Computing a Centerpoint of a Finite Planar Set of Points in Linear Time

Shreesh Jadhav
Asish Mukhopadhyay *
email: am@iitk.ernet.in

Abstract

The notion of a centerpoint of a finite set of points in two and higher dimensions is a generalisation of the concept of the median of a (finite) set of points on the real line. In this paper, we present an algorithm for computing a centerpoint of a set of n points in the plane. The algorithm has complexity $O(n)$ which significantly improves the $O(n \log^3 n)$ complexity of the previously best known algorithm. We use suitable modifications of the ham-sandwich-cut algorithm and the prune-and-search technique to achieve this improvement.

1 Introduction.

We all have an intuitive idea as to what phrases like "the very center of the square" or "the very center of the city" mean. The notion of the center of a set of points is an attempt to capture this intuition in a quantitative way [YB61]. The center of a set S of n points in a d-dimensional Euclidean space is the set of points c such that, for any hyperplane containing c there are at least $\lceil n/(d+1) \rceil$ of the points of S in each closed half-space determined by the hyperplane. A centerpoint of S is a member of this set. We can alternatively view this definition of a centerpoint as a generalisation of the familiar concept of the median of a set of real numbers. As shown in [Ede87], a centerpoint always exists.

Just as the computation of the median is an important subproblem in many algorithms involving a finite set of real numbers, the computation of a centerpoint of a finite set of points in higher dimensions will be of fundamental importance for geometric algorithms that require a balanced partitioning of the given point set.

The concept of center is also closely related to the concept of the k-hull of a set of points, S, which generalises the notion of the convex hull of S. The k-hull of S is the set of points p such that for any hyperplane containing p there are at least k points of S in each closed half-space determined by the hyperplane. The center of S is its $\lceil n/(d+1) \rceil$-hull. This connection was exploited by Cole et al to construct an $O(n \log^5 n)$ algorithm for computing a centerpoint [CSY87]. In a subsequent paper, by using a refinement of Megiddo's parametric searching technique, Cole improved this to $O(n \log^3 n)$ [Meg83a, Col87].

If we are prepared to settle for less, viz. an approximate centerpoint, then several fast algorithms exist. Megiddo showed that we can find an approximate centerpoint for a planar set of points in linear time [Meg85]. Any line containing this approximate centerpoint has at least $\lceil n/4 \rceil$ points in each closed half-plane determined by it. Indeed we can generalise his method to obtain a linear time algorithm in higher dimensions also. Matoušek showed how an approximate centerpoint, as close to the exact one as we want, can be found in linear time [Mat91].

In this paper we propose an optimal algorithm for computing a centerpoint of a planar set of points. We use an interesting modification of Megiddo's prune-and-search technique [Meg83b]. The modification consists of adding a few extra points in each pruning step so that we retain a subspace of the original solution space, while ensuring a net deletion of points. In addition to this we use the linear time ham-sandwich-cut algorithm of Lo and Steiger [LS91] as an oracle.

The paper is organised as follows. In Section 2 we introduce a few pertinent definitions and discuss the existence of a centerpoint. Section 3 contains a discussion on what points to prune. In the next section we describe the method used to find these points. The algorithm is presented in Section 5. Section 6 contains an analysis of the time complexity of the algorithm. Conclusions and directions to further research are given in

*Department of Computer Science, Indian Institute of Technology, Kanpur.

9th Annual Comptutational Geometry,5/93/CA,USA

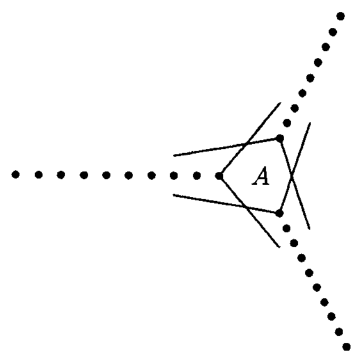

Figure 1: An Extreme Configuration of 30 Points

Section 7. In an appendix we modify our algorithm to show that it suffices to use ham-sandwich-cuts for separable sets only [Meg85]. This gives a more efficient algorithm than when we use the ham-sandwich-cut algorithm of Lo and Steiger.

2 Preliminaries.

Let S be a set of n points in a d-dimensional Euclidean space. The center A of S is a subset E^d such that any closed half-space intersecting A contains at least $\lceil n/(d+1) \rceil$ points of S. From this definition of A we can alternatively characterise the center as the intersection of all open half-spaces such that their complements contain less than $\lceil n/(d+1) \rceil$ points.

A centerpoint of S is a member of A.

From the definition of the center above, it follows that any hyperplane which contains a centerpoint has at least $\lceil n/(d+1) \rceil$ points of S in each closed half-space determined by it.

A centerpoint exists for every configuration of points. We have the following theorem which can be proved using Helly's theorem.

Theorem 2.1 [Ede87] *Every finite set of points in d-dimensional Euclidean space admits a centerpoint.*

The bound $\lceil n/(d+1) \rceil$ in the definition of the centerpoint above is tight. We can always find a configuration of n points in E^d such that the center is empty if $\lceil n/(d+1) \rceil$ is replaced by a larger number.

We show a configuration of 30 points in the plane for which all points in E^2 admit a line which contains no more than 10 points on one of its sides (Fig. 1).

We will use the notation S_H to denote the set of points of S contained in the half-plane H i.e. the set $S \cap H$, and S_{GH} to denote the set of points of S common to the half-planes G and H i.e. the set $S \cap G \cap H$.

3 What to Prune.

We would attempt to compute a centerpoint by applying prune-and-search. An application of Megiddo's prune-and-search paradigm to a given problem consists of constructing a reduced search space, which contains the solution(s) to the problem, and then throwing away a constant fraction of the input set with respect to this reduced search space.

We cannot attempt to compute the center of a set of points by this technique because we cannot guarantee that a pruned set has the same center as the original set. However, it might be possible to do something less. We can try to prune away a fraction of points of the original set, ensuring that the center of the pruned set is contained in the center of the original set. Then the relevant question is, which points to prune. To answer this, let us try to understand how we can instead add points so that the center of the enlarged set is a superset of the center of the original set.

A triangle has the well-known property that any line which intersects it has at least one of the vertices of the triangle in each closed half-plane determined by it. Suppose that the center of S is contained inside a triangle. If we enlarge S by adding to it the vertices of this triangle then the center of S is contained in the center of this bigger set. We prove this in the theorem below.

Theorem 3.1 *If P is the set of vertices of a triangle T, which contains the center A of S, then A is a subset of the center of $S \cup P$.*

Proof: By the definition of the center every closed half-plane H intersecting A contains at least $\lceil n/3 \rceil$ points of S.

Since the triangle T contains the center A, the half-plane H intersects T. Therefore H contains at least one of the vertices of P, and the number of points of

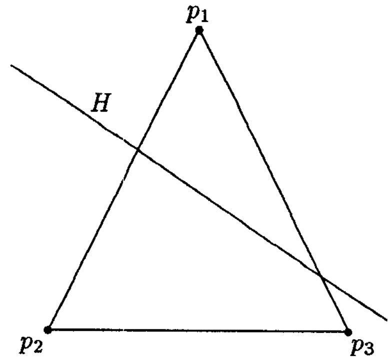

Figure 2: Deletion of Triplets of Points From Set S

$S \cup P$ in H is at least $\lceil n/3 \rceil + 1$. Thus A is contained in the center of $S \cup P$. ∎

The next theorem tells us how we can find such a triangle. We first note that if a closed half-plane contains a centerpoint of S then it necessarily contains at least $\lceil |S|/3 \rceil$ points.

Let $|S \cup P| = n$ in the rest of this section.

Theorem 3.2 *Let H_1, H_2, H_3 be three closed half-planes, each containing less than $\lceil n/3 \rceil$ points of $S \cup P$ and situated so that the intersection of their complements is a bounded open triangle A; $P = \{x_1, x_2, x_3\}$, where $x_1 \in H_1 \cap H_2, x_2 \in H_2 \cap H_3$ and $x_3 \in H_3 \cap H_1$. Then any centerpoint of S is also a centerpoint of $S \cup P$.*

Proof: Clearly, the triangle formed by the set of points in P encloses the triangle A formed by the boundaries of the half-planes, and hence also the center of S. The result follows from the previous theorem (Fig. 2). ∎

For the purpose of our algorithm we need to be able to use the above theorem in a "subtractive" way, that is, start with the set S and get to the set $S - P$. We would also like to do it many times, so that a fraction of the points of the set we start with can be discarded. This however may not always be possible. There are configurations of points for which we cannot determine P for any choice of half-planes satisfying the conditions of our theorem. An example of such a configuration is n points evenly arranged on a circle.

To get around this difficulty we enlarge the scope of the above theorem, allowing for the choice of four half-planes instead of three. The details are provided in the theorem below.

Theorem 3.3 *Let H_1, H_2, H_3, H_4 be four closed half-planes, each containing less than $\lceil n/3 \rceil$ points of $S \cup P$ and situated so that the intersection of their complements is a bounded open quadrilateral or triangle A; $P = \{x_1, x_2, x_3, x_4\}$, where $x_1 \in H_1 \cap H_2, x_2 \in H_2 \cap H_3, x_3 \in H_3 \cap H_4$ and $x_4 \in H_4 \cap H_1$. Then any centerpoint of $S \cup \{p\}$ is also a centerpoint of $S \cup P$ where p is either the intersection of the diagonals of the quadrilateral P if P defines a quadrilateral or it is the interior point among the points of P if P defines a non-convex quadrilateral.*

Proof: Let c be a center-point of $S \cup \{p\}$. Any closed half-plane H containing c contains at least $\lceil n/3 - 1 \rceil$ points of $S \cup \{p\}$ ($|S \cup \{p\}| = n - 3$).

Our first observation is that c lies inside A. To see how, suppose this is not the case. Then c lies in at least one of the half-planes H_1, H_2, H_3 and H_4. Let H_i be that half-plane. Also, let H_c be that closed half-plane which is a proper subset of H_i and contains c in its boundary. Since H_i contains less than $\lceil n/3 \rceil$ points of $S \cup P$, therefore $H_c \cap (S \cup \{p\})$ contains less than $\lceil n/3 \rceil - 2 + 1$ points i.e. when two points of P are dropped and the point p is added to H_c. This leads to a contradiction as every closed half-plane containing c contains at least $\lceil n/3 - 1 \rceil$ points of $S \cup \{p\}$.

We have to show that a hyperplane H which contains c contains at least $\lceil n/3 \rceil$ points of $S \cup P$. Depending on whether the points in P form a convex or a non-convex quadrilateral, three different cases arise.

Case 1 The four points in P form a non-convex quadrilateral.

This case is a trivial application of *Theorem 3.1*. The three convex vertices form a triangle that encloses the center of $S \cup \{p\}$. Thus by *Theorem 3.1* any centerpoint of $S \cup \{p\}$ is also a centerpoint of $S \cup P$.

Case 2 The four points of P form a convex quadrilateral but the intersection point p of the diagonals does not belong to H (Fig. 3).

Since the quadrilateral contains the center of $S \cup \{p\}$ the half-plane H intersects it. Thus H contains at least $\lceil n/3 \rceil$ points of $S \cup P$ since it contains

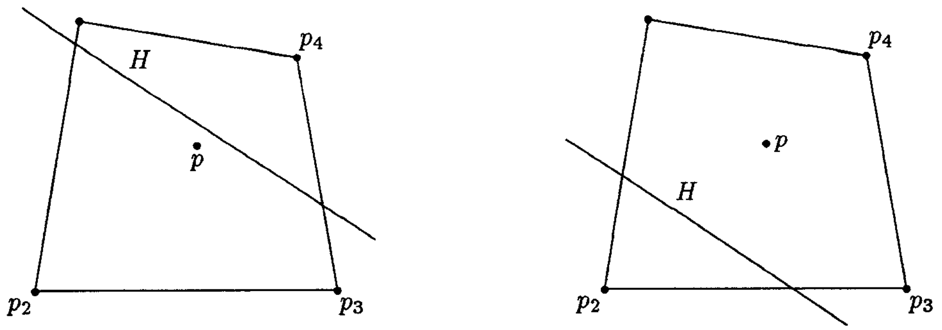

Figure 3: Substitution of Quadruples of Points by Point p in S

at least one of the four vertices of the quadrilateral.

Case 3 The four points of P form a convex quadrilateral and the intersection point of the diagonals belongs to H (Fig. 3).

In this case H contains at least two points of P. We can therefore delete p from H and still claim that H contains at least $\lceil n/3 \rceil$ points of $S \cup P$.

Thus in all the cases H contains at least $\lceil n/3 \rceil$ points of $S \cup P$. Since the total number of points in $S \cup P$ is n, c is a center-point of $S \cup P$ as well. ∎

The last theorem is the cornerstone of our pruning mechanism. In the next section we will show how to use ham-sandwich-cuts to make a clever choice of four half-planes so that we can prune a fraction of of the input set by applying this theorem repeatedly.

At this point it would be pertinent to point out a salient and novel feature of the last theorem. Unlike any previous application of the prune-and-search paradigm, in this case, we will have to add new points to the input set, ensuring at the same time a net deletion of points in each iteration.

4 How to Prune.

To live up to the claim we made in the last section, we will find a set of four closed half-planes L, R, U and D that satisfy the conditions of the last theorem so that a fraction of the input set can be pruned.

We fix L as follows. We determine the $(\lceil n/3 \rceil - 1)$ smallest of the orthogonal projections of all the points of S onto an arbitrary reference line. The closed half-plane, determined by the perpendicular line passing through this point, that contains $\lceil n/3 \rceil - 1$ points on the left is chosen to be L. This takes $O(n)$ time.

The half-planes U and D are determined by using a slightly generalised form of the ham-sandwich-cut algorithm for separable point sets. We show how to determine U only. The half-plane D is computed in a similar manner. Let A be the set of points in L and B the rest of the points of S. We use the ham-sandwich-cut algorithm to partition set A in the ratio 1:3 and set B in the ratio 3:5 so that one of the closed half-planes determined by the partioning line contains exactly $\lceil n/3 \rceil - 1$ points of S. This again takes $O(n)$ time. We ensure at the same time that each of the sets $L \cap U$ and $L \cap D$ contains $\lfloor n/12 \rfloor - 1$ points.

Determining the half-plane R is somewhat tricky. We must make sure that each of the sets $R \cap D$ and $R \cap U$ contains at least $\lfloor n/12 \rfloor - 1$ points and R contains less than $\lceil n/3 \rceil$ points. Let T be the set $(\bar{U} \cap \bar{D}) \cup (U \cap D)$. Add the points in $U - D$ and $D - U$ twice respectively to T to form the multisets A and B. Both of these multisets contain n points.

To see this, suppose that $U - D$ contains x points of S. Then the sets $U \cup D$ and $\bar{U} \cap \bar{D}$ contain $\lceil n/3 \rceil - x - 1$ and $\lfloor 2n/3 \rfloor - x + 1$ points of S respectively. Since the points in $U - D$ are counted twice the total number of points in A is $2x + \lceil n/3 \rceil - x - 1 + \lfloor 2n/3 \rfloor - x + 1 \; (= n)$.

The points in T are included both in A and B so these are also counted twice. Using the algorithm of Lo and Steiger [LS91] we now make a ham-sandwich-cut of the sets A and B, and choose R as that closed half-plane which contains $\lceil n/3 \rceil - 1$ points of A and $\lceil n/3 \rceil - 1$ points of B. Since each point is counted twice the total number of points of S in R is $\lceil n/3 \rceil - 1$. We prove in a subsequent section that both the sets $R \cap U$ and $R \cap D$ contain at least $\lfloor n/12 \rfloor - 1$ points.

Remark: In an appendix to this paper we sketch a modification of the above procedure which uses the ham-sandwich-cut algorithm for separable sets only.

Moreover, for non-separable sets, a ham-sandwich-cut with an arbitrary ratio may not always exist. However, in the case mentioned above, the R-cut always exist. We give a proof of this in another appendix.

Assuming that each of the sets $L \cap U$, $L \cap D$, $R \cap U$ and $R \cap D$ contains at least $\lfloor n/12 \rfloor - 1$ points of S, it is clear what we must do. We select any quadruple

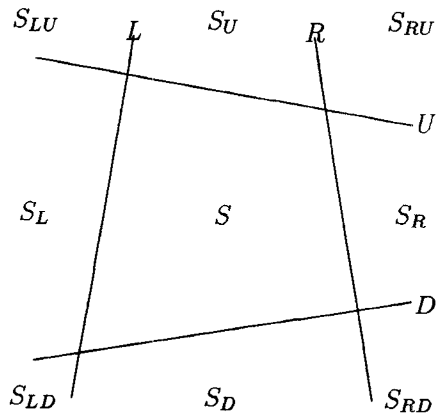

Figure 4: Partitioning of Set S by Half-Planes U, R, D, and L

of points, one each from these sets. If these form a convex quadrilateral we delete them from S and add to it the intersection of the diagonals. Otherwise we delete the convex vertices but retain the concave one. We can repeat this procedure on the reduced set of points thus obtained until one of the sets S_{LU}, S_{RU}, S_{RD} and S_{LD} is empty, since the half-planes L, U, D and R continue to satisfy the conditions of the theorems of the previous section. We call it `REPLACE-POINTS(`P`: Points)`, where P is the set of all such points. We note that this reduces the size of S by at least one fourth.

5 The Center-Point Algorithm.

After the discussion in the previous sections, the procedure for finding a centerpoint is now quite clear.

In each iteration we compute the points that are to be discarded or replaced. By throwing away these points we reduce the size of the set by a non-zero fraction.

When the size of the set becomes so small that no more points can be discarded we call a halt to the pruning procedure and compute a centerpoint by any straightforward algorithm.

The formal algorithm is given below.

```
Algorithm CENTERPOINT(S: Points)
begin

  repeat
     Compute the half-planes L, U, D and R
     Compute the sets S_L, S_U,S_D and S_R
     P ← S_LU ∪ S_LD ∪ S_RU ∪ S_RD
     REPLACE-POINTS(P)
  until P = φ

  Find a centerpoint by any straightforward
  method
end.
```

6 Analysis of the Center-Point Algorithm.

We first show that each of the sets S_{RU} and S_{RD} contains at least $\lfloor n/12 \rfloor - 1$ points. By construction we know that the each set, S_{LU} and S_{LD}, contains $\lfloor n/12 \rfloor - 1$ points.

Let S'_U and S'_D denote the sets $S_U - S_{LU}$ and $S_D - S_{LD}$ respectively. Let S' denote the set $S - \{S_L \cup S_U \cup S_D\}$.

Theorem 6.1 *Each of the sets S_{LU}, S_{RU}, S_{RD} and S_{LD} contains at least ϵn points of S, where* $0 < \epsilon < 1$.

Proof: We have already seen that each of the sets S_{LU} and S_{LD} contains $\lfloor n/12 \rfloor - 1$ points of S.

Each of the sets S'_U and S'_D (inclusive of S_{RU} and S_{RD} respectively) contains $\lceil n/3 \rceil - \lfloor n/12 \rfloor$ points.

The sets S_L, $S'_U \cup S'_D$ and S' are disjoint. Therefore

$$|S_L| + |S'_U \cup S'_D| + |S'| = n$$

If S_c denotes the set $S'_U \cap S'_D$, from the above equation we have

$$|S'| - |S_c| = n - (3\lceil n/3 \rceil - 1] - 2\lfloor n/12 \rfloor)$$

Now there are two cases to consider. When $S_{RU} \cup S_{RD}$ contains S_R, both S_{RU} and S_{RD} contain at least $\lfloor n/6 \rfloor - 1$ points. In the second case the set S_c includes the set S_R and we have

$$|S'| + |S_{RU} \cup S_{RD}| \geq \lceil n/3 \rceil - 1$$

Therefore,

$$\begin{aligned} & |S_{RU}| + |S_{RD}| \\ & \geq |S_c| + |S_{RU} \cup S_{RD}| \\ & \geq 4\lceil n/3 \rceil - n - 2\lfloor n/12 \rfloor - 2 \end{aligned}$$

of points, so

$$|S_{RU}| = |S_{RD}| \geq \lfloor n/12 \rfloor - 1$$

Thus all the sets in question contain at least $\lfloor n/12 \rfloor - 1$ points. We have also shown that ϵ is approximately 1/12. ∎

Combining the earlier theorems we get the result

Theorem 6.2 *A point in the center of a set can be computed in linear time.*

Proof: In each iteration at least $3n/12$ ($= n/4$) points are deleted.

If $T(n)$ is the running time of the algorithm for an input of size n, then it satisfies the following recurrence.

$$\begin{aligned} T(n) &= T(n - 3n/12) + O(n) \\ \Rightarrow \quad T(n) &= T(3n/4) + O(n) \end{aligned}$$

Since $T(n) = O(n)$ from the above recurrence, the claim of the theorem follows. ∎

7 Conclusions.

We have presented an optimal algorithm for computing a centerpoint of a finite set of points in the plane, thus providing one more example of the power and versatility of the Prune-And-Search paradigm.

It would be worth exploring how this speeds up algorithms which uses the centerpoint computation as a basic subroutine.

Acknowledgements

The comments made by Tamal K. Dey, Edgar Ramos and Gunter Röte are gratefully acknowledged. We thank all of them for having pointed out several errors in the initial drafts of the paper. We also thank Herbert Edelsbrunner for his encouragement.

References

[Col87] R. Cole. Slowing down sorting networks to obtain faster sorting algorithms. *J. ACM*, 34(1):200–208, 1987.

[CSY87] R. Cole, M. Sharir, and C. Yap. On k-hulls and related problems. *SIAM J. Comput.*, 16:61–77, 1987.

[Ede87] H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer Verlag, 1987.

optimal geometric algorithms using partial sorting networks. Technical Report TRCS-93-165, Indian Institute of Technology, Kanpur, 1993. Submitted to the Third National Seminar on Theoretical Computer Science, Kharagpur, India.

[LS91] Chi Yuan Lo and William Steiger. An optimal-time algorithm for ham-sandwich cuts in the plane. In *Proc. of the Second Canadian Conference on Computational Geometry*, pages 5–9, 1991.

[Mat91] Jiří Matoušek. Approximations and optimal geometric divide-and-conquer. In *Proc. 23rd Ann. ACM Symposium on Theory of Computing*, pages 505–511, 1991.

[Meg83a] N. Megiddo. Applying parallel computation algorithms in the design of serial algorithms. *J. ACM*, 30(4):751–758, 1983.

[Meg83b] N. Megiddo. Linear time algorithms for linear programming in R^3 and related problems. *SIAM J. of Comput.*, 12(4):759–776, 1983.

[Meg85] N. Megiddo. Partitioning with two lines in the plane. *J. Algorithms*, 3:430–433, 1985.

[YB61] I. M. Yaglom and V. G. Boltyanskiĭ. *Convex Figures.* Holt, Rinehart and Winston, 1961.

A A Proof of Existence of an R-Cut.

In this section we prove the existence of a ham-sandwich-cut for a ratio other than 1:1. We give a sufficient condition for this which is applicaple to the configuration discussed in this paper. It can also be easily seen how the algorithm by Lo and Steiger can be wmodified for this application.

Suppose we want to find a ham-sandwich-cut which divides two given sets A and B in ratios r_A and r_B respectively. Then,

Lemma A.1 *If there exist two cuts H^1 and H^2 which divide A, B in ratios r_A^1, r_B^1 and r_A^2, r_B^2 respectively such that*

$$r_A^1 \geq r_A \geq r_A^2$$

and

$$r_B^1 \leq r_B \leq r_B^2$$

then there exists a ham-sanwich-cut which divides the set A and B in ratios r_A and r_B respectively.

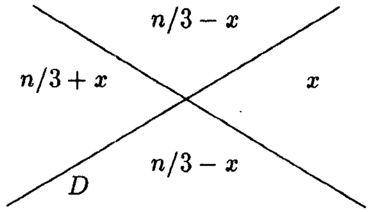

Figure 5: Distribution of Points for an R-Cut.

Proof: The proof of this Lemma is constructive. We dualise the problem of ham-sandwich-cut[LS91]. In the dual problem we have to find intersection of A-level corresponding to r_A and B-level corresponding to r_B. A k^{th}-level is defined as the piece-wise linear curve that intersects every vertical line at k^{th}-median.

The point corresponding to H^1 is below the A-level corresponding to r_A and above the B-level corresponding to r_B. At the point corresponding to H^2 the reverse is true. As both of these levels interchange their relative positions they will intersect in the vertical strip between the dual points of H^1 and H^2. This intersection point is the required dual of the general ham-sandwich-cut. ∎

Let the boundaries of U and D be the cuts H_1 and H_2 respectively as stated in *Lemma A.1.* The ratios r_A and r_B are both 1/2. Let the size of S_{UD} be x. The sizes of other sets would be as shown in the Fig. 5. For the sake of simplicity floor and ceiling functions are dropped from various terms. Then,

$$
\begin{aligned}
r_A^1 &= \frac{2n/3 - x}{n/3 + x} \\
r_A^2 &= \frac{x}{n - x} \\
r_B^1 &= \frac{x}{n - x} \\
r_B^2 &= \frac{2n/3 - x}{n/3 + x}
\end{aligned}
$$

Clearly, for $0 \leq x \leq n/3$,

$$r_A^1 \geq r_A \geq r_A^2$$

and

$$r_B^1 \leq r_B \leq r_B^2$$

Therefore an R-cut exists for the said configuration in the paper.

B An Alternative Pruning Procedure.

Instead of computing R by the method given in section 4, we can use the following alternative procedure. The half-planes U and D are computed as before. But to compute R we start with the half-plane U instead of L. We find a ham-sandwich-cut such that $U \cap R$ contains $\lfloor n/12 \rfloor - 1$ points of S. Further, let R and D be computed maximally such that it may not be possible to locate another half-planes which have S_{RU} and S_{LD} respectively as proper subsets.

By construction, the sets S_{LU}, S_{LD} and S_{RU} contain $\lfloor n/12 \rfloor - 1$ points each. We have to show that S_{RD} contains at least $\lfloor n/12 \rfloor - 1$ points.

Theorem B.1 *There are at least $\lfloor n/12 \rfloor - 1$ points in S_{RD}.*

Proof: There are several different cases to consider. We prove the statement for the configuration shown in Fig. 6 only. For the sake of simplicity ceiling and floor functions are dropped from various terms.

Let the size of S_{DU} be x and size of S_{RD} be $n/12 + x - y$.

Then number of points in every partition would be as shown in the Figure. All the sizes of sets must be greater than or equal to zero. Therefore,

$$x - y > 0$$

and hence,

$$S_{RD} = n/12 + x - y > n/12$$

We can prove the statement of theorem for other configurations of hyperplanes L, U, D and R similarly. ∎

Since all the ham-sandwich-cuts in this procedure are computed for separable sets, this algorithm is more efficient than the one described in the paper.

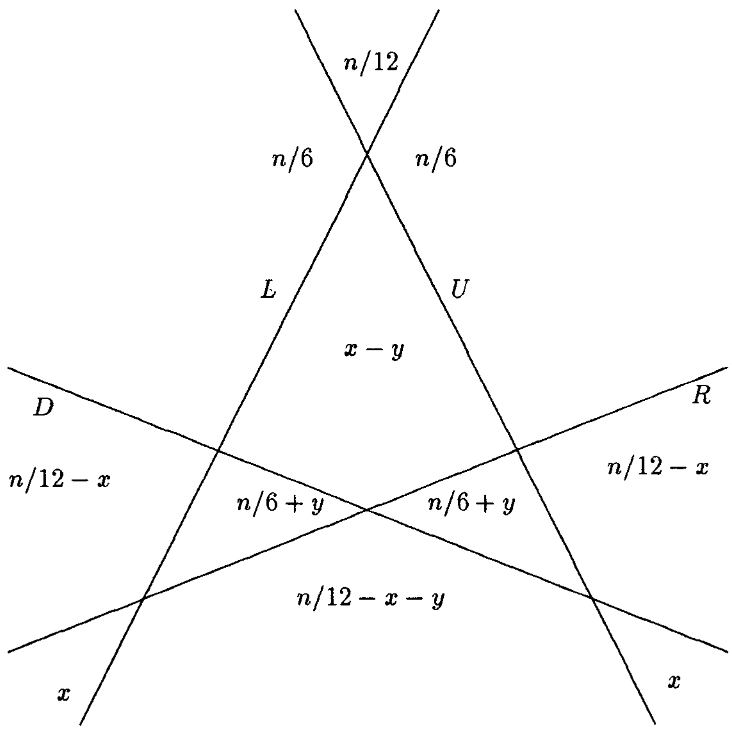

Figure 6: Distribution of Points for the Configuration in the Proof of *Theorem B.1.*