates how many time steps it takes before p can be in P^*_ε. At each time step we then consider only points whose time stamps have expired. Recall that there are $\lceil 4\pi/\varepsilon \rceil$ different classes of canonical wedges, corresponding to the orientation of their angular bisector. We treat each of these classes separately. Consider one such class, and assume without loss of generality that its angular bisector is pointing vertically upward. We wish to maintain the set $P^{*,\mathrm{up}}_\varepsilon$ of points whose upward canonical wedge is empty. Define $W^{\mathrm{down}}(p)$ to be the wedge with apex p that is the mirrored image of the upward canonical wedge of p, and let $\mathcal{W}^{\mathrm{down}}(t) := \{W^{\mathrm{down}}(p(t)) : p \in P\}$ be the set of all such downward wedges. Then a point q lies in the upward canonical wedge of p if and only if $p \in W^{\mathrm{down}}(q)$.

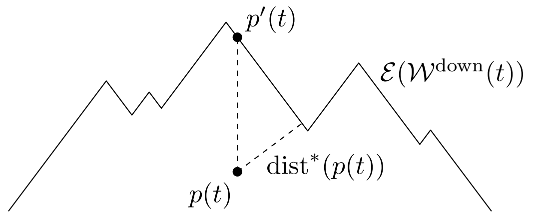

Figure 2: The point $p'(t)$ is the projection of $p(t)$ on $\mathcal{E}(\mathcal{W}^{\mathrm{down}}(t))$.

This implies the following lemma.

Lemma 2 *Let $\mathcal{E}(\mathcal{W}^{\mathrm{down}}(t))$ denote the upper envelope of the downward wedges at time t. Then $p \in P^{*,\mathrm{up}}_\varepsilon(t)$ if and only if $p(t)$ is a vertex of $\mathcal{E}(\mathcal{W}^{\mathrm{down}}(t))$.*

Because of the bounded speed of the points, we know that points far from the upper envelope $\mathcal{E}(\mathcal{W}^{\mathrm{down}})$ need a lot of time before they can appear on the envelope. Hence, we can use the distance from p to $\mathcal{E}(\mathcal{W}^{\mathrm{down}})$ to define its time stamp. To be able to compute time stamps quickly, we will not use the Euclidean distance from p to $\mathcal{E}(\mathcal{W}^{\mathrm{down}})$ but an approximation of it. Let $p'(t)$ be the vertical projection of $p(t)$ onto $\mathcal{E}(\mathcal{W}^{\mathrm{down}}(t))$; see Figure 2. Then our approximated distance is defined as $\mathrm{dist}^*(p(t)) := \mathrm{dist}(p(t), p'(t)) \cdot \sin(\varepsilon/4)$. Note that $\mathrm{dist}^*(p(t))$ is equal to the distance from $p(t)$ to the boundary of the downward wedge $W^{\mathrm{down}}(p'(t))$. Because $W^{\mathrm{down}}(p'(t))$ is completely below (or on) $\mathcal{E}(\mathcal{W}^{\mathrm{down}})$, the actual Euclidean distance from $p(t)$ to $\mathcal{E}(\mathcal{W}^{\mathrm{down}})$ is at least $\mathrm{dist}^*(p(t))$. Hence, we can safely use $\mathrm{dist}^*(p(t))$ to define the time stamps. Thus, when we compute the time stamp of a point p at time t we set

$$t^{\mathrm{up}}(p) := \min(\lfloor \mathrm{dist}^*(p(t))/2d_{\max} \rfloor + 1, n).$$

Lemma 3 *If a point p receives time stamp $t^{\mathrm{up}}(p)$ at time t_i, then p cannot be on $\mathcal{E}(\mathcal{W}^{\mathrm{down}}(t_j))$ for $t_i < t_j < t_i + t^{\mathrm{up}}(p)$.*

The final time stamp of a point p is defined as the minimum over all time stamps computed for p for the $\lceil 4\pi/\varepsilon \rceil$ different wedge orientations. The algorithm for maintaining the Euclidean 2-center can now be summarized as follows. Initially (at time $t = t_0$) we compute a time stamp $t(p)$ for every point p, which is the minimum over the time stamps for the $\lceil 4\pi/\varepsilon \rceil$ canonical wedge orientations. Then at each time step $t = t_i$ we take the set $Q(t)$ of points whose time stamps expire at time t. For each canonical orientation we use a simple sweep-line algorithm to compute in $O(|Q(t)| \log |Q(t)|)$ time the envelope of the mirrored wedges of the points in $Q(t)$. Since there are $\lceil 4\pi/\varepsilon \rceil$ different orientations this takes $O((1/\varepsilon)|Q(t)| \log |Q(t)|)$ time in total. The collection of all points $p \in Q(t)$ that are a vertex of at least one of the envelopes is the set $P^*_\varepsilon(t)$. We then solve the Euclidean 2-center problem on $P^*_\varepsilon(t)$ using an algorithm for static points, giving us two disks disk(q_1, r) and disk(q_2, r). (To get the best running time, we use Chan's algorithm [6] for this.) If $2r - \varepsilon r/2 \leqslant \text{dist}(q_1, q_2) \leqslant 2r + \varepsilon r$ then we report disk$(q_1, r + \varepsilon r)$ and disk$(q_2, r + \varepsilon r)$, otherwise we report disk(q_1, r) and disk(q_2, r). Finally, we compute new time stamps for the points in $Q(t)$. Since we already have all the envelopes this can be done in $O((1/\varepsilon)|Q(t)| \log |Q(t)|)$ time in total.

The running time of our algorithm strongly depends on the number of points in $Q(t)$. Although in the worst case $|Q(t)|$ may be large, we can show that it is small on average. The proof of the following lemma is similar to a proof in a previous paper [4, Lemma 6].

Lemma 4 *The number of points in $Q(t)$ is $O((1/\varepsilon^2)k\Delta_{2,k} \log n)$ amortized.*

Using that the static 2-center algorithm by Chan [6] on m points runs in $O(m \log^2 m(\log \log m)^2)$ time, we obtain the following theorem.

Theorem 5 *Let P be a set of n moving points that adheres to the Displacement Assumption with parameters k and $d_{\max}$, let $\Delta_{2,k}$ denote the maximum $(2, k)$-spread of P at any time t, and let $0 < \varepsilon \leqslant \pi/4$. Then we can maintain a $(1 + \varepsilon)$-approximation of the Euclidean 2-center for P in the black-box KDS model in $O((k/\varepsilon^3)\Delta_{2,k} \log^3 n(\log \log n)^2))$ amortized time per time step and using $O(n)$ space.*

3 Conclusions

We have shown how to maintain an approximation of the Euclidean 2-center problem in amortized sublinear time in the black-box model under certain assumptions on the distribution of the points. In the solution presented here the centers can "jump" between time steps. That is, between two consecutive time steps the distances between the centers can be very large compared to $d_{\max}$. For clustering this is not a problem, but for facility location problems this is undesirable. Durocher and Kirkpatrick [9] show a lower bound of $\sqrt{2}$ on the approximation ratio when the centers move with bounded speed. They provide an approximation scheme that achieves an approximation ratio of $8/\pi \approx 2.55$. We have also investigated bounded speed approximations for Euclidean kinetic 2-center problem in the black-box model. In the full paper we show how to obtain a 2.28-approximation for the Euclidean 2-center, such that the centers move at most $4\sqrt{2}d_{\max}$ per time step. There, we also study the rectilinear version of the kinetic 2-center problem (with and without speed restriction for the centers).

References

[1] P.K. Agarwal and C.M. Procopiuc. Exact and approximation algorithms for clustering. *Algorithmica* 33:201–226, 2002.

[2] M. Bādoiu, S. Har-Peled and P. Indyk. Approximate clustering via core-sets. In *Proc. 34th ACM Symp. Th. Comp.*, pages 250–257, 2002.

[3] J. Basch, L.J. Guibas and J. Hershberger. Data structures for mobile data. In *Proc. 8th ACM-SIAM Symp. Disc. Alg.*, pages 747–756, 1997.

[4] M. de Berg, M. Roeloffzen and B. Speckmann. Kinetic convex hulls and Delaunay triangulations in the black-box model. In *Proc. 27th ACM Symp. Comp. Geom.*, pages 244–253, 2011.

[5] M. de Berg, M. Roeloffzen and B. Speckmann. Kinetic compressed quadtrees in the black-box model with applications to collision detection for low-density scenes. In *Proc. 20th Euro. Symp. Alg.*, pages 383–394, 2012.

[6] T.M. Chan. More planar two-center algorithms. *Comp. Geom.* 13:189–198, 1999.

[7] B. Degener, J. Gehweiler and C. Lammersen. Kinetic facility location. *Algorithmica* 57:562-584, 2010.

[8] Z. Drezner. On the rectangular p-center problem. *Naval Res. Log. Quart.* 34:229–234, 1987.

[9] S. Durocher and D. Kirkpatrick. Bounded-velocity approximations of mobile Euclidean 2-centres. *Int. J. Comp. Geom. Appl.* 18(3):161–183, 2008.

[10] J. Erickson. Dense point sets have sparse Delaunay triangulations. *Disc. Comp. Geom.* 30:83–115, 2005.

[11] S. Friedler and D. Mount. Approximation algorithm for the kinetic robust K-center problem. *Comp. Geom.* 43:572–586, 2010.

[12] J. Gao, L.J. Guibas and A. Nguyen. Deformable spanners and applications. In *Proc. 20th ACM Symp. Comp. Geom.*, pages 190–199, 2004.

[13] M. Hoffmann. A simple linear algorithm for computing rectangular three-centers. In *Proc. 11th Can. Conf. Comp. Geom.*, pages 72-75, 1999.

[14] N. Megiddo and K.J. Supowit. On the complexity of some common geometric location problems. *SIAM J. Comp.* 13(1):182196, 1984.

[15] D. Nussbaum. Rectilinear p-piercing problems. In *Proc. 1997 Int. Symp. Symb. Algebraic Comp.*, pages 316–324, 1997.

[16] M. Segal. On piercing of axis-parallel rectangles and rings. In *Proc. 1997 Int. Symp. Symb. Algebraic Comp.*, pages 430–442, 1997.

[17] M. Segal. Lower bounds for covering problems. *J. Math. Model. Alg.* 1(1):17–29, 2002.

[18] M. Sharir. A near-linear algorithm for the planar 2-center problem. *Disc. Comp. Geom.* 18:125–134, 1997.

[19] M. Sharir and E. Welzl. Rectilinear and polygonal p-piercing and p-center problems. In *Proc. 12th ACM Symp. Comp. Geom.*, pages 122–132, 1996.

[20] J.J. Sylvester. A question in the geometry of situation. *Quart. J. Math.* 1:79, 1857.