

Erdős-Szekeres is NP-hard in 3 dimensions - and what now?

Christian Knauer*

Daniel Werner^{†‡}

Abstract

The Erdős-Szekeres theorem states that, for every k , there is a number n_k such that every set of n_k points in general position in the plane contains a subset of k points in convex position. If we ask the same question for subsets whose convex hull does not contain any other point from the set, this is not true: as shown by Horton, there are sets of arbitrary size that do not contain an empty 7-gon.

These questions have also been studied extensively from a computational point of view, and polynomial time algorithms for finding the largest (empty) convex set have been given for the planar case. In higher dimension, it is not known how to compute such a set efficiently. In this paper, we show that already in 3 dimensions no polynomial time algorithm exists for determining the largest (empty) convex set (unless $P=NP$), by proving that the corresponding decision problem is NP-hard. This answers a question by Dobkin, Edelsbrunner and Overmars from 1990.

As a corollary, we derive a similar result for the closely related problem of testing weak ε -nets in \mathbb{R}^3 . Answering a question by Chazelle et al. from 1995, our reduction shows that the problem is co-NP-hard.

Finally, we make several suggestions for further research on the subject.

1 Preliminaries

The Erdős-Szekeres theorem [8] is one of the major theorems from combinatorial geometry and one of the earliest results in geometric Ramsey theory.

Theorem 1 (Erdős and Szekeres, 1935) *For every k there is a number n_k such that every planar set of n_k points in general position contains k points in convex position.*

Exact values of n_k are known only for very few cases and subject to extensive research, also for the higher dimensional cases.

A closely related question is the following: is the theorem still true if we ask for sets whose convex

hull is empty, i.e., does not contain any other point from the set? That this is not the case was shown by Horton [11]: in the plane there are arbitrary large sets which do not contain empty 7-gons. Nicolás [15] and Gerken [9] independently solved the long standing open problem whether or not there is always an empty 6-gon. See Matoušek [12, Chapter 3] for further details and references.

Both these questions generalize to dimension larger than 2 in the obvious way, and clearly the numbers n_k do not increase when the dimension gets larger (proof: project to \mathbb{R}^2). See the surveys by Bárány and Károlyi [2] or Morris and Soltan [14] for further references and (more or less) recent progress on the subject.

The corresponding computational problems have also received a lot of attention in the past (e.g., [1], [6], [7], [13]). Polynomial time algorithms are known for both problems in the plane. The fastest algorithm is given in [7], and the question is stated whether a polynomial time algorithm for determining the largest empty convex set also exists in \mathbb{R}^3 .

1.1 Our results

In this paper, we will consider the following decision problems:

Definition 1 (ERDŐS-SZEKERES) *Let P be a set of points in \mathbb{R}^d and $k \in \mathbb{N}$. Is there a set $Q \subseteq P$ of k points in convex position?*

and

Definition 2 (LARGEST-EMPTY-CONVEX-SET) *Let P be a set of points in \mathbb{R}^d and $k \in \mathbb{N}$. Is there a set $Q \subseteq P$ of k points in convex position whose convex hull does not contain any other point from P ?*

Using the reduction technique from Giannopoulos et al. [10], it is an easy exercise to show that both problems are NP-hard if the dimension is not fixed. For people familiar with parameterized complexity: the problem is even W[1]-hard with respect to the dimension d . This means that it is very unlikely to admit an algorithm with running time $O(f(d)n^c)$ for any computable function f and constant c .

Still, this does not exclude the possibility that in every fixed dimension, the problem can be solved with a running time of, say, $O(n^{d+1})$. In this paper, we

*Institut für Informatik, Universität Bayreuth, Germany
christian.knauer@uni-bayreuth.de

†Institut für Informatik, Freie Universität Berlin, Germany,
daniel.werner@fu-berlin.de

‡This research was funded by Deutsche Forschungsgemeinschaft within the Research Training Group (Graduiertenkolleg) “Methods for Discrete Structures”.

show that this cannot be the case (under standard complexity theoretic assumptions):

Theorem 2 *The problems LARGEST-EMPTY-CONVEX-SET and ERDŐS-SZEKERES are NP-hard in \mathbb{R}^3 .*

The first part of the theorem, hardness of LECS, is shown in Sec. 2. In Sec. 3, the proof is adapted to ES. In Sec. 4, we derive a similar result for testing weak ε -nets and red-blue discrepancy. Finally, in Sec. 5, we make several suggestions for further research on the subject.

2 The reduction

We will show that the problems is NP-hard by a reduction from the following problem:

Definition 3 (ISNUD) *Given a set of pairwise non-overlapping unit disks in \mathbb{R}^2 , decide whether there are k disks such that no two of them touch.*

Here, non-overlapping means that the interiors of the disks are pairwise disjoint. As shown by Cerioli et al. [3], the problem ISNUD is NP-hard. The reduction has the additional property that there is no K_3 in the underlying geometric graph, which we will use in the proof of Lemma 3.

We will now reduce this problem to LECS and show how to adapt it to ES in the next section.

For a given instance \mathcal{D} of unit disks in the plane, we will create a set of points in \mathbb{R}^3 . These points will almost lie on the elliptic paraboloid, in a sense to be made precise later.

For a point $x = (x_1, x_2) \in \mathbb{R}^2$, let

$$\text{lift}: (x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2)$$

denote the standard lifting transform to the paraboloid.

Let D_c denote the n centers of the disks in \mathcal{D} . Let L denote the set of all points $\hat{x} = \text{lift}(x)$, for $x \in D_c$.

We now want to forbid certain pairs of points to lie in empty convex positions, namely those for which the corresponding disks intersect. Thus, for a pair of intersecting disks d, d' and their centers $c_d, c_{d'}$, we add a blocking point

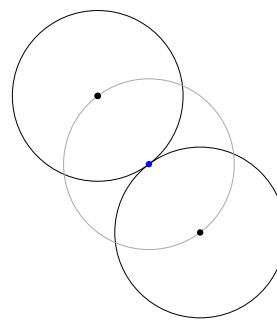
$$b_{dd'} = \frac{1}{2} (\text{lift}(c_{d'}) + \text{lift}(c_d)).$$

The set B then consist of all the points

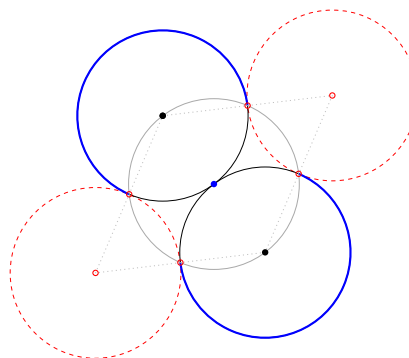
$$\{b_{dd'} \mid d \cap d' \neq \emptyset\},$$

and we set $P = L \uplus B$.

Thus, we have created $O(|\mathcal{D}|)$ points and the reduction is linear in the input size. The main property of the reduction is captured in the following lemma.



(a) Two intersecting disks and the projected blocking point



(b) Because the dashed circles do not appear in the construction, all projected blocking points lie on the bold arcs.

Figure 1: Finding an empty circle

Proposition 1 *Let Q be a set of points and h be a hyperplane through p such that $hx \geq 0$ for all $x \in Q$. Then p is in $\text{ch}(Q)$ if and only if it is in $\text{ch}(Q \cap h)$.*

Lemma 3 *A blocking point $b_{dd'}$ is contained in the convex hull of a set $Q \subseteq L$ if and only if \hat{c}_d and $\hat{c}_{d'}$ are contained in Q .*

Proof. \Leftarrow : by definition

\Rightarrow : We show that there is a hyperplane that contains $b_{dd'}$, \hat{c}_d , and $\hat{c}_{d'}$ and has all other points strictly on the positive side. Here we will make use of the fact that our instance consists of non-overlapping unit disks — otherwise, the claim would not hold.

Let C be the circle with center $\text{proj}(b_{dd'})$ through c_d and $c_{d'}$. Because all disks are non-overlapping unit disks, this circle does not contain any other points from D_c . Further, because the underlying graph does not contain a K_3 , the circle does not contain any (projection of) a blocking point. See Fig. 1. We then take as h the unique hyperplane whose intersection with the paraboloid projects to the circle C . This hyperplane contains all three points, and because C does not contain any points, all other points from L and thus B lie strictly above h . The claim then follows from Proposition 1. \square

The following states that whether or not a set is in empty convex position will depend only on which points we choose from L . The set B can always be added without destroying this property.

Proposition 2 *The sets L and B each are in empty convex position, and $\text{ch}(L) = \text{ch}(L \cup B)$.*

Proof. By construction, all points of L lie on the paraboloid. The points from B can be separated from each other by the hyperplane defined in the previous proof. As all of them are convex combinations of points in L , we have $\text{ch}(B) \subseteq \text{ch}(L)$. \square

Corollary 4 *A set $L' \uplus Q' \subseteq P$ is in convex position if and only if no point of $Q' \subseteq Q$ is contained in the convex hull of $L' \subseteq L$.*

The main lemma then reads as follows:

Lemma 5 *There is an independent set of size m among the unit disks if and only if there are $m + |B|$ points in empty convex position.*

Proof. \Rightarrow : Let I , $|I| = m$, be an independent set among the set of disks. Let $\hat{I} \subset L$ denote the corresponding lifted centers. We claim that $S = \hat{I} \cup B$ is in empty convex position. Indeed, by Corollary 4, no point of $L - \hat{I}$ is in the convex hull of S . Further, by Lemma 3, if some point $b \in B$ was in $\text{ch}(S)$, this would mean that there are two points in \hat{I} that contained b . Thus, the corresponding disks would touch, and I would not be an independent set. This means that there are $m + |B|$ points in empty convex position.

\Leftarrow : Now assume that there is no independent set of size m . This means that for any choice of m disks, two of them touch. Now take any set S of $m + |B|$ points. As there are only $|L| + |B|$ points in total, this must contain at least m points from L . Thus, some two of them belong to disks that intersect. By Lemma 3, their convex hull contains a point of B . Thus, S is not in empty convex position. \square

3 Adaption to Erdős-Szekeres

We now show how this exact reduction can be applied to ERDŐS-SZEKERES. One direction of Lemma 5 is clear, since we have shown how an independent set of size m results in an empty convex set of size $m + |B|$. For the other direction, we need to show that if there is *any* (not necessarily empty) convex set of $m + |B|$ points, then there is also an independent set of size m among the disks.

Lemma 6 *There is an independent set of size m among the unit disks if and only if there are $m + |B|$ points in convex position.*

Proof. \Rightarrow : An empty convex set is convex.

\Leftarrow : Let S be a set of $m + |B|$ points in convex position with $|S \cap B| < |B|$. We show how to construct a set S' in convex position of the same size such that $|S' \cap B| = |S \cap B| + 1$.

Let $I = S \cap L$, and let D_I be the corresponding set of disks. Observe that, if $|S \cap B| < |B|$, then $|I| > m$. If all disks from D_I are independent, we are done. Otherwise, let d and d' be two disks from D_I that intersect. The point $b_{dd'}$ cannot be part of S , for otherwise S would not be in convex position. If we thus set $S' = I - \{d\} \cup B \cup \{b_{dd'}\}$, by Corollary 4 the set is still in convex position and we have $|S'| = |S|$ and $|S' \cap B| > |S \cap B|$. Thus, after finitely many steps we end up with a set of $m + |B|$ points which contains all points from B . In particular, the set contains no point from B in the convex hull. By Lemma 5, this means that the disks corresponding to the m points from L do not intersect. Thus, we have an independent set of size m . \square

This finishes the proof of Thm. 2.

4 Testing weak ε -nets and red-blue discrepancy

Here we shortly mention that the hardness proofs also show hardness for two closely related problems. Recall that a range space is a pair (X, \mathcal{R}) , where $\mathcal{R} \subset 2^X$. If X is a set of points in \mathbb{R}^d and \mathcal{R} is the set of all convex sets determined by them, a weak ε -net for (X, \mathcal{R}) is a set of points S such that $|S \cap R| \neq \emptyset$ whenever $|R \cap X| \geq \varepsilon|X|$, for all $R \in \mathcal{R}$. See Matoušek [12, Chapter 10] for further details.

This leads to the following decision problem as follows:

Definition 4 (ε -NET-VERIFICATION) *Given a set of points $P \subset \mathbb{R}^d$, another set $S \subset \mathbb{R}^d$ and an $\varepsilon > 0$. Is S an ε -net for P with respect to all convex sets?*

Chazelle et al. [4] give a polynomial time algorithm for the problem in the plane and ask whether it is solvable in polynomial time in \mathbb{R}^3 .

A closely related concept is that of red-blue discrepancy: Given a set R of red and a set B of blue points, the *discrepancy* of a set C is defined as

$$D(C) = ||R \cap C| - |B \cap C||.$$

The discrepancy of the set $P = R \cup B$ is then defined as $D(P) = \max_C D(C)$. The corresponding decision problem RED-BLUE-DISCREPANCY asks whether the discrepancy of a given set is at least some value $k \in \mathbb{N}$.

Now observe that the set of blocking points B determines an (m/n) -net¹ for the set of lifted points L if and only if there is *no* empty convex set of size

¹Recall that n was the number of unit disks.

$m + |B|$ among the points of $L \uplus B$. A similar argument holds for RED-BLUE-DISCREPANCY. Our reduction thus also shows the following:

Theorem 7 *The problem ε -NET-VERIFICATION is co-NP-hard in \mathbb{R}^3 and RED-BLUE-DISCREPANCY is NP-hard in \mathbb{R}^3 .*

5 Conclusion and open problems

The major open question is to find an approximation algorithm for the problems ERDŐS-SZEKERES and LARGEST-EMPTY-CONVEX-SET. The obvious approach (projecting to \mathbb{R}^2 and solving the problem there) does not work very well: as shown by Chazelle et al. [5], there are polytopes whose projection in any direction has $\Theta(\log n)$ vertices on the convex hull. This leads to a polynomial time $(\log n)/n$ -approximation, which only very few people will be happy about. Thus, the question for a more intelligent (probably constant-factor) approximation algorithm remains and seems to be very challenging.

In addition to this, the most interesting question is the following: Is LARGEST-EMPTY-CONVEX-SET in \mathbb{R}^3 fixed parameter tractable with respect to the size of the solution? That is, can we decide whether there are k points in empty convex position in time $O(f(k) \cdot n^c)$ for some computable function f and constant c ? More generally, given a point set P in \mathbb{R}^d , can we decide whether there is an empty convex set of size k in time $O(f(k)n^{O(d)})$?

Observe that due to the Erdős-Szekeres theorem itself, the problem ERDŐS-SZEKERES is trivially fixed-parameter tractable: Given a point set P and a $k \in \mathbb{N}$, if $n := |P| \leq 2^k$, we use a brute force algorithm, i.e., simply try all subsets of size k . This takes time $\binom{n}{k} \approx n^k \leq (2^k)^k$. If $n > 2^k$, we simply answer yes. In any case, the running time is bounded by 2^{k^2} , and thus we have an algorithm with running time $O(f(k)n)$. Here, the question for a polynomial size problem kernel is of interest.

Acknowledgements We thank Günter Rote and Nabil Mustafa for pointing us out to [5] and [4], respectively.

References

- [1] D. Avis and D. Rappaport. Computing the largest empty convex subset of a set of points. In *Proceedings of the first annual Symposium on Computational geometry*, SCG '85, New York, NY, USA, 1985. ACM.
- [2] I. Bárány and G. Károlyi. Problems and Results around the Erdős-Szekeres Convex Polygon Theorem. In *Discrete and Computational Geometry*, volume 2098 of *Lecture Notes in Computer Science*. Springer, 2001.
- [3] M. Cerioli, L. Faria, T. Ferreira, and F. Protti. On minimum clique partition and maximum independent set on unit disk graphs and penny graphs: complexity and approximation. *Electronic Notes in Discrete Mathematics*, 18, 2004.
- [4] B. Chazelle, H. Edelsbrunner, D. Eppstein, M. Grigni, L. Guibas, M. Sharir, and E. Welzl. Algorithms for weak epsilon-nets, 1995.
- [5] B. Chazelle, H. Edelsbrunner, and L. J. Guibas. The complexity of cutting complexes. *Discrete & Computational Geometry*, 4, 1989.
- [6] V. Chvátal and G. Klinešek. Finding largest convex subsets. *Congressus Numeratum* 29, 1980.
- [7] D. Dobkin, H. Edelsbrunner, and M. Overmars. Searching for empty convex polygons. *Algorithmica* 5, 1990.
- [8] P. Erdős and G. Szekeres. A combinatorial problem in geometry. *Compositio Math.* 2, 1935.
- [9] T. Gerken. Empty convex hexagons in planar point sets. *Discrete Comput. Geom.*, 39, 2008.
- [10] P. Giannopoulos, C. Knauer, M. Wahlström, and D. Werner. Hardness of discrepancy computation and epsilon-net verification in high dimension. *Journal of Complexity*, 2012. In press; available online.
- [11] J. D. Horton. Sets with no empty convex 7-gons. *C. Math. Bull.* 26, 1983.
- [12] J. Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.
- [13] J. S. B. Mitchell, G. Rote, G. Sundaram, and G. Woeginger. Counting convex polygons in planar point sets. *Information Processing Letters* 56, 1995.
- [14] W. Morris and V. Soltan. The Erdős-Szekeres problem on points in convex position – a survey. *Bull. Amer. Math. Soc.* 37, 2000.
- [15] C. M. Nicolás. The empty hexagon theorem. *Discrete Comput. Geom.*, 38, September 2007.