Counting Large Distances in Convex Polygons: A Computational Approach

Filip Morić and David Pritchard

Abstract In a convex n-gon, let $d_1 > d_2 > \cdots$ denote the set of all distances between pairs of vertices, and let m_i be the number of pairs of vertices at distance d_i from one another. Erdős, Lovász, and Vesztergombi conjectured that $\sum_{i \le k} m_i \le kn$. Using a new computational approach, we prove their conjecture when $k \le 4$ and n is large; we also make some progress for arbitrary k by proving that $\sum_{i \le k} m_i \le (2k-1)n$. Our main approach revolves around a few known facts about distances, together with a computer program that searches all distance configurations of two disjoint convex hull intervals up to some finite size. We thereby obtain other new bounds, such as $m_3 \le 3n/2$ for large n.

1 Introduction

Given a set S of n points in the plane, let $d_1 > d_2 > \cdots$ be the set of all distances between pairs of points in S. It was shown by Hopf and Pannwitz in 1934 [5] that the distance d_1 (the diameter of S) can occur at most n times, which is tight (e.g., for a regular polygon of odd order). In 1987, Vesztergombi [6] showed that the second-largest distance, d_2, can occur at most $\frac{3}{2}n$ times; she subsequently [7] considered the version of the problem when the points are in convex position and showed that in this case the number of second-largest distances is at most $\frac{4}{3}n$. She also showed that both results are tight up to additive constants.

F. Morić
Ecole Polytechnique Fédérale de Lausanne, Chair of Combinatorial Geometry,
EPFL SB IMB DCG, MA C1 585 (Bâtiment MA), CH-1015 Lausanne, Switzerland

D. Pritchard (✉)
Centre for Education in Math and Computing, University of Waterloo,
200 University Ave W, N2L 3G1 Waterloo, ON, Canada
e-mail: daveagp@gmail.com

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_22,

Let m_i denote the number of times that d_i occurs. It is known that $m_k \leq 2kn$ [6], and moreover that $m_k \leq kn$ for point sets in convex position [7], while the following open conjecture would imply $m_k \leq 2n$.

Conjecture 1.1 (Erdős, Moser [2, 7]). The number of unit distances generated by n points in convex position cannot exceed $2n$.

A lower bound of $2n-7$ for this conjecture is known due to Edelsbrunner and Hajnal [3].

For the rest of the chapter, we consider only point sets in convex position. One natural question is to find how large $m_{\leq k} := \sum_{i \leq k} m_i$, i.e., the number of *top-k* distances, can be in terms of n. The conjectured value is

Conjecture 1.2 (Erdős, Lovász, Vesztergombi [4]). The number of top-k distances generated by n points in convex position is at most kn; i.e., $m_{\leq k} \leq kn$.

Odd regular polygons prove $m_{\leq k} = kn$ is possible. In [4], the bound $m_{\leq k} \leq 3kn$ is proven, and $m_{\leq 2} \leq 2n$ was shown in [7], verifying Conjecture 1.2 for $k = 2$.

In this chapter, we give improved upper bounds on m_k and $m_{\leq k}$ for convex point sets, and more generally bounds for sums of the form $\sum_{t \in T} m_t$. Our first result is the following.

Theorem 3. *For any $k \geq 1$, the number of top-k distances generated by n points in convex position is at most $(2k-1)n$; i.e., $m_{\leq k} \leq (2k-1)n$.*

Thus, we close about half of the gap toward Conjecture 1.2.

Next, by combining several known conditions on distances for convex point sets, and by using a computer program to carry out an exhaustive search on a finite abstract version of the problem, we prove the following.

Theorem 4. *The distances generated by n points in convex position satisfy the following bounds, for large enough n:*

- $m_{\leq 3} \leq 3n, m_{\leq 4} \leq 4n$;
- $m_3 \leq \frac{3}{2}n, m_4 \leq \frac{13}{8}n$;
- $m_1 + m_3 \leq 2n, m_2 + m_3 \leq \frac{9}{4}n$.

In particular, we verify Conjecture 1.2 for $k \leq 4$ and n large. For m_3 and $m_2 + m_3$, the bound is as good as can be obtained by our abstract version of the problem, as witnessed by periodic patterns achieving $m_3 = \frac{3}{2}n$ and $m_2 + m_3 = \frac{9}{4}n$, but we do not know if any convex polygon can realize these distances; we elaborate in Sect. 6.

The proof of Theorem 4 uses a computer program to make certain types of automatic deductions, as well as the following lemma to eliminate long distances "near" the boundary.

Lemma 1.5. *For any $k \geq 1$ and $\ell \geq 0$, there is a constant $C(k, \ell)$ such that the following holds: In a convex polygon, if there are ℓ or fewer vertices between some vertices a and b such that $|ab| \geq d_k$, then the number of top-k distances satisfies $m_{\leq k} \leq n + C(k, \ell)$.*

The detailed bound we obtain is of the form $C(k,\ell) = O(k^2(k+\ell)^2)$. In an earlier version of this chapter,[1] we proved results like "$m_{\leq 3} \leq 3n + O(1)$," which are weaker for large n but better for small n, using the following alternative lemma.

Lemma 1.6. *For any $k \geq 1$ and $\ell \geq 0$, there is a constant $C'(k,\ell)$ such that the following holds. In a convex polygon, at most $C'(k,\ell)$ diagonals ab have both (i) ℓ or fewer vertices between a and b and (ii) $|ab| \geq d_k$.*

In the latter, $C'(k,\ell) = O(k\ell^2)$. We do not think either lemma is tight.

In Sect. 2, we describe *levels*, a key element in our approach. In Sect. 3, we collect geometric facts used by the algorithm. We prove Lemma 1.5 in Sect. 3.1. The proof of our main result, Theorem 4, consists of the algorithmic approach described in Sect. 4 together with our computational results stated in Sect. 5. We conclude with suggestions for future work.

2 Levels

We use the term *diagonal* to mean any line segment connecting two points of S, including sides of the convex hull of S. We will partition the diagonals into n *levels* in the following way. Let $S = \{a_1, a_2, \ldots, a_n\}$ be the vertex set of our convex polygon, ordered clockwise. Then *level i* is the set of diagonals

$$L_i := \{a_j a_k \mid j + k \equiv i \bmod n\},$$

where the index i can be taken modulo n. Equivalently, consider an auxiliary regular n-gon $b_1 b_2 \ldots b_n$; then two diagonals $a_i a_j$ and $a_k a_l$ lie in the same level when the corresponding segments $b_i b_j$ and $b_k b_l$ are parallel. We illustrate this in Fig. 1a.

Levels are used in the following way to prove Theorem 3 (i.e., $m_{\leq k} \leq (2k-1)n$).

Proof of Theorem 1.3. In the next section, we prove Lemma 3.5: In any level, there are at most $2k-1$ diagonals of length $\geq d_k$. Since there are at most n levels, we are done. □

3 Geometric Facts

To begin this section, we collect four geometric facts from the literature [1, 4, 7], which will be used in our computer program. For completeness, we include the proofs. The first two facts were used in [4, 7].

Fact 3.1. If $abcd$ is a convex quadrangle, then $|ab| + |cd| < |ac| + |bd|$.

---

[1] http://arxiv.org/abs/1103.0412v1.

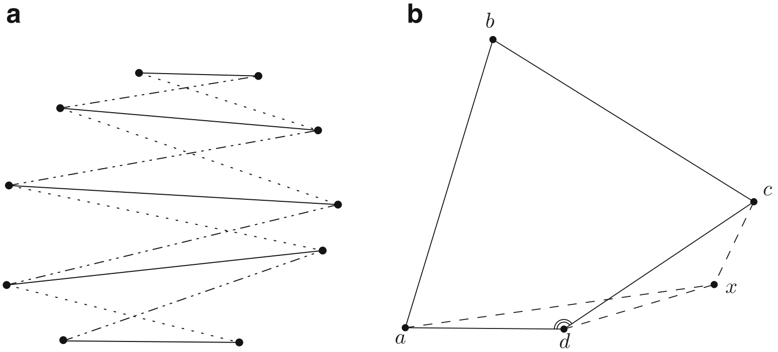


Fig. 1 (**a**) Three consecutive levels of diagonals in a convex decagon. (**b**) Proof of Fact 3.2

Proof. Let p be the intersection point of the diagonals ac, bd. Then, by the triangle inequality,

$$|ab| + |cd| < |ap| + |bp| + |cp| + |dp| = |ac| + |bd|. \qquad \square$$

Fact 3.2. If a, b, c, d are vertices of a convex polygon in clockwise order, then at least one of these four cases must occur:

- $|ax| > |ad|$ for all vertices x of the polygon between c and d, including c;
- $|bx| > |bc|$ for all vertices x of the polygon between c and d, including d;
- $|cx| > |bc|$ for all vertices x of the polygon between a and b, including a;
- $|dx| > |ad|$ for all vertices x of the polygon between a and b, including b.

Proof. Since the sum of the angles of quadrilateral $abcd$ is 2π, at least one angle is nonacute. Without loss of generality, let $\angle cda \geq \frac{\pi}{2}$. Then for any vertex x of the polygon between c and d, we have that $\angle xda \geq \angle cda \geq \frac{\pi}{2}$, and, thus, $|ax| > |ad|$ (see Fig. 1b). □

The special case $i = j$ of the following fact appears in [4].

Fact 3.3. If a, b, c, d are vertices of a convex polygon listed in clockwise order, such that $|bc| \geq d_i$ and $|ad| \geq d_j$, where d_i and d_j are the ith- and jth-largest distances among vertices of the polygon, then either between a and b or between c and d there are no more than $i + j - 3$ other vertices of the polygon.

Proof. Let's denote without loss of generality $a = a_1, b = a_x, c = a_y, d = a_z$. We will show $\min\{x - 1, z - y\} \leq i + j - 2$, which proves the lemma. We use induction on $i + j$ see Fig. 2. The base case $i = j = 1$ amounts to saying that any two noncrossing d_1's must share a vertex, which follows by Fact 3.1.

For the inductive step, we apply Fact 3.2. Suppose that the first of the four cases happens, so $d' := a_{z-1}$ satisfies $|ad'| > |ad|$; the other cases are similar. Consequently, $|ad'| \geq d_{j-1}$. By induction, $\min\{x - 1, (z - 1) - y\} \leq i + (j - 1) - 3$, from which the desired result follows (Fig. 2). □

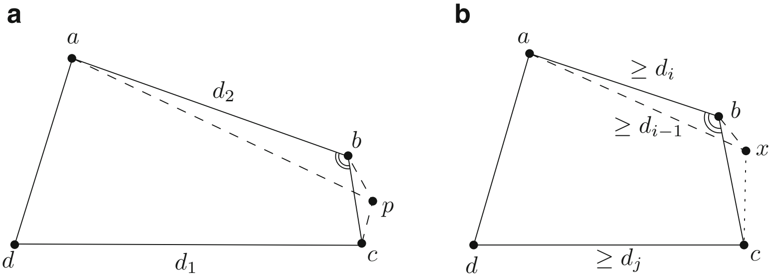


Fig. 2 (**a**) Proof of Fact 3.3, base case $i = 2$, $j = 1$; (**b**) proof of Fact 3.3, inductive step

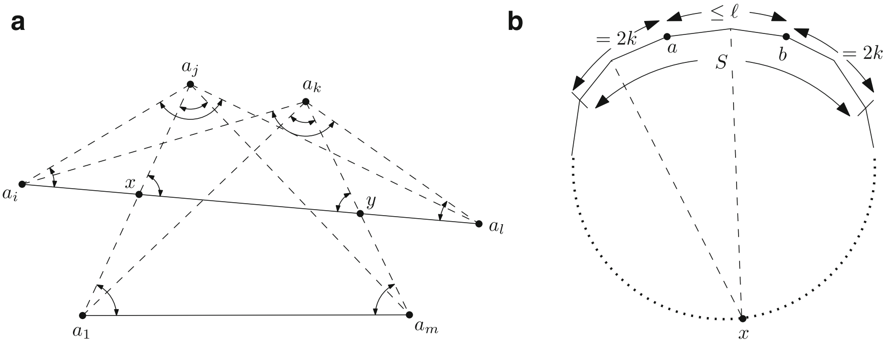


Fig. 3 (**a**) Proof of Fact 3.4; (**b**) proof of Lemma 1.5

The following is a strengthening of a result of Altman, obtained by removing all nonessential conditions from the hypothesis of [1, Lemma 1] but using the same proof. (He considered only the case where $|a_1a_m| = d_1$.)

Fact 3.4. Let $a_1 \ldots a_n$ be a convex polygon. If $1 \le i < j \le k < \ell < m$ and $|a_1a_m| \ge \max\{|a_1a_k|, |a_ja_m|\}$, then $|a_ia_\ell| > \min\{|a_ia_k|, |a_ja_\ell|\}$.

Proof. Suppose for the sake of contradiction that $|a_ia_\ell| \le \min\{|a_ia_k|, |a_ja_\ell|\}$. Denote by x and y the points where a_1a_j and a_ma_k intersect a_ia_ℓ (see Fig. 3a). Repeatedly using the fact that when s, s' are two sides of a triangle, $|s| > |s'|$ iff the angle opposite s is larger than the angle opposite s', we have

$$\begin{aligned}\angle a_jxa_\ell + \angle a_kya_i &> \angle a_ja_ia_\ell + \angle a_ka_\ell a_i \ge \angle a_ia_ja_\ell + \angle a_\ell a_ka_i \\ &> \angle a_1a_ja_m + \angle a_1a_ka_m \ge \angle a_ja_1a_m + \angle a_ka_ma_1\,.\end{aligned}$$

However, $\angle a_jxa_\ell + \angle a_kya_i = \angle a_ja_1a_m + \angle a_ka_ma_1$, which gives a contradiction. □

3.1 Counting Lemmas

First, we complete the proof of Theorem 3, using Fact 3.3.

Lemma 3.5. *In any level, there are at most* $2k-1$ *diagonals of length* $\geq d_k$.

Proof. Without loss of generality (by relabeling), we consider the level L_0. The diagonals of this level are $a_j a_{-j}$, with indices modulo n, for $0 < j < n/2$. Let $m > 0$ (resp., M) be the minimal (resp., maximal) j such that $|a_j a_{-j}| \geq d_k$. Then, by Fact 3.3, we see that $M - m - 1 \leq k + k - 3$. So the number of top-k diagonals in L_0 is bounded by $|\{m, m+1, \ldots, M\}| = M - m + 1 \leq 2k - 1$, which gives the corollary. □

Next, we give the proof of Lemma 1.5, which is needed in order to argue that our computational approach is correct.

Proof. We want to show that if $|ab| \geq d_k$, and a and b are separated by at most ℓ vertices, then the number of top-k distances satisfies $m_{\leq k} \leq n + O(k^2(k+\ell)^2)$. Let S be the interval obtained from this $[a,b]$ by extending onto $2k$ further points in both directions. By Fact 3.3, all edges of length $\geq d_k$ have at least one endpoint in S. Note $|S| = O(k+\ell)$.

We will show an upper bound of $n + O(k^2(k+\ell)^2)$ on the number of edges sx of length $\geq d_k$, with $s \in S, x \in V \backslash S$. This will complete the proof since the only other top-k distance edges must lie with both endpoints in S, and there are at most $O(k+\ell)^2$ such edges.

The key observation is that in the bipartite graph between S and $V \backslash S$ consisting of these edges, all but a constant number of vertices in $V \backslash S$ have degree 1. Specifically, if $sx, s'x$ are both edges in this graph, then the location of x is uniquely determined by $s, s', |sx|$, and $|s'x|$; it follows that $\sum_x \binom{\deg(x)}{2}$ is at most $O((k+\ell)^2 k^2)$, and consequently $\sum_{x:\deg(x)>1} \deg(x) = O((k+\ell)^2 k^2)$. We are then done by counting the endpoints of degree-1 vertices, of which there are at most n. □

4 The Algorithm

The algorithm we use to prove Theorem 4 examines distances among finite configurations of points in the plane. Informally, we examine all possible configurations of a bounded size, where a configuration includes all occurrences of top-k distances in a few consecutive levels, and we try to establish that not too many top-k distances can occur per level, averaged over a small interval of levels. Thus, ultimately, the argument in our proof decomposes any global point set into local configurations of bounded size.

4.1 The Goal

Our computational goal will be to bound the number of long distances that can occur in a consecutive sequence of several levels. We begin by reproving (for large n) Vesztergombi's result on counting the second-largest distances; it illustrates the type of computational result we need.

Proposition 4.1. *We have* $m_2 \leq \frac{4}{3}n$ *for large enough* n.

Proof. We prove the theorem for $n \geq 3 \cdot C(16,2)$ with C as in Lemma 1.5. Let a *special diagonal* be a diagonal of length d_2 or longer, whose endpoints are separated by at most 16 vertices. If there is any special diagonal, we are done by Lemma 1.5. So we may assume there are no special diagonals.

Using our computer program, we establish the following lemma.

Lemma 4.2. *In every point set* S *without special diagonals, for every level* i*, at least one of the following is true:*

- *at most* $1 = \lfloor 1 \cdot \frac{4}{3} \rfloor$ *diagonal in level* i *has length* d_2*;*
- *at most* $2 = \lfloor 2 \cdot \frac{4}{3} \rfloor$ *diagonals in levels* i *and* $i+1$ *have length* d_2*;*
- *at most* $4 = \lfloor 3 \cdot \frac{4}{3} \rfloor$ *diagonals in levels* $i,\ldots,i+2$ *have length* d_2*;*
- *at most* $5 = \lfloor 4 \cdot \frac{4}{3} \rfloor$ *diagonals in levels* $i,\ldots,i+3$ *have length* d_2*.*

Now let's see how this gives the desired result. Taking $i = 1$, the four cases above establish that for some $1 \leq \gamma_1 \leq 4$, the number of d_2's in levels $1,\ldots,\gamma_1$ is at most $\frac{4}{3}\gamma_1$. Applying the same logic to $i = \gamma_1 + 1$, we get that there is some $1 \leq \gamma_2 \leq 4$ such that the number of d_2's in levels $\gamma_1 + 1,\ldots,\gamma_1 + \gamma_2$ is at most $\frac{4}{3}\gamma_2$.

We continue to further define γ_i's in the same way until $\sum_{i=1}^{x} \gamma_i \equiv \sum_{i=1}^{y} \gamma_i \pmod{n}$ for some $x < y$. Summing a contiguous subset of these bounds, the number of d_2's in levels from $1 + \sum_{i=1}^{x} \gamma_i$ to $\sum_{i=1}^{y} \gamma_i$ is at most $\frac{4}{3}$ per level on average. But this sum counts each of the n levels an equal number of times, so the number of d_2's overall is at most $\frac{4}{3}n$. □

The computer program's goal is thus to prove a general version of Lemma 4.2: Given a *target ratio* α and *target distances* (a subset of $\{d_1, d_2, \ldots, d_k\}$), find a constant m so that every level i admits $1 \leq m' \leq m$ such that $\leq m' \cdot \alpha$ target lengths occur in levels $i,\ldots,i+m'$. The program searches for a point set with $> \alpha$ target diagonals in level 1, $> 2\alpha$ in level 2, etc. If the search terminates, the above proof shows the number of target distances is $\leq \alpha n$. The hypothesis that no special diagonals exist is used only indirectly by the program, explained below.

Our algorithm works with *configurations* consisting of two disjoint intervals of points, and an assignment of a distance from $\{d_1, d_2, \ldots, d_k, \text{“} < d_k\}$ to each diagonal spanning the two intervals. We thereby obtain analogues of Lemma 4.2 by checking all possible configurations up to some finite size. For this to work, Fact 3.2 is crucial since it implies that all of the top-k distances in ℓ consecutive levels have all of their endpoints in two intervals of bounded size. We use an incremental branch-and-bound search: It exhaustively searches all possibilities, but in an efficient way where large sections of the search space can be eliminated at once. Each individual step of the algorithm corresponds to an application of one of the Facts 3.1–3.4. The lack of special diagonals allows us to focus on *disjoint* interval pairs. The Java implementation is available at

http://sourceforge.net/projects/convexdistances/.

4.2 Configurations

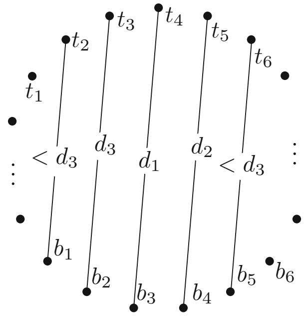
In more detail, our algorithm maintains a set of *configurations*. Each configuration has two disjoint intervals of points from S; then for each diagonal generated by one point from each interval, the configuration stores a set of possible values for the distance between those two points. Arbitrarily name one interval the *top* and denote its points as $\{t_i\}_i$, with t_{i+1} following t_i in clockwise order, and name the other interval the *bottom* with points $\{b_i\}_i$, and b_{i-1} following b_i in clockwise order. Then we denote the set of possible distances between t_i and b_j as $D[i,j]$; in each configuration $D[i,j]$ is a subset of $\{1,2,\dots,k,\infty\}$, where $x \in D[i,j]$ means that d_x is a possible value for the distance $|t_ib_j|$, while $\infty \in D[i,j]$ means that it is possible for $|t_ib_j|$ to be shorter than d_k. (So typical steps in our program use special cases to reason with "d_∞" distances correctly.) Reiterating, a configuration consists of a top interval of indices, a bottom interval of indices, and for each top-bottom pair a subset of $\{1,2,\dots,k,\infty\}$.

We assume that t_ib_j is in level number $j-i$ (modulo n), which is without loss of generality. To gain some intuition and exhibit the notation, it is helpful to look at a couple of examples. Our examples will be drawn from actual point sets and therefore each $D[i,j]$ will be just a singleton, in contrast to the larger sets $D[i,j]$ typically occurring in the algorithm. The first example, shown in Fig. 4, is a regular polygon of odd order. The second example, shown in Fig. 5, exhibits the extremal construction of Vesztergombi for second distances [7].

4.3 Methodology

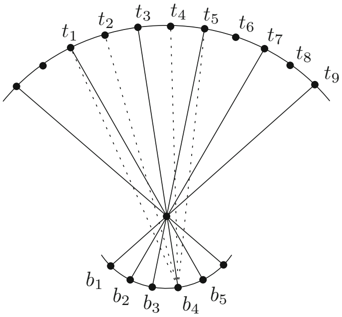
Here is an example of a typical step in the algorithm, shown in Fig. 6. Suppose some configuration includes points t_1,t_2,b_2,b_1, suppose that $D[1,1] = D[2,2] = \{2\}$, $D[1,2] = \{2,3,\infty\}$ and that $D[2,1] = \{1,2,3,\infty\}$. Then, using Fact 3.1, we know that $|t_1b_2| + |t_2b_1| > |t_1b_1| + |t_2b_2|$. As the right-hand side equals $2d_2$ and the maximum possible length of t_1b_2 is d_2, we can deduce that $|t_2b_1| > d_2$ and so we may update the configuration via $D[2,1] := \{x \in D[2,1] \mid x < 2\} = \{1\}$.

The program uses Facts 3.1–3.4 in ways analogous to the above example. Whenever one of the facts is applicable, we use it to reduce the size of one set D in the configuration. We use Fact 3.4 only when a_1,a_i,a_j lie in the top interval and a_k,a_l,a_n lie in the bottom, or vice versa.



	$i=1$	2	3	4	5	6
$j=1$	∞	$\textcircled{\infty}$	∞	3	2	1
$j=2$	∞	∞	$\textcircled{3}$	2	1	1
$j=3$	∞	3	2	$\textcircled{1}$	1	2
$j=4$	3	2	1	1	$\textcircled{2}$	3
$j=5$	2	1	1	2	3	$\textcircled{\infty}$
$j=6$	1	1	2	3	∞	∞

Fig. 4 *Left*: an odd regular polygon, with a *top* and *bottom* interval. *Right*: the corresponding values of D, where entry x in column i, row j indicates $D[i,j]=\{x\}$. One level is illustrated on the *left* and *circled* on the *right*



	$i=1$	2	3	4	5	6	7	8	9
$j=1$	∞	∞	∞	∞	∞	∞	2	2	1
$j=2$	∞	∞	∞	∞	2	2	1	2	2
$j=3$	∞	∞	2	2	1	2	2	∞	∞
$j=4$	2	2	1	2	2	∞	∞	∞	∞
$j=5$	1	2	2	∞	∞	∞	∞	∞	∞

Fig. 5 *Left*: an illustration of Vesztergombi's construction with $m_2=\frac{4}{3}n-O(1)$. Some diagonals of lengths d_1 and d_2 are shown (*solid and dotted*, respectively). *Right*: the corresponding configuration; again, entry x in column i, row j indicates $D[i,j]=\{x\}$

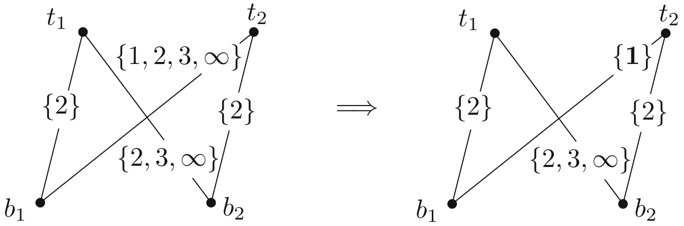


Fig. 6 A typical step of the algorithm, using Fact 3.1

Our algorithm also makes use of another easy observation. In any instance S, it cannot be true that both $d_1+d_3>d_2+d_2$ and $d_1+d_3<d_2+d_2$. Hence, using Fact 3.1, a quadruple t,t',b',b (in that cyclic order) with $|tb|=|t'b'|=d_2, |tb'|=d_1, |t'b|=d_3$ cannot co-exist with another quadruple $\hat{t},\hat{t}',\hat{b}',\hat{b}$ with $|\hat{t}\hat{b}|=d_1, |\hat{t}'\hat{b}'|=d_3, |\hat{t}\hat{b}'|=|\hat{t}'\hat{b}|=d_2$. More generally, given a configuration, we can deduce from any i,j,i',j' with each $D[i,j], D[i,j'], D[i',j], D[i',j']$ singletons other than $\{\infty\}$ that an

inequality of the form $d_w + d_x > d_y + d_z$ is true; in testing a configuration for validity, our program will reject any configuration where a contradiction arises from the set of all such pairwise inequalities. This is done by testing the associated digraph of $\binom{k+1}{2}$ pairs for acyclicity. (We also include arcs of the form $d_x + d_y > d_x + d_z$ whenever $y < z$.)

In some situations none of these facts is applicable; say, for example, if each $D[i, j]$ is equal to $\{1, 2, \infty\}$, we cannot conclude any further information. In this case, we use an approach that is similar to recursion or *branch-and-bound* in this situation, which works as follows. Find some i, j with $|D[i, j]| > 1$, and let X denote $D[i, j]$. We then replace this configuration with two new configurations: Each of the new ones is almost identical to the original, except that in one we take $D[i, j] = \min_{x\in X} x$ and in the other we take $D[i, j] = X\backslash\{\min_{x\in X} x\}$. In a little more detail, while we are examining the levels from 1 to L, we only perform branching on diagonals in levels 1 to L, (i.e., only when $1 \le j - i \le L$) and any other nonsingleton $D[i, j]$ does not entail branching. This was faster in practice than branching on every $D[i, j]$.

4.4 Initializing and Growing Configurations

Recall that our theorems are all of the following form, for a set T of positive integers and some real α:

$$\sum_{t\in T} m_t \le \alpha n + O(1). \qquad (\spadesuit)$$

We call a *target distance* any distance d_t with $t \in T$. We use k to represent the largest number in T.

We begin this detailed section by explaining why it suffices to examine configurations of bounded size to bound the number of target distances in L consecutive levels. The key tool is Fact 3.3. Namely, suppose t_0b_1 is any diagonal in level 1 with length $|t_0b_1| \ge d_k$, and consider any top-k distance diagonal e in levels $1, \ldots, L$. If e crosses t_0b_1, then t_0 (resp., b_1) is within L steps along the boundary from an endpoint of e (resp., the other endpoint of e). If e and t_0b_1 don't cross, one endpoint of e is at most $2k$ steps from t_0 or b_1 by Fact 3.3, and the other endpoint of e is at most $2k + L$ points away from the other of t_0 or b_1. Summarizing, in either case, e has one endpoint in the interval I_t consisting of vertices at most $2k + L$ steps from t_0, and e's other endpoint lies in the interval I_b consisting of vertices at most $2k + L$ steps from b_1; and this holds for all top-k distance diagonals e in levels $1, \ldots, L$.

Our program makes valid deductions whenever these intervals are disjoint, which is false only when t_0 and b_1 are within $2(2k + L)$ steps of one another on the boundary. Set $\ell = 2(2k + L)$ and define a *special diagonal* to be one with length $\ge d_k$ and at most ℓ vertices between its endpoints. Recall that $|t_0b_1| \ge d_k$, so the program's deductions are valid unless there was a special diagonal. This explains the choice of $16 = 2(2 \cdot 2 + 4)$ in Proposition 4.1 and justifies our general approach.

In the rest of this section, we explain some of the implementation details. The program begins working with a configuration consisting of a single diagonal t_0b_1 of length $\geq d_k$, and we assume without loss of generality that there are no diagonals $t_i b_{i+1}$ such that $i < 0$ and $|t_i b_{i+1}| \geq d_k$. Thus, the top and bottom intervals begins as the singleton sets $\{t_0\}, \{b_1\}$.

We will now enlarge these configurations. Reviewing our proof strategy, the program must enumerate all possible configurations such that level 1 has more than α diagonals of a target length, *and* levels 1 and 2 together have more than 2α, etc., with the hope being that once the number of levels is high enough, we find that no such configurations exist, since this would give a result like Lemma 4.2.

Note that, by our choice of t_0 and b_1, which normalize our indices, in any convex point set, all level-1 diagonals of the target distances are of the form $t_i b_{i+1}$ for $i > 1$, and by Fact 3.3, they also satisfy $i \leq 2k-2$. So crucially, their possible positions are confined to an interval of bounded size. We now determine which of these diagonals have target lengths by *exhaustive guessing*, a term that simply means trying all possibilities. In detail, first, exhaustively guess the smallest $i > 0$ for which $t_i b_{i+1}$ is a target distance, then the second-smallest, etc. When the top and bottom intervals are enlarged, each new $D[i,j]$ is set to $\{1,\ldots,k,\infty\}$ by default, meaning that no assumptions are made on the distance. When i is guessed as a minimal new level-1 diagonal for which $t_i b_{i+1}$ is a target distance, rather than the defaults, we set $D[i,i+1] = T$ and $D[i',i'+1] := \{1,\ldots,k,\infty\} \setminus T$ for all new $i' < i$.

After each new diagonal is added, we reapply Facts 3.1–3.4 in order to make additional deductions and eliminate any impossible configuration; and we split any nonsingleton sets D in the first level, as described earlier.

After this exhaustive guessing, we have collected all possible configurations. We keep only those for which level 1 has more than α diagonals of the target lengths. If any exist, we grow them in all possible ways to 2-level configurations, using exhaustive guessing like that explained above, except that we expand "to the left" before expanding "to the right" (for level 1, only rightward expansion was needed due to our choice of t_0 and b_1). Again, we prune those that have no more than 2α target distance in the first two levels.

We repeat the process described in the previous paragraph over and over, increasing the number of levels by 1 each time. If the program terminates eventually, it implies a result of the form like Lemma 4.2 and consequently that (♠) holds for this choice of T and α. We give a high-level review of the algorithm in Fig. 7.

5 Results: Proof of Theorem 4

Each row in Table 1 corresponds to an execution of our program that terminated. In other words, each execution establishes that an analogue of Lemma 4.2 holds, and we consequently deduce Theorem 4 using reasoning as in the proof of Proposition 4.1. Each line proves

$$\sum_{t \in T} m_t \leq \alpha n \text{ for } n > C(k, 2(2k+L))/(\alpha - 1), \tag{$\clubsuit$}$$

- Initialize a configuration with intervals $\{t_0\}, \{b_1\}$ and $D[0,1]$ set to T (all target distances)
- For $L = 1, 2, \ldots$
  - Extend the configurations by exhaustively guessing all diagonals of target lengths in level L, extending leftwards first if $L > 1$, and then rightwards in all cases.
  - Keep only configurations with more than αL target distances in levels $1, \ldots, L$.
  - Stop if no configurations remain.
- Upon extending a configuration, **check** it:
  - Use Facts 3.1–3.4 to perform deductions.
  - Check that distance pairs are consistent.
  - If $|D[i,j]| > 1$ for some diagonal $t_i b_j$ in one of the first L levels, partition it into two configurations and **check** both (recursively).

Fig. 7 Sketch of the algorithm

Table 1 The terminating executions of our program, each one proving (♣) for that α and T. *Tight* means convex point sets are known with $\sum_{t \in T} m_t = \alpha n - O(1)$, and *abstractly tight* means some periodic configuration has $\sum_{t \in T} m_t = \alpha n$, but we could not realize it convexly in the plane

T	α	L	Time (s)	Tightness of result
$\{1,2\}$	2	2	<1	Tight (odd regular)
$\{2\}$	4/3	4	<1	Tight [7]
$\{1,2,3\}$	3	3	<1	Tight (odd regular)
$\{3\}$	3/2	9	5	Abstractly tight, Fig. 8
$\{2,3\}$	9/4	6	1	Abstractly tight, Fig. 9
$\{1,3\}$	2	4	<1	Tight (odd regular)
$\{1,2,3,4\}$	4	3	68	Tight (odd regular)
$\{4\}$	13/8	27	50,890	Unknown

where k is the largest element of T, and C is the constant from Lemma 1.5. Note that the first two lines of Table 1 correspond to results that were already known. The running times are from a computer with a 2-GHz processor. The program was written in Java and is available on SourceForge.[2] For $T = \{1,2,3,4,5\}$ or $T = \{5\}$, the program ran out of memory before obtaining any reasonable result.

6 Abstract Tightness

Our computer program can also generate tight examples. In Fig. 8, we show two periodic configurations with $m_3 = \frac{3}{2}n$ with periods of six and eight levels, respectively. (No other example has period lower than 14.) We were not able to embed these examples as convex point sets in the plane, and at the same time we

---

[2] http://sourceforge.net/projects/convexdistances/.

∞	∞	∞	∞	∞	∞	∞	∞	3	3	1	2	1	3	3
∞	∞	∞	∞	∞	∞	3	3	2	3	2	3	3	∞	∞
∞	∞	∞	∞	3	3	1	2	1	3	3	∞	∞	∞	∞
∞	∞	3	3	2	3	2	3	3	∞	∞	∞	∞	∞	∞
3	3	1	2	1	3	3	∞	∞	∞	∞	∞	∞	∞	∞

∞	∞	∞	∞	∞	∞	3	3	3
∞	∞	∞	∞	3	3	1	2	3
∞	∞	∞	3	2	3	2	3	∞
∞	∞	3	2	1	3	3	∞	∞
∞	∞	3	3	3	∞	∞	∞	∞
3	3	1	2	3	∞	∞	∞	∞

Fig. 8 Two unrealized periodic configurations with $m_3 = \frac{3}{2}n$. *Rows* and *columns* are two intervals of vertices, and entry i (resp., ∞) means distance d_i (resp., $< d_3$)

∞	∞	∞	3	2	2	2	3	∞
∞	∞	3	2	1	2	3	∞	∞
∞	3	2	2	2	3	∞	∞	∞

Fig. 9 An unrealized periodic configurations with $m_2 + m_3 = \frac{9}{4}n$

did not disprove that they were embeddable. Based on our attempts, it seems like there is no simple periodic embedding respecting the natural symmetries of the distance configurations. A disproof of realizability could be used in the program to get stronger results. For $m_2 + m_3 = \frac{9}{4}n$, we also have an abstractly tight periodic example that we could not realize (Fig. 9).

7 Future Directions

Our program is essentially a depth-first search; each configuration examined by the program has a unique "parent" configuration from which it was grown. Thus, it would be possible to rewrite the program so as to use a smaller amount of memory and thereby possibly obtain results with smaller α or larger k; and a distributed implementation should also be straightforward.

It would be good to come up with constructions exhibiting better lower bounds. For example, no construction is known where m_3/n is asymptotically greater than 4/3.

Our approach constitutes an abstract generalization of the original problem of bounding sums of the m_i's in convex point sets. Vesztergombi [7] considered an abstraction as well, using only a subset of the facts we applied here. Can Conjecture 1.1 of Erdős and Moser be violated in either of these abstractions?

Finally, can the functions C, C' in Lemmas 1.5 and 1.6 be improved?

Acknowledgements We thank the referees for useful feedback, and K. Vesztergombi for helpful discussions. EPFL, Lausanne, Switzerland. We gratefully acknowledge support from the Swiss National Science Foundation (Grant No. 200021-125287/1) and an NSERC Post-Doctoral Fellowship.

References

1. E. Altman, On a problem of P. Erdős. Am. Math. Mon. **70**(2), 148–157 (1963)
2. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
3. H. Edelsbrunner, P. Hajnal, A lower bound on the number of unit distances between the vertices of a convex polygon. J. Comb. Theor. A **56**(2), 312–316 (1991)
4. P. Erdős, L. Lovász, K. Vesztergombi, On the graph of large distances. Discrete Comput. Geom. **4**, 541–549 (1989)
5. H. Hopf, E. Pannwitz, Aufgabe Nr. 167. Jahresbericht Deutsch. Math.-Verein. **43**, 114 (1934)
6. K. Vesztergombi, On large distances in planar sets. Discrete Math. **67**, 191–198 (1987)
7. K. Vesztergombi, On the distribution of distances in finite sets in the plane. Discrete Math. **57**, 129–145 (1985)