# 2D Closest Pair Problem: A Closer Look

Ovidiu Daescu\*

Ka Yaw Teo\*

## Abstract

A closer look is taken at the well-known divide-andconquer algorithm for finding the closest pair of a set of points in the plane under the Euclidean distance. An argument is made that it is sufficient, and sometimes necessary, to check only the next three points following the current point associated with the *y*-sorted array in the combine phase of the algorithm.

#### 1 Introduction

The closest pair of points problem is a fundamental problem in computational geometry and has received significant attention over the years. The input for the problem consists of a set  $P = \{p_1, p_2, \ldots, p_n\}$  of n points in  $\mathbb{R}^d$ , where d is typically treated as a constant, and the objective is to find two points p and q in P such that  $d(p,q) = \min\{d(p_i, p_j)|p_i, p_j \in P, p_i \neq p_j\}$ , where d(p,q) represents the Euclidean distance between points p and q. It is well known that the problem can be solved optimally in  $O(n \log n)$  time for any constant dimension d using a divide and conquer approach.

In this paper, the two-dimensional (2D) case of the problem is considered, and a tight geometric bound is derived on a specific step of the combine phase. The description of the algorithm is given in [1, 2, 3], and it can be summarized as follows: (1) Divide P into two equal parts,  $P_L$  and  $P_R$ , by a vertical line  $l: x = x_m$ , where  $x_m$  is the median x-coordinate of the points in P; (2) Recursively find the closest pair of points in  $P_L$ and  $P_R$ , respectively; (3) Let  $\delta$  be the minimum of the two distances returned in the previous step; that is, no two points in  $P_L$  or  $P_R$  can be closer than  $\delta$ . Then, the distance for the closest pair in P is either  $\delta$  or given by a pair of points (p, q) where  $p \in P_L$  and  $q \in P_R$ . In order to find the distance in the later case, denoted as  $\delta_{L,R}$ , let  $Y_{\delta}$  represent the points p in P, with  $x_p \in$  $[x_m - \delta, x_m + \delta]$ , sorted in non-decreasing order by ycoordinate. Then,  $\delta_{L,R}$  can be found by traversing  $Y_{\delta}$  in the sorted order and, for each point  $p_i \in Y_{\delta}$ , computing the distances from  $p_i$  to the next five points following  $p_i$  in  $Y_{\delta}$  (see Exercise 33.4 in [2]). It has been posed to the authors of the current paper, as an open problem, to prove or disprove whether it suffices to check only the next four points following  $p_i$  in  $Y_{\delta}$ . The main result is the following theorem.

**Theorem 1** It is sufficient, and sometimes necessary, to check only the next three points following  $p_i$  in  $Y_{\delta}$ .

Note that a similar and easier analysis follows if any point  $p \in Y_{\delta}$  is required to satisfy  $x_p \in (x_m - \delta, x_m + \delta)$ , since we are looking for a pair of points that gives  $\delta_{L,R} < \delta$ . Nevertheless, for the sake of argument, the subsequent analysis has been performed in line with the original algorithm proposed in [1, 2, 3], which considers all points p in  $Y_{\delta}$  with  $x_p \in [x_m - \delta, x_m + \delta]$ .

### 2 A Closer Look

In the current section, we prove Theorem 1. We start by taking care of inputs that have overlapping points, defined as points with the same x- and y-coordinates. Overlapping points imply that the closest pair has a zero distance.

Consider a  $2\delta \times \delta$  axis-aligned box *B* centered at  $x_m$ , and assume that *B* is placed with its bottom edge on the *x*-axis and  $x_m = 0$ . As shown in [2], at most four points in *B* belong to  $P_L$ , and at most four to  $P_R$ .

**Lemma 2** Let  $p_i$  and  $p_j$  be two overlapping points in  $Y_{\delta}$ , with i < j. Then, for each point in  $Y_{\delta}$ , it is necessary and sufficient to check the next three points in  $Y_{\delta}$  to detect the overlapping pair of points.

**Proof.** Let  $p_i$  be the current point in  $Y_{\delta}$ , and assume that  $p_i \in P_L$ . There can be at most three other points in B with the same y-coordinate as  $p_i$ . That happens when the x-coordinate of  $p_i$  is  $x_m$ , there is a point in  $P_L$  at  $(x_m - \delta, 0)$ , and two points in  $P_R$  at  $(x_m + \delta, 0)$  and  $(x_m, 0)$ , respectively. Thus, it suffices to check the next three points following  $p_i$  in  $Y_{\delta}$  to find the overlapping points. On the other hand, if  $p_{i+1}$ ,  $p_{i+2}$ , and  $p_{i+3}$  are in the aforementioned order, then  $p_{i+3}$  has to be checked in order to identify the overlapping points.  $\Box$ 

Notice that if B is defined as an open box, where any point  $p \in Y_{\delta}$  satisfies  $x_p \in (x_m - \delta, x_m + \delta)$ , then the overlapping points can be found by checking only the next point following  $p_i$  in  $Y_{\delta}$ .

From now on, assume that the input set P contains no overlapping points. Let  $B_L$  and  $B_R$  denote the left and right sides of B, respectively, as partitioned by vertical

<sup>\*</sup>Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA. {daescu, ka.teo}@utdallas.edu



Figure 1: Illustration for Subcase A of Case I. (a)-(d) The positions of points  $p_2$ ,  $p_3$ , and  $p_4$  in  $B_R$  change as point  $p_2$  varies from (0, 0) to  $(\delta, 0)$ . The circular arc (of radius  $\delta$ ) centered at p indicates that only two points in  $B_R$  are located  $\leq \delta$  from p.

line  $y = x_m$ .  $B_L$  and  $B_R$  are  $\delta \times \delta$  squares. Let  $p = p_i$  be the current point in  $Y_{\delta}$ , and assume that  $p \in P_L$ .

Given that the maximum number of points of separation of at least  $\delta$  in  $B_L$  (or  $B_R$ ) is four, at most three points in  $Y_{\delta}$  with an array index greater than *i* lie within  $B_L$ . In other words, at most three points coming after *p* in  $Y_{\delta}$ , not necessarily in a consecutive order, are in  $B_L$ .

This observation results in four different cases that must be considered separately - (I) three points after pare in  $B_L$ , (II) two points after p are in  $B_L$ , (III) one point after p is in  $B_L$ , and (IV) no point after p is in  $B_L$ (this case is trivial and omitted herein). In each of these cases, the worst-case scenario is determined, and that is the minimum number of points following p in  $Y_{\delta}$  that must be examined in order to identify the closest pair of points correctly. For simplicity of notation, Y is used instead of  $Y_{\delta}$ , and  $p_1, p_2, \ldots$  in place of  $p_{i+1}, p_{i+2}, \ldots$ hereafter.

As shall be seen shortly, by looking at the four cases,

 $p_5$  cannot be a candidate for the closest pair with p. It is then required to prove that either (i) one of  $\{p_1, p_2, p_3\}$ is closer to p than  $p_4$ , or (ii) if  $p_4$  is closer to p than any one of  $\{p_1, p_2, p_3\}$ , then one of  $\{p_1, p_2, p_3\}$  is closer to  $p_4$ than p, and so  $(p, p_4)$  cannot be the closest pair.

### Case I: Three points after p are in $B_L$

Suppose that p is at the bottom right corner of  $B_L$  (the case where p is at the bottom left corner is trivial). We have the following subcases.

**Subcase A:** Three points after p are in  $B_R$ . Consider Figure 1, where  $p_2$ ,  $p_3$ , and  $p_4$  define an equilateral triangle of side length  $\delta$  (i.e., worst case, in which the points are at their closest distance from each other in  $B_R$ ). The first point in  $B_R$  can have the same y-coordinate as p (i.e., worst case, given that choosing a larger y-coordinate for the first point in  $B_R$  would simply increase the distances between the points in  $B_R$  and p), and it can be either the first or second point following



Figure 2: Illustration for Subcase B of Case II when p = (0, 0). (a)-(d) The locations of points  $p_2$ ,  $p_3$ , and  $p_5$  in  $B_R$  change as point  $p_2$  varies from (0, 0) to  $(\delta, 0)$ . Based on the circular arc (of radius  $\delta$ ) centered at p, only two points in  $B_R$  are located  $\leq \delta$  from p.

p in Y (i.e., labeled as  $p_1$  or  $p_2$ ). Since it is irrelevant to the subsequent argument, assume that it is  $p_2$ . When  $p_2$ varies from (0, 0) to  $(\delta/2, 0)$ ,  $d(p, p_3)$  is always greater than or equal to  $\delta$ , and  $d(p, p_4)$  is always greater than  $d(p, p_2)$  (Figure 1 (a), (b)). As  $p_2$  varies from  $(\delta/2, 0)$ to  $(\delta, 0)$ ,  $d(p, p_4)$  is always greater than or equal to  $\delta$ , but  $d(p, p_3)$  can become smaller than  $d(p, p_2)$  (Figure 1 (c), (d)). Thus, in Subcase A, p has to be compared with the three following points in Y.

**Subcase B:** Two points after p are in  $B_R$ . In this case,  $p_2$  and  $p_3$  can be located within the interior of  $B_R$  to be deemed competitive, and thus p must be compared with the next three following points in Y.

Overall, in Case I, only three points following p in Y need to be taken into consideration.

## Case II: Two points after p are in $B_L$

In this case, p can be located anywhere along the bottom edge of  $B_L$ .

**Subcase A:** Four points after p are in  $B_R$ . This case is trivial; the point at the leftmost bottom corner of  $B_R$  is the closest point to p.

**Subcase B:** Three points after p are in  $B_R$ . Refer to Figure 2, where the triangles defined by  $(p, p_1, p_4)$  and  $(p_2, p_3, p_5)$ , respectively, are equilateral with side length  $\delta$ .

First, assume that p is at the lower right corner of  $B_L$ . The worst case occurs when the first point in  $B_R$ , either  $p_1$  or  $p_2$  (say  $p_2$ ), has the same y-coordinate as p. When  $p_2$  varies from (0, 0) to  $(\delta/2, 0)$ ,  $d(p, p_3)$  is always greater than or equal to  $\delta$ , and  $d(p, p_5)$  is always greater than  $d(p, p_2)$  (Figure 2 (a), (b)). As  $p_2$  varies from  $(\delta/2, 0)$  to  $(\delta, 0)$ ,  $d(p, p_5)$  is always greater than or equal to  $\delta$ , but  $d(p, p_3)$  can become smaller than  $d(p, p_2)$  (Figure 2 (c), (d)). Thus, p has to be compared with the next three following points in Y.

When p is located somewhere on the bottom edge of  $B_L$  other than the bottom right corner of  $B_L$ , points



Figure 3: Illustration for Subcase B of Case II when  $p \neq (0, 0)$ . (a) When the first two points in  $B_R$  are located on the bottom edge,  $d(p, p_2)$  can be  $\leq \delta$ . (b)-(d) When  $p_1$  is between (0, 0) and  $(\delta/2, 0)$ ,  $d(p, p_4) \geq (p, p_1)$ . When  $p_1$ is between  $(\delta/2, 0)$  and  $(\delta, 0)$ ,  $d(p_3, p)$  can be  $\leq d(p_1, p)$ . (e)-(f)  $p_4$  is at its closest to p. When p is between (0, 0)and  $(-\delta/2, 0)$ ,  $d(p, p_4) \geq d(p, p_1)$ . When p is between  $(-\delta/2, 0)$  and  $(-\delta, 0)$ ,  $d(p, p_4) \geq \delta$ . Thus, Only three points following p in Y need to be considered in the worst case.



Figure 4: Illustration for Subcase C of Case II when p = (0, 0). (a)  $p_3$  is located exactly  $\delta$  from p and  $p_1$ , respectively. (b)  $p_3$  is placed slightly higher than and to the left of that in (a). In both scenarios,  $p_4$  does not need to be taken into consideration, given that  $d(p_3, p_4) \leq d(p, p_4)$ .



Figure 5: Illustration for Subcase C of Case II when  $p \neq (0, 0)$ . The shaded region indicates the possible location of the second point in  $B_R$ .

 $p_1$  and  $p_2$  can be placed at the corners on the bottom edge of  $B_R$ , and  $p_4$  is always farther than  $\delta$  from p(Figure 3 (a)). When  $p_1$  varies from (0, 0) to ( $\delta/2$ , 0), as illustrated in Figure 3 (b) and (c),  $d(p, p_4)$  is always greater than  $d(p, p_1)$ . As  $p_1$  changes from ( $\delta/2$ , 0) to ( $\delta$ , 0),  $d(p, p_4)$  is always greater than or equal to  $\delta$ , but  $d(p, p_3)$  can become smaller than  $d(p, p_1)$  (Figure 3 (d)). The worst case happens when  $p_4$  is located such that it has the smallest x- and y-coordinates possible (i.e.,  $p_4$  is at its closest to p), as shown in Figure 3 (e). In such case, when p is between (0, 0) and ( $-\delta/2$ , 0),  $d(p, p_4) \geq d(p, p_1)$ . When p is between ( $-\delta/2$ , 0) and ( $-\delta$ , 0),  $d(p, p_4)$  is always greater than or equal to  $\delta$ (Figure 3 (f)).

Altogether, in Subcase B, only three following points after p in Y need to be examined.

**Subcase C:** Two points after p are in  $B_R$ . As shown in Figure 4, with the assumption that p is located at

the bottom right corner of  $B_L$ , the shaded region corresponds to possible locations of  $p_4$  such that  $d(p, p_4) \leq \delta$ . The chosen location of  $p_2$  is of the smallest y-coordinate, so that the area of the shaded region is maximized (i.e., worst case). A different location of  $p_2$  would only diminish the shaded region. In addition,  $p_3$  is placed such that  $d(p, p_3) \geq \delta$  and  $d(p_1, p_3) \geq \delta$ .

**Lemma 3** There exists a configuration of points in Y such that  $p_4$  is closer to p than any of  $\{p_1, p_2, p_3\}$ .

**Proof.** Refer to Figure 4. 
$$\Box$$

**Lemma 4** If  $p_4$  is closer to p than any of  $\{p_1, p_2, p_3\}$  then  $d(p_3, p_4) \leq d(p, p_4)$ .

**Proof.** At first, consider the scenario when p,  $p_1$ , and  $p_3$  form an equilateral triangle of side  $\delta$ , as shown in Figure 4 (a). Let  $p_4$  be any point in the shaded region,



Figure 6: Illustration for Case III. (a) When  $p_1$  is between (0, 0) and  $(\delta/2, 0)$ ,  $d(p, p_1) \leq d(p, p_4)$ . (b) When  $p_1$  is between  $(\delta/2, 0)$  and  $(\delta, 0)$ ,  $d(p, p_3)$  can be  $< \delta$ .

let a be the intersection point of the two circular arcs (of radius  $\delta$ ) in  $B_R$ , and let o be the intersection between line segment  $p_3a$  and l. Obviously,  $d(p_3, a) = \delta$ , and, for any point b on the open line segment oa,  $d(p_3, b) \leq d(p, b)$ . Let b be the intersection point between  $pp_4$  and  $p_3a$ . Then,  $d(p, p_4) = d(p, b) + d(b, p_4) \geq d(p_3, b) + d(b, p_4) \geq d(p_3, p_4)$ . As a result,  $(p, p_4)$  does not need to be considered when the current point in Y is p.

Assume now that  $p_3$  is moved upwards and to the left, while having  $d(p_1, p_3) \ge \delta$  (Figure 4 (b)). Consider a horizontal line passing through  $p_3$ . Point o is the intersection of the horizontal line with l, point b is where  $d(p_1, b) = \delta$ , and point a is where  $d(p, a) = \delta$ . Let c be a point on line segment oa. Notice that  $\angle p_3 po \le \pi/4$ . Thus,  $d(p_3, o) \le d(p, o)$ . Given that  $d(p, p_3) \ge \delta$  and ab is parallel with  $pp_1$ ,  $d(a, b) = d(p, p_1) = \delta$  (since  $d(p_1, b) = d(p, a) = \delta$ ). So,  $d(p_3, a) \le \delta$ . As a result, for any point c on the open line segment oa,  $d(p_3, c) \le d(p, c)$ . This implies that, for any point  $p_4$  in the shaded region,  $d(p_3, p_4) \le d(p, p_4)$ . Consequently,  $(p, p_4)$  does not need to be checked.

When p is placed to the left of the lower right corner of  $B_L$ , as illustrated in Figure 5, the second point in  $B_R$  can be located in the shaded region with a distance  $\leq \delta$  (i.e.,  $p_2$ ,  $p_3$ , or  $p_4$ ). Consider the case that the second point in  $B_R$  is  $p_4$ , and a is then  $p_3$ . We claim that  $d(p_3, p_4) \leq d(p, p_4)$ , which can be proven using a similar argument as that in Lemma 4.

Hence, in Subcase C, the current point p has to be compared to only the next three points in Y.

## Case III: One point after p is in $B_L$

If p is situated at the lower right corner of  $B_L$ , the argument is essentially the same as that in Case I, but

without the two points at the top edge of  $B_L$ . Thus, only three points following p in Y have to be examined.

Consider that p is located away from the bottom right corner of  $B_L$ . If there are four points in  $B_R$ , only two following points after p need to be checked, given that the first two points in  $B_R$  are located on the bottom edge of  $B_R$  (i.e., one at each lower corner). Suppose that there are three points in  $B_R$ . As shown in Figure 6 (a), when  $p_1$  is located between (0, 0) and ( $\delta/2$ , 0),  $d(p, p_1) \leq d(p, p_4)$ . When  $p_1$  is between ( $\delta/2$ , 0) and ( $\delta$ , 0),  $p_3$  can be less than  $\delta$  from p. Hence, only the three following points after p in Y must be checked in the worst case.

This concludes the proof of Theorem 1.

#### 3 Acknowledgment

The authors would like to thank Bogdan Armaselu and Shane St. Luce for their insightful comments on the problem. They also thank Balaji Raghavachari for raising the question of whether the fifth point after the current point in the *y*-sorted array can ever be relevant. This work has been partially supported by NSF award IIP1439718 and CPRIT award RP150164.

#### References

- J. L. Bentley and M. I. Shamos. Divide-and-conquer in multidimensional space. In *Proceedings of the Eighth Annual ACM Symposium on Theory of Computing*, pages 220–230, 1976.
- [2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2009.
- [3] F. P. Preparata and M. I. Shamos. Computational Geometry: An Introduction. Springer-Verlag New York, Inc., 1985.