Efficient Approximation Algorithm With Partition Technique
For The Diameter Of A Set Of Points In 2D Plane

Zhipeng Wang¹*, Jieying Hong², and Wei Niu¹

¹Ecole Centrale Pekin, Beihang University, Beijing 100191, China
²ESSEC Asia-Pacific, ESSEC Business School, Singapore 139408, Singapore
*ZhipengWang@buaa.edu.cn

Abstract. We apply the partition technique to build an efficient approximation algorithm to
compute the diameter of a set of points in 2D Euclidean plane. The optimal time of our algorithm
is \(O(N + \frac{1}{\delta} \log \frac{1}{\delta})\), up to a \((1 + \delta)\) factor, for 2-dimensional \(N\)-point set. The error bounds
are proved strictly. Compared to the prior works, our algorithm is without complex data structure
and easy to be implanted. The partition technique is general and may be applied in higher-
dimensional space.

1. Introduction
Given a finite set of points \(P\) in a 2D Euclidean plane \(\mathbb{R}_2\), its diameter, denoted by \(d_P\), is defined as the
maximum distance between two points of \(P\). Computing the diameter of a point set is a fundamental
problem in computer science. It has been proved that in a Euclidean plane, finding the accurate value of
the diameter of a set of \(N\) points can be reduced to formulating the convex hull of them, with a lower
bound of complexity \(O(N \log N)\) [1, 2, 3].

In the science of big data, this classical problem encounters new challenges. For big data, the number
of points \(N\) can be huge, and one usually expects linear approximate algorithms to replace the
\(O(N \log N)\) complexity. In a mathematical language, this problem can be described as: given an
arbitrary positive number \(0 < \delta < 1\), we aim at outputting an approximate diameter \(d_\delta\) in linear time, so
that \(1 - \delta \leq d_\delta / d_P \leq 1 + \delta\).

More generally, we can use the notation of [4]. Let \(P\) be a set of \(N\) points in \(D\)-dimensional Euclidean
space where \(D\) is a constant. Let \(0 < \delta < 1\) be a user-specified parameter which is not necessarily a
constant and let \(E = \lceil \frac{1}{\delta} \rceil\). There has been a long series of prior results of approximation algorithms.
Table 1 gives a comparison of different \((1 + \delta) - \text{approximation}\) algorithms to compute the diameter of
\(P\), in chronological order.

In this paper, we present a new algorithm to compute an \((1 + \delta) - \text{approximation}\) algorithm of the
diameter of \(P\) with almost \(O(N + E^D/2 \log E^{D/2})\) complexity when \(D = 2\) using partition technique.
Table 1. Comparison of different \((1 + \delta) - \text{approximation}\) algorithms of the diameter of \(P\)

| Author          | complexity                        | main method                      |
|-----------------|-----------------------------------|----------------------------------|
| Agarwal1992[5]  | \(O(E^3N)\)                      | Random-sampling approach         |
| Barequet2001[6] | \(O(N + E^{3/2})\)               | Rounding to Grid                 |
| Chan2000[7]     | \(O(N + E^{3D/2})\)              | Rounding to Grid + Cones         |
| Chan2000[7]     | \(O(N + E^D)\)                   | Cones + Dimension Reduction      |
| Arya2014[8]     | \(O(N + E^{D/2}N^{1/2})\)        | \(\varepsilon - \text{Dependencies}\) |
| Chan2017[4]     | \(O(N + E^{D/2})\)               | Chebyshev Polynomials            |
| This work       | \(O(N + E^{D/2} \log E^{D/2})\)  | Partition technique              |

2. An approximation algorithm for the diameter in a Euclidean plane

We first give our main algorithm to compute the diameter in a Euclidean plane based on partition technique and its accuracy and efficiency will be analyzed later.

**Input:** A finite set of points \(P\) in \(\mathbb{R}^2\)

**Procedure:**

1. Choose a point \(O \in P\) arbitrarily as the origin, and then divide the plane into \(6n\) same regions with \(n \in \mathbb{Z}^*_+\).

2. In each region \(S_i\) \((i = 0, \ldots, 6n-1)\), find the farthest point from the origin and let \(r_i\) denote the distance between the farthest point of \(S_i\) and the origin. By using the origin \(O\) as the center of a circle and \(r_i\) as the radius, obtain \(6n\) sector regions, as Figure 1.

3. Let \(p_i\) \((i = 0, \ldots, 6n-1)\) be the midpoint of the arc of each sector region, and compute the largest distance \(d_o\) of these \(6n\) midpoints \(p_i\)s.

**Output:** \(d_o = \max(r_0, \ldots, r_{6n-1}, d_o)\).

The main theorem of this paper is as follow, which shows the relationship between \(d_o\) of the point set \(P\) and the largest distance \(d_P\). Note that the virtual points \(p_i\) are not necessarily real points in \(P\).

**Theorem 1.** If \(d_P\) is the diameter of a finite set of points \(P\) in \(\mathbb{R}^2\), and \(d_o\) is the largest distance of the \(6n\) virtual midpoints \(p_i\) as defined above, then the following statement holds:

\[
\left(\frac{1}{2} + 2\cos\left(\frac{(2n-1)\pi}{3n}\right)\right)^{\frac{1}{2}} \leq \frac{d_P}{d_o} \leq \left(\frac{1 + \cos\left(\frac{(2n-2)\pi}{3n}\right)}{1 + \cos\left(\frac{(2n-1)\pi}{3n}\right)}\right)^{\frac{1}{2}}.
\]

We will prove the lower and upper bounds of \(d_P/d_o\) respectively.

2.1. Lower bound of \(d_P/d_o\)

Without loss of generality, suppose that an endpoint of the line segment \(d_P\) is in the region \(S_0\), and then we denote the opposite angle region by \(S_{6n}\) and denote the other regions clockwise by \(S_i\) \((i = 0, \ldots, 6n-1)\). Note that \(S\) is exactly \(S_{6n}\). Let the line passing through the origin \(O\) and the midpoint of the arc of the region \(S\) be the x-axis, then we can set up the Cartesian coordinate system in the plane, as Figure 2. The coordinate of the midpoint \(p_i\) of the arc in each region \(S_i\) is \((-r_i \cos \frac{i\pi}{3n}, r_i \sin \frac{i\pi}{3n})\), where \(i \in [0, 6n-1]\) and \(r_i\) is the radius of the sector region \(S_i\).
Figure 1. Diagram of the partition for the point set. Solid points are real points in the set $P$. Empty points are virtual points in the set $p$.

Before giving the proof of the lower bound of $d_p/d_o$, we firstly bring out the following lemmas.

**Lemma 1.** If an endpoint of the line segment $d_o$ is in the region $S(i.e., S_{2n})$ as we supposed above, then the other endpoint of $d_o$ cannot be obtained in the regions $S_{2n+1}, \ldots, S_{3n-1}$.

**Proof.** Denote $R = \max(r_0, \ldots, r_{6n-1})$, then $R \leq d_o$. For the point $p_i(-r_i \cos \frac{i\pi}{3n}, r_i \cos \frac{i\pi}{3n})$ in the region $S_{2n+1}, \ldots, S_{3n-1}(i.e., i \in [2n+1, 3n+1])$, and the point $p_{3n}(-r_{3n}, 0)$ in the region $S$, we have

$$p_i p_{3n} = r^2 + r_i^2 + 2r_i \cos \frac{i\pi}{3n} < r^2 + r_i^2 - rr_i \leq R^2 \leq d_o^2.$$  

**Lemma 2.** If an endpoint of the line segment $d_o$ is in the region $S(i.e., S_{3n})$ as we supposed above, then the other endpoint of $d_o$ cannot be obtained in the regions $S_{3n+1}, \ldots, S_{4n-1}$.

The proof of Lemma 2 is similar. Moreover, the cases that an endpoint of $d_o$ is in the regions $S_0, S_{6n-1}, \ldots, S_{4n}$ are equivalent to those in the regions $S_0, S_1, \ldots, S_{2n}$. Therefore, if we suppose that an endpoint of the line segment $d_o$ is in the region $S$, then from Lemma 1 and Lemma 2, we only need to consider the $2n+1$ cases where the other endpoint of $d_o$ is in the regions $S_0, S_1, \ldots, S_{2n}$.

**Case 1:** $i \in [0, 2n-1]$ As we supposed above, the two endpoints of $d_o$ are in the region $S$ and the region $S_i (i = 0, 1, \ldots, 2n-1)$, then there certainly exist a point $q_i (-r_i \cos \theta_i, r_i \sin \theta_i)$ on the arc of $S_i$ where

$$\theta_i \in \left[-\frac{\pi}{6n}, \frac{\pi}{6n}\right] \quad \text{and a point } q_2 (-r_i \cos \theta_i, r_i \sin \theta_i) \text{ on the arc of } S_j (i = 0, 1, \ldots, 2n-1),$$

where

$$\theta_i \in \left[-\frac{\pi}{6n}, \frac{\pi}{6n}\right].$$

Then

$$d_o \leq r^2 + r_i^2 + 2rr_i \cos \frac{i\pi}{3n} \quad \text{and} \quad d_o \geq q_i q_2 \geq \left(r^2 + r_i^2 + 2rr_i \cos \frac{i\pi}{3n}\right)^{\frac{1}{2}}.$$

Then

$$\frac{d_p}{d_o} \geq \left(\frac{r^2 + r_i^2 + 2rr_i \cos \frac{(i+1)\pi}{3n}}{r^2 + r_i^2 + 2rr_i \cos \frac{i\pi}{3n}}\right)^{\frac{1}{2}} \geq \left(1 + \frac{2 \left(\cos \frac{(i+1)\pi}{3n} - \cos \frac{i\pi}{3n}\right)}{2 + 2 \cos \frac{i\pi}{3n}}\right)^{\frac{1}{2}} = \left(1 + \cos \frac{(i+1)\pi}{3n}\right)^{\frac{1}{2}}.$$
Let \( f(x) = \left( 1 + \cos \left( \frac{(x+1)\pi}{3n} \right) \right) \left( 1 + \cos \frac{x\pi}{3n} \right)^{-1} \) where \( x \in [0, 2n-1] \). \( f'(x) < 0 \) in the interval \([0, 2n-1]\). It leads up to the following inequalities:

\[
d_p/d_o \geq \sqrt{f(i)} \geq \sqrt{f(2n-1)} = \left( 2 + 2\cos \left( \frac{(2n-1)\pi}{3n} \right) \right)^{1/2}.
\]

**Case II:** \( i = 2n \) Therefore \( d_o = p_{2n}p_{3n} = r^2 + r_i^2 - rr_i \) \( \leq R \). Since \( d_p \geq R \), we have \( d_p/d_o \geq 1 \).

Concluding the two cases, the lower bound of \( d_p/d_o \) is obtained: \( d_p/d_o \geq \left( 2 + 2\cos \left( \frac{(2n-1)\pi}{3n} \right) \right)^{1/2} \).

2.2. **Upper bound of \( d_p/d_o \)**

Similar to the proof of the lower bound, supposing that an endpoint of \( d_p \) is in the region \( S_i \) for \( i \in [0, 2n] \) need to be considered.

**Case I:** \( i \in [0, 2n-1] \) As we supposed above, there exist a point \( m_i (\hat{r} \cos \theta, \hat{r} \sin \theta) \) in the region \( S_i \) and another point \( m_{1 - i} (-\hat{r}, \cos \theta, -\hat{r} \sin \theta) \) in the region \( S_{1 - i} \), where \( \hat{r} \in [0, r], \hat{r}_{1 - i} \in [0, r_{1 - i}] \), \( \theta \in [-\pi/6n, \pi/6n] \), \( \theta_{1 - i} \in [i\pi/3n - \pi/6n + \pi/6n, i\pi/3n] \). Then \( \theta + \theta_{1 - i} \in [(i-1)\pi/3n, (i+1)\pi/3n] \). Let \( a = \cos \left( \frac{(i-1)\pi}{3n} \right) \in (-1, 1] \), we have \( m_i m_{1 - i} = r^2 + r_i^2 + 2rr_i \cos (\theta + \theta_{1 - i}) \leq h(\hat{r}) = r^2 + r_i^2 + 2a \hat{r}r_i \). In the case \( a \in [0, 1] \),

\[
h(\hat{r}) \leq r^2 + r_i^2 + 2ar_i .
\]

Thus \( d_p/d_o \leq \left( r^2 + r_i^2 + 2ar_i \cos \left( \frac{(i-1)\pi}{3n} \right) \right)/\left( r^2 + r_i^2 + 2ar_i \cos \left( \frac{i\pi}{3n} \right) \right)^{1/2} \). In the case \( a = \cos \left( \frac{(i-1)\pi}{3n} \right) \in (-1, 0] \), compute that \( h(\hat{r}) = 2(\hat{r} + ar_i) \).

a) If \( r_i \geq 2r \), then \( h(\hat{r}) \leq 0 \) when \( \hat{r} \in [0, r] \). And we can get the maximum of \( h(\hat{r}) \):

\[
h(\hat{r})_{\text{max}} = h(0) = r_i^2 \leq r^2 \leq R^2 .
\]

Thus \( d_p = m_i m_{1 - i} \leq R \), that leads that \( d_p = m_i m_{1 - i} = R \) and \( r_{1 - i} = R, r = 0 \).

At this moment, \( d_o = p_{i} p_{3n} = R \). Therefore, we have \( d_p/d_o \leq 1 \).

b) If \( 0 \leq r_i < 2r \), the values of \( h'(\hat{r}) \) leads \( h(\hat{r})_{\text{max}} = h(0) \) or \( h(\hat{r}) \). If \( h(\hat{r})_{\text{max}} = h(0) \), same as the case a), \( d_p/d_o \leq 1 \). If \( h(\hat{r})_{\text{max}} = h(\hat{r}) = r^2 + r_i^2 + 2ar_{1 - i} \), let \( t(\hat{r}_{1 - i}) = r^2 + r_i^2 + 2ar_{1 - i} \), and \( t(\hat{r}_{1 - i}) = (2\hat{r} + ar) \) where \( \hat{r}_{1 - i} \in [0, r_{1 - i}] \). If \( r_{1 - i} < 2r \), then \( t(\hat{r}_i) < 0 \), and \( t(\hat{r}_{1 - i})_{\text{max}} = t(0) = r^2 \). In this case, \( d_p = m_i m_{1 - i} = \sqrt{h(\hat{r})} \leq \sqrt{t(0)} = r \leq R \), then \( d_p/d_o \leq 1 \). If \( r_{1 - i} \geq 2r \), \( t(\hat{r}_i) \geq 0 \) and \( t(\hat{r}_{1 - i})_{\text{max}} = t(\hat{r}_i) = r^2 + r_i^2 + 2ar_i \). Here we have \( d_p = \sqrt{h(\hat{r})} \leq \sqrt{t(\hat{r}_i)} \). Thus

\[
d_p/d_o \leq \left( r^2 + r_i^2 + 2ar_i \cos \left( \frac{(i-1)\pi}{3n} \right) \right)/\left( r^2 + r_i^2 + 2ar_i \cos \left( \frac{i\pi}{3n} \right) \right)^{1/2} \).
\]

**Case II:** \( i = 0 \) \( d_o = p_j p_{3n} = r + r_i \), \( d_p = m_j m_{3n} = \left( r^2 + r_i^2 + 2rr_i \cos \left( \theta_{1 - i} \right) \right)^{1/2} \leq \hat{r} + \hat{r}_i \leq r + r_i \).

Therefore, \( d_p/d_o \leq 1 \).
The upper bound of $d_p / d_o$ can be concluded: 
\[
\frac{d_p}{d_o} \leq \left(1 + \cos \left(\frac{2(n-2)\pi}{3n}\right)\right) \left(1 + \cos \left(\frac{(2n-1)\pi}{3n}\right)\right)^{-\frac{1}{3}},
\]
and the inequality is obtained by the similar computation of lower bound. Thus, **Theorem 1** is proved.

### 2.3. Remarks

The complexity of visiting all $N$ points in set $P$ is linearly $O(N)$. The complexity of calculating $d_o$ of $6n$ virtual points will be $O(n \log n)$ as was introduced in section 1. The total complexity is $O(N + n \log n)$. Let $d = d_o \left(2 + 2\cos \left(\frac{(2n-1)\pi}{3n}\right)\right)^{\frac{1}{2}}$, then $1 \leq d_p / d \leq 1 + \delta = \left(2 + 2\cos \left(\frac{(2n-2)\pi}{3n}\right)\right)^{\frac{1}{2}}$. If $n$ is big enough, 
\[
sin \frac{2\pi}{3n} = 0, \cos \frac{2\pi}{3n} = 1, \delta = \left(2 - 2\cos \frac{2\pi}{3n} + \sqrt{3}\sin \frac{2\pi}{3n}\right)^{\frac{1}{3}} - 1 = \left(1 + \frac{2\pi}{\sqrt{3}n}\right)^{\frac{1}{3}} - 1 = \frac{\pi}{\sqrt{3}n}.
\]
Then, an $(1 + \delta) - approximation$ algorithm with complexity $O(N + (1/\delta) \log (1/\delta))$ is obtained.

### 3. Conclusions

As a fundamental problem of big data, linear approximation algorithms for the diameter of a set of points will be potentially useful. By introducing the partition technique, which is totally different from the main methods of other approximation algorithms, we introduce an $(1 + \delta) - approximation$ algorithm with complexity $O(N + n \log n) = O(N + (1/\delta) \log (1/\delta))$. Compared to the work of [4] whose main technique is Chebyshev Polynomials, the implementation of our algorithm is extremely simple and does not require any complicated data structure. In practice $n$ will be much smaller than $N$, therefore $O(n \log n)$ even $O(n^2)$ will be negligible compared to $O(N)$. The present contribution is a preliminary attempt of partition technique in 2D plane. This new partition technique might also be extended in higher dimensional space, but a division of hyper-sphere [9, 10, 11] will be required. In those situations, other partition schemes may be more efficient and the related accuracy will also be more complicated. In addition, we note that the accuracy of our work is not the best one among all approximation algorithms for the diameter of a set of points. Other partitions might be chosen to improve the accuracy. These problems are expected to be investigated in a future work.

### References

[1] A. C.-C. Yao, “On constructing minimum spanning trees in k-dimensional spaces and related problems,” SIAM Journal on Computing, vol. 11, no. 4, pp. 721–736, 1982.

[2] F. Preparata, “Shamos. mi: Computational geometry. an introduction.texts and monographs in computer science. corr. 3rd printing,” 1985.

[3] J. L. Bentley, F. P. Preparata, and M. G. Faust, “Approximation algorithms for convex hulls,” Communications of the ACM, vol. 25, no. 1, pp. 64–68, 1982.

[4] T. M. Chan, “Applications of chebyshev polynomials to low-dimensional computational geometry,” in LIPics-Leibniz International Proceedings in Informatics, vol. 77. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[5] P. K. Agarwal, J. Matousek, and S. Suri, “Farthest neighbors, maximum spanning trees and related problems in higher dimensions,” Computational Geometry, vol. 1, no. 4, pp. 189–201, 1992.

[6] G. Barequet and S. Har-Peled, “Efficiently approximating the minimum-volume bounding box of a point set in three dimensions,” Journal of Algorithms, vol. 38, no. 1, pp. 91–109, 2001.

[7] T. M. Chan, “Approximating the diameter, width, smallest enclosing cylinder, and minimum-width annulus,” in Proceedings of the sixteenth annual symposium on Computational
geometry. ACM, 2000, pp. 300–309.

[8] S. Arya and T. M. Chan, “Better -dependencies for offline approximate nearest neighbor search, euclidean minimum spanning trees, and -kernels,” in Proceedings of the thirtieth annual symposium on Computational geometry. ACM, 2014, p. 416.

[9] P. W. Cooper, “The hypersphere in pattern recognition,” Information and control, vol. 5, no. 4, pp. 324–346, 1962.

[10] S. Katsuki and D. M. Frangopol, “Hyperspace division method for structural reliability,” Journal of Engineering Mechanics, vol. 120, no. 11, pp. 2405–2427, 1994.

[11] K. Sato and A. Yamaji, “Uniform distribution of points on a hypersphere for improving the resolution of stress tensor inversion,” Journal of structural geology, vol. 28, no. 6, pp. 972–979, 2006.