# An Optimal Generalization of the Centerpoint Theorem, and its Extensions

Nabil H. Mustafa
Max-Planck Institute
Saarbrucken, Germany.
nmustafa@mpi-sb.mpg.de

Saurabh Ray
Max-Planck Institute
Saarbrucken, Germany.
saurabh@mpi-sb.mpg.de

## ABSTRACT

We prove an optimal generalization of the centerpoint theorem: given a set $P$ of $n$ points in the plane, there exist two points (not necessarily among input points) that hit all convex objects containing more than $4n/7$ points of $P$. We further prove that this bound is tight. We get this bound as part of a more general procedure for finding small number of points hitting convex sets over $P$, yielding several improvements over previous results.

## Categories and Subject Descriptors

G.2 [**Combinatorics**]: Discrete Geometry

## General Terms

Algorithms, Theory

## Keywords

Combinatorial geometry, weak $\epsilon$-nets, centerpoint theorem, discrete geometry, extremal methods, hitting convex sets

## 1. INTRODUCTION

The centerpoint theorem is one of the fundamental combinatorial results in discrete geometry, with applications in geometric algorithms [4, 9, 11], large-scale computing [8], multivariate data analysis [5] and several others. It states the following:

**Centerpoint Theorem [10, 6].** Given a set $P$ of $n$ points in the plane, there exists a point $c$ such that any convex object containing more than $2n/3$ points of $P$ contains $c$[1]. Furthermore, this bound is tight.

In this paper we look at a generalization of the above statement to more than one point. For example, is it possible to find two points $c_1$ and $c_2$ in the plane such that any convex object containing at least $n/2$ points must contain either $c_1$ or $c_2$? We present a general procedure that gives the following results: one can hit all convex objects containing more than $4n/7$ points with 2 points. Furthermore, we prove that this bound is tight. Similar results are derived for larger number of points. In particular, we show that if each convex object contains more than $20n/41$ points, then five points suffice. This improves a natural way of adding five points [2] which gives the worse $n/2$-bound: find two lines (using the ham-sandwich theorem [6]) which partition the point set into four regions with $n/4$ points in each. Add the intersection point of the lines along with the centerpoints of the four regions.

## Related results

Aronov *et al.* [2] prove that all convex objects containing greater than $5n/8$ points of $P$ can be hit by two points. They also construct inputs where regardless of how one picks the two points, there exists a convex object containing at least $5n/9$ points that is not hit. In this paper, we improve both their results to get the optimal result of $4n/7$. We similarly improve their results for all small numbers of points (see Section 3 for specific improvements).

Our problem is related to two other areas of work. In the weak $\epsilon$-net problem [1, 3, 7], given a parameter $\epsilon > 0$ and a point set $P$, one would like to compute a small set of (not necessarily input) points that hit all convex objects containing at least $\epsilon n$ points of $P$. Clearly, for $\epsilon > 2/3$, the centerpoint is the desired weak $\epsilon$-net. Our work can be seen as computing small weak $\epsilon$-nets.

The other related area of work is the so-called *Gallai-type* problems [6], an example of which is the following: Given a set of closed disks in the plane such that every pair intersects, what is the smallest number of points needed to hit all these disks? In this case, the answer which is both necessary and sufficient, is four. In our problem, we are looking to hit considerably more general objects (convex objects), with the added constraint that one first fixes $n$ input points, and each convex object has to contain a constant proportion of these points.

## 2. MAIN THEOREM

We first present some definitions. Given a set $P = \{p_1, \ldots, p_n\}$ of $n$ points in $\mathbb{R}^2$, define the following:

$$\epsilon(P,Q) = \min\{\epsilon \mid |\mathcal{C} \cap Q| \neq \emptyset \, \forall \text{ convex sets } \mathcal{C} \; s.t. \; |\mathcal{C} \cap P| > \epsilon n\}$$

and let $\epsilon_i(P) = \min_{Q,|Q|=i} \epsilon(P,Q)$. Set $\epsilon_i = \sup_P \epsilon_i(P)$. In other words, given any $P$, consider the set of all convex objects containing $\epsilon_i n$ points of $P$. Then they can be hit by $i$ points. These $i$ points are said to form a *weak* $\epsilon_i$-net for $P$. From the centerpoint theorem, it follows that $\epsilon_1 = 2/3$.

---

[1] This theorem can be equivalently stated as: there exists a point $c$ such that any halfspace containing $c$ contains at least $n/3$ points of $P$.

*SCG'07,* June 6–8, 2007, Gyeongju, South Korea.

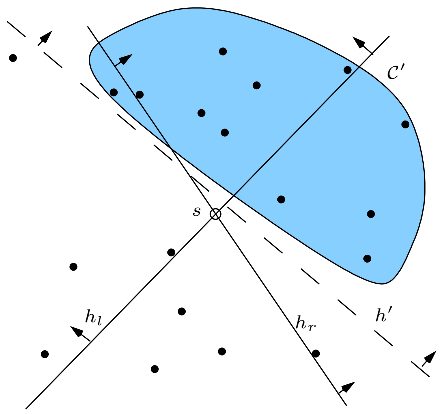


**Figure 1: Illustration of Theorem 2.1**

For a point $p$, let $x(p)$ and $y(p)$ denote its $x$- and $y$-coordinates. For a convex object $\mathcal{C}$, let $x(\mathcal{C})$ (resp. $y(\mathcal{C})$) denote the smallest $x$-coordinate (resp. $y$-coordinate) of a point in $\mathcal{C}$, i.e. $y(\mathcal{C}) = \min_{p\in\mathcal{C}} y(p)$.
We now present our main result.

THEOREM 2.1. *Given a set $P$ of $n$ points in $\mathbb{R}^2$, and two integers $c \ge 0$ and $d \ge 0$,*

$$\epsilon_{c+2d+1} \le \frac{\epsilon_c \cdot (1+\epsilon_d)}{\epsilon_c + 1 + \epsilon_d \cdot \epsilon_c},$$

*where we define $\epsilon_0 = 1$.*

*Construction.* Let $a$,$b \in [0,1]$ be two reals to be fixed later.
Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be the set of all halfspaces which contain more than $an$ points of $P$. Define $\mathcal{H}^2 = \{(h_i, h_j) \mid |P \cap (h_i \cap h_j)| \ge bn$, where $h_i, h_j \in \mathcal{H}\}$ to be the set of all pairs of halfspaces in $\mathcal{H}$ whose intersection contains at least $bn$ points of $P$. Take the pair, say $h_l$ and $h_r$, such that

1. $(h_l, h_r) \in \mathcal{H}^2$
2. $h_l \cap h_r$ has the highest lowest-intersection point of any pair of halfspaces in $\mathcal{H}^2$, i.e., $y(h_l \cap h_r) = \max_{(h_i,h_j)\in\mathcal{H}^2} y(h_i \cap h_j)$

Now construct and return the set $Q = \{s\} \cup Q_c \cup Q_l \cup Q_r$, where

1. $s = y(h_l \cap h_r)$ is the lowest point common to $h_l$ and $h_r$.
2. $Q_c$ is a $\epsilon_c$-net for the point set $P \setminus (P \cap h_l \cap h_r)$ using $c$ points.
3. $Q_l$ is a $\epsilon_d$-net for the point set $P \setminus (P \cap h_l)$ using $d$ points, and $Q_r$ is a $\epsilon_d$-net for the point set $P \setminus (P \cap h_r)$ using $d$ points.

LEMMA 2.1. *$Q$ is a $a$-net for $P$, and has size $c + 2d + 1$.*

PROOF. The size of $Q$ is obvious, and we show that it is an $a$-net for the value required in the statement of the theorem. We first need the following crucial fact.

CLAIM 2.1. *Let $\mathcal{C}'$ be a convex object which does not contain $s$ and intersects $h_l \cap h_r$. Then, either $|P \cap \mathcal{C}' \cap h_l| < bn$ or $|P \cap \mathcal{C}' \cap h_r| < bn$.*

PROOF. For contradiction, assume that $\mathcal{C}'$ intersects both $h_l$ and $h_r$ in more than $bn$ points of $P$. Since $\mathcal{C}'$ does not contain $s$, by the Separation theorem [6], there exists a halfspace $h'$ such that $\mathcal{C}' \subseteq h'$, and $h'$ does not contain $s$. Since $\mathcal{C}'$ intersects $h_l \cap h_r$, i) $h'$ intersects $h_l \cap h_r$, and ii) $h'$ contains more than $an$ points of $P$, and iii) $|P \cap h' \cap h_l| \ge bn$ and $|P \cap h' \cap h_r| \ge bn$.
Note that any $h'$ which satisfies condition i) above must have either $y(h' \cap h_l) > y(h_l \cap h_r)$ or $y(h' \cap h_r) > y(h_l \cap h_r)$ (See Figure 1). Otherwise, if $h'$ intersects $h_l \cap h_r$, and intersects both $h_l$ and $h_r$ below $y(s)$, then $h'$ must contain $s$, a contradiction. Say $y(h' \cap h_l) > y(h_l \cap h_r) = y(s)$. But then, from conditions ii) and iii) above, we get a contradiction to the fact that $h_l$ and $h_r$ were pair with the highest $y(\cdot)$ value by looking at the pair $h'$ and $h_l$. $\square$

We now show that any convex object $\mathcal{C}'$ containing $an$ points must contain a point of $Q$ by one of these cases:

1. $\mathcal{C}'$ contains $s$, so is hit by $Q$.
2. $\mathcal{C}'$ does not intersect $h_l \cap h_r$. Since $|P \cap (h_l \cap h_r)| \ge bn$, $\mathcal{C}'$ contains $an$ points from the remaining set $P \setminus (P \cap h_l \cap h_r)$, which has size $(1-b)n$. If $an \ge \epsilon_c(1-b)n$, then $\mathcal{C}'$ is hit by $Q_c$.
3. $\mathcal{C}'$ does not contain $s$ and yet intersects $h_l \cap h_r$. Then, by Claim 2.1, either $\mathcal{C}' \cap h_l \le bn$ or $\mathcal{C}' \cap h_r \le bn$. Then it must contain at least $an - bn$ points from either $P \setminus (P \cap h_l)$, or $P \setminus (P \cap h_r)$. If $an - bn \ge \epsilon_d(1-a)n$, then $\mathcal{C}'$ is hit either by $Q_l$ or $Q_r$.

Therefore, if

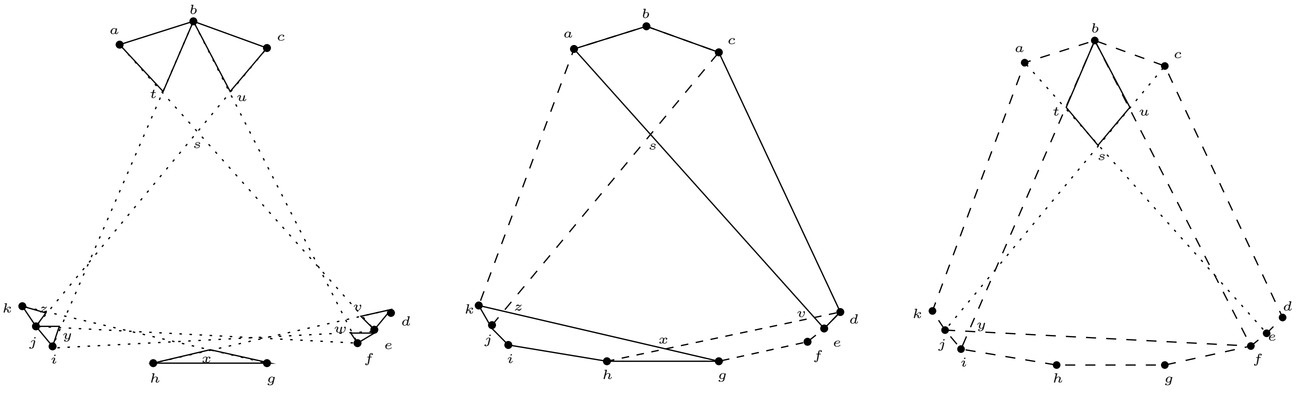
$$an \ge \epsilon_c(1-b)n \quad \text{and} \quad an - bn \ge \epsilon_d(1-a)n \tag{1}$$

then any convex set is hit by $Q$. Solving (1) yields

$$\epsilon_{c+2d+1} \le a = \frac{\epsilon_c \cdot (1+\epsilon_d)}{\epsilon_c + 1 + \epsilon_d \cdot \epsilon_c},$$

completing the proof of Lemma 2.1 and hence Theorem 2.1.

**Remark:** The above method actually gives another elementary proof of the centerpoint theorem in any dimension. The proof for two dimensions, as in the method of Theorem 2.1: consider all halfspaces containing more than $2n/3$ points, and take the pair with the highest lowest-intersection point (i.e., highest $y(\cdot)$ value). This is the required point, since any convex object not containing this point cannot intersect the intersection of the halfspaces (Claim 2.1),



(a) (b) (c)

**Figure 2: (a) One of the seven (bold) triangles contains a point of the weak $\epsilon$-net (b) One of the four triangles $jzk$, $gxh$, $dve$ or $asc$ contains a point of the weak $\epsilon$-net (c) $jyi$ contains a point of the weak $\epsilon$-net**

which contains more than $n/3$ points of $P$. Hence, such a convex object can only contain the remaining points of $P$, of which there are fewer than $2n/3$. This follows from Theorem 2.1 by setting $c = d = 0$ to get $\epsilon_1 = 2/3$! The proof for $d$-dimensions is exactly the same: consider sets of $d$ halfspaces, each of which contains more than $dn/(d+1)$ points and choose the set with the highest lowest-intersection point (w.r.t. any dimension).

## 3. CONSEQUENCES OF MAIN THEOREM

Improving upon previous work [2], we completely resolve the 2-point case.

COROLLARY 3.1. *Given a set $P$ of $n$ points in $\mathbb{R}^2$, the set of all convex objects which contain more than $4n/7$ points of $P$ can be hit by* two *points (i.e., $\epsilon_2 \leq 4/7$). Furthermore, there exist arbitrarily large point sets such that the set of all convex objects containing $4n/7$ points cannot be hit by two points.*

PROOF. The upper bound follows from Theorem 2.1 by setting $c = 1, d = 0$.

For any $n$, we construct a point set $P$ of size $n$ such that for any two given points $p$ and $q$ in the plane there is a convex set which avoids both the points and contains $4n/7$ points of $P$.

Consider the vertices of a regular heptagon each representing a set of $n/7$ points contained in a sufficiently small disk. Let $a, b, c, d, e, f$ and $g$ be the sets in clockwise order. Our set $P$ is the union of these sets.

If one of the points $p$ or $q$ is arbitrarily close to one of the 7 sets, say the set $a$, then the other point cannot hit the convex hulls of the sets $b \cup c \cup d \cup e$, $d \cup e \cup f \cup g$ and $f \cup g \cup a \cup b$ simultaneously since they don't have a common intersection. Now, assume that neither $p$ nor $q$ is arbitrarily close to any of the 7 sets. Consider the line $l$ passing through the points $p$ and $q$. If $l$ does not pass through any of the 7 sets then one of the closed halfspaces defined by $l$ contains 4 of the sets whose convex hull is not hit by either $p$ or $p$. Otherwise, one of the closed halfspaces defined by $l$ contains 3 of sets and they along with one of the sets which $l$ passes through define a set of $4n/7$ points whose convex hull is not hit by either $p$ or $q$. ☐

COROLLARY 3.2. *Given $P$, the set of all convex objects which contain more than $8n/15$ points of $P$ can be hit by* three *points (i.e., $\epsilon_3 \leq 8/15$). Furthermore, there exist arbitrarily large point sets such that the set of all convex objects containing $5n/11$ points cannot be hit by three points.*

PROOF. The upper bound follows from Theorem 2.1 by setting $c = 2, d = 0$. The lower bound construction is as follows.

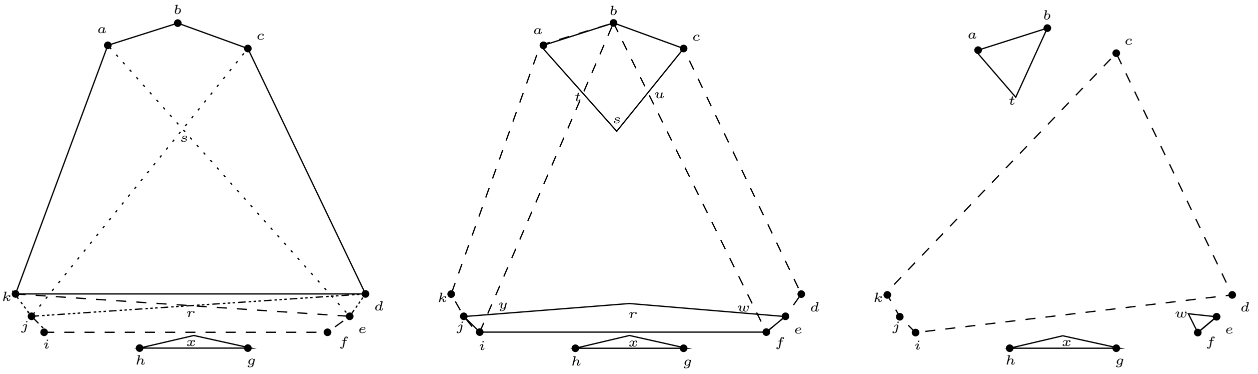
For any $n$, we construct a point set $P$ of size $n$ such that for any three given points in the plane there is a convex set containing $5n/11$ points of $P$ which avoids all the three points. Figure 2(a) shows such a point set. Each of the 11 points in the figure represents a set of $n/11$ points contained in a sufficiently small disk.

Assume that there are three points which hit all convex sets containing $5n/11$ points of $P$. We first show that these points cannot be arbitrarily close to any of the 11 sets in the point set. Observe that if all the three points are arbitrarily close to one of the 11 sets in the point set, then they cannot hit the convex region formed by the rest of the 8 sets. Also, if two of the points $p, q$ and $r$ are arbitrarily close to one of the sets, then the rest of the 9 sets can be used to define two convex sets containing $5n/11$ points each and sharing only one of the 11 sets. A single point hitting both these sets should be arbitrarily close to the shared set implying that all the three points are arbitrarily close to one of the sets. If only one of the points is arbitrarily close to a set, say the set $k$, we take the rest of the 10 sets and consider the convex sets $defgh$, $fghij$ and $jabcd$. Since two points hit all the three sets, one of the points should be contained in the region $hxfg$. Now, consider the sets $hijab$ and $bcdef$. The third point must hit both these regions and therefore must be arbitrarily close to the set $b$.

Assuming that none of the points is arbitrarily close to one of the 11 sets, we show that if there exists a set of three points which hits all convex sets containing $5n/11$ points from $P$ then one of those points is contained in one of the bold triangles shown in Figure 2(a).

Consider the four convex sets $jkabc$, $abcde$, $defgh$ and $ghijk$ (see Figure 2(b)) containing $5n/11$ points each. In order to hit all the four regions, one of the three points must be in one the four triangles $jzk$, $gxh$, $dve$ or $asc$. If there is a point in one of the triangles $jzk$, $gxh$ or $dve$, we are done. So, assume that there is a point in the triangle $asc$. There cannot be two points in this region since then the remaining one point cannot hit the disjoint regions $ahijk$ and $cdefg$ simultaneously.

If the point in $asc$ is in one of the triangles $atb$ or $buc$ (see Figure 2(c)), we are done again. So, we assume that it is in the region $stbu$. But then, the regions $abijk$, $fghij$ and $bcdef$ must be hit by the other two points, and one of those must be in the triangle $jyi$



(a) (b) (c)

**Figure 3: (a)** $efijr$ **contains a point of the weak** $\epsilon$**-net (b) Either** $abt$ **and** $efw$ **contain one point each or** $buc$ **and** $ijy$ **contain one point each. (c)** $abt$**,** $efw$ **and** $hxg$ **contain one point each. Hence** $cdijk$ **cannot be hit.**

(see Figure 2(c)).

Hence, one of the bold triangles shown in Figure 2(a) must contain one of three weak $\epsilon$-net points.

Assume that the triangle $hxg$ contains one of the points (the other cases are analogous). Since the regions $abcdk$, $efijk$ and $defij$ must be hit by two points, the region $efijr$ must contain one of the points (see Figure 3(a)). Now, since the regions $abcjk$ and $abcde$ must be hit by one point (see Figure 2(a)), the region $abcs$ contains a point.

Also, since the regions $abijk$ and $bcdef$ must be hit (see Figure 3(b)), either the regions $abt$ and $efw$ contain one point each or the regions $buc$ and $ijy$ contain one point each. Since the cases are symmetric, let us assume that the regions $abt$ and $efw$ contain one point each.

But then, the region $cdijk$ does not contain any point (see Figure 3(c)) although it contains $5n/11$ points of $P$. Hence, it is not possible to hit all the convex regions containing $5n/11$ points of $P$ using 3 points. ☐

Aronov [2] show that $\epsilon_4 \leq 4/7$. We actually are able to hit sets containing $4n/7$ points by just two points (Corollary 3.1). For $\epsilon_4$, Theorem 2.1 yields $16/31$, again improving upon Aronov et al.'s result. Improving upon a result of Alon et al. [1], Aronov et al. [2] showed that if each convex set contains $n/2$ points, then they can be hit by five points. Theorem 2.1 yields an improvement (set $c = 2$, and $d = 1$).

COROLLARY 3.3. $\epsilon_4 \leq 16/31$.

COROLLARY 3.4. $\epsilon_5 \leq 20/41$.

## 4. CONCLUSIONS

We presented a general technique for computing small number of points that hit all convex sets containing points of $P$. This then gives an optimal generalization to two points, and improves bounds of previous work for larger number of points. One intriguing open problem is whether the bound can be closed for the three-point case. Our work leaves a gap ($5/11 < \epsilon_3 \leq 8/15$), and it would be nice to get an optimal bound there.

## 5. ACKNOWLEDGEMENTS

The first author would like to thank Sathish Govindarajan and Shakhar Smorodinsky for enlightening discussions on these topics. The second author would like to thank Raimund Seidel and Hans Raj Tiwary for helpful suggestions.

## 6. REFERENCES

[1] Noga Alon, Imre Bárány, Zoltán Füredi, and Daniel J. Kleitman. Point selections and weak e-nets for convex hulls. *Combinatorics, Probability & Computing*, 1:189–200, 1992.

[2] B. Aronov, F. Aurenhammer, F. Hurtado, S. Langerman, D. Rappaport, S. Smorodinsky, and C. Seara. Small weak epsilon nets. In *Proc. 17th Canadian Conference on Computational Geometry*, 2005.

[3] Bernard Chazelle, Herbert Edelsbrunner, Michelangelo Grigni, Leonidas Guibas, Micha Sharir, and Emo Welzl. Improved bounds on weak eps-nets for convex sets. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 495–504, 1993.

[4] Kenneth L. Clarkson, David Eppstein, Gary L. Miller, Carl Sturtivant, and Shang-Hua Teng. Approximating center points with iterative radon points. *Int. J. Comput. Geometry Appl*, 6(3):357–377, 1996.

[5] D. L. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20:1803–1827, 1994.

[6] Jirí Matousek. *Lectures in Discrete Geometry*. Springer-Verlag, New York, NY, 2002.

[7] Jirí Matousek and Uli Wagner. New constructions of weak epsilon-nets. *Discrete & Computational Geometry*, 32(2):195–206, 2004.

[8] G Miller, S Teng, W. Thurston, and S.A. Vavasis. Automatic mesh partitioning. *Workshop on Sparse Matrix Computations: Graph Theory Issues and Algorithms*, 1993.

[9] Gary L. Miller, Shang-Hua Teng, William P. Thurston, and Stephen A. Vavasis. Separators for sphere-packings and nearest neighbor graphs. *Journal of ACM*, 44:1–29, 1997.

[10] Janos Pach and Pankaj K. Agarwal. *Combinatorial Geometry*. John Wiley & Sons, New York, NY, 1995.

[11] F. Frances Yao. A 3-space partition and its applications. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 258–263, 1983.