SERIE B — INFORMATIK

On the minimum number of empty polygons in planar point sets*

Pavel Valtr[†‡]

B 92–01
January 1992

Abstract

We describe a configuration (related to Horton's constructions) of n points in general position in the plane with less than $1.8n^2$ empty triangles, less than $2.42n^2$ empty quadrilaterals, less than $1.46n^2$ empty pentagons, and less than $n^2/3$ empty hexagons. It improves the constants shown by Bárány and Füredi.

*Work on this paper was partially done when the author participated at the workshop Uniformity and Irregularity of Partitions, University Bielefeld, Bielefeld, Germany
†Department of Applied Mathematics, Charles University, Malostranské nám. 25, 118 00 Praha 1, Czechoslovakia.
‡Graduiertenkolleg "Algorithmische Diskrete Mathematik", Fachbereich Mathematik, Freie Universität Berlin, Arnimallee 2-6, W 1000 Berlin 33, Germany, supported by "Deutsche Forschungsgemeinschaft", grant We 1265/2-1

1 Introduction

We say that a set P of points in the plane is in *general position* if no three points of P lie on a line. Erdős and Szekeres [ES 35] proved that for any k there is an integer $n(k)$ such that any set of $n(k)$ points in general position in the plane contains k points which are vertices of a convex k-gon.

We call a subset A of k points in P *an empty k-gon* if the convex hull of A contains no point of P in its interior. Erdős [Er 75] asked whether the following sharpening of the Erdős-Szekeres theorem is true. Is there an $N(k)$ such that any set of $N(k)$ points in general position in the plane contains an empty k-gon? He pointed out that $N(4) = 5$ and Harborth [Ha 78] proved $N(5) = 10$. On the other hand, Horton [Ho 83] showed that $N(k)$ does not exist for $k \geq 7$. The question about the existence of $N(6)$ is still open.

Denote by $f_k(P)$ the number of empty k-gons in P and let $f_k(n) = \min\{f_k(P) : |P| = n \text{ and } P \text{ is in general position}\}$. Katchalski and Meir [KM 87] proved that there is a constant $K < 200$ such that for any $n \geq 3$

$$\binom{n-1}{2} \leq f_3(n) \leq Kn^2.$$

Horton [Ho 83] constructed configurations giving $f_k(n) = 0$, for $k \geq 7$. Bárány and Füredi [BF 87] proved

$$n^2 - O(n \log n) \leq f_3(n) \leq 2n^2,$$

$$\frac{1}{4}n^2 - O(n) \leq f_4(n) \leq 3n^2,$$

$$\left\lfloor \frac{n}{10} \right\rfloor \leq f_5(n) \leq 2n^2,$$

$$f_6(n) \leq \frac{1}{2}n^2.$$

They proved the upper bounds only when n is a power of 2. However, one can prove them with a bit more effort for any integer n. To show the upper bounds Bárány and Füredi used the construction of Horton [Ho 83] giving $f_k(n) = 0$, for $k \geq 7$.

In Section 2 we describe two simple random configurations where the expected number of empty triangles is $2n^2 + o(n^2)$.

In Section 3 we show a construction giving the following better upper bounds:

$$f_3(n) < 1.8n^2, \qquad f_4(n) < 2.42n^2,$$

$$f_5(n) < 1.46n^2, \qquad f_6(n) < \frac{1}{3}n^2.$$

Note that the construction in Section 3 is a simplified version of a complicated construction which gives still a bit better estimations (see also remark at the end of the paper).

2 Random constructions

Bárány and Füredi [BF 87] proved that the following random construction gives a similar upper bound of $f_3(n)$ as Horton's construction.

Theorem 1 *Let $I_1, I_2, ..., I_n$ be parallel unit intervals in the plane, $I_i = \{[x,y] : x = i, 0 \leq y \leq 1\}$. For any i, choose a random point p_i from I_i. Then the expected number of empty triangles in the set $P = \{p_1, p_2, ..., p_n\}$ is at most $2n^2 + \mathcal{O}(n \log n)$.*

In the following we show that another random construction gives a similar result:

Theorem 2 *Let K be a bounded convex area in the plane. Let P be a set of n points placed randomly (with uniform distribution) and independently inside K. Then the expected number of empty triangles in P is at most $2n^2 - 2n$.*

Proof. Without loss of generality, assume the area of K equals 1. Consider two points p_i, p_j from the set $P = \{p_1, p_2, ..., p_n\}$, and denote the Euclidean distance between p_i and p_j by l. Define the axes so that $p_i = [0,0]$ and $p_j = [l,0]$. Let S_{ij} be the strip of width l between the $y-$axis and the line $x = l$. For all triangles $p_ip_jp_k$ with the longest side p_ip_j, the vertex p_k lies obviously inside S_{ij}. The expected number of points p_k from $P \cap S_{ij}$ such that $p_ip_jp_k$ is an empty triangle can be easily estimated. For any real number y, define the line segment $I_y = \{[x,y] : 0 \leq x \leq l\}$, and let $|I_y \cap K|$ denote the length of the line segment $I_y \cap K$. If $|y| > \frac{2}{l}$ then $I_y \cap K = \emptyset$ (otherwise the area of K exceeds 1). For any $k, 1 \leq k \leq n, k \neq i, k \neq j$,

$$\text{Prob}(p_k \in S_{ij} \text{ and } p_ip_jp_k \text{ is an empty triangle}) =$$

$$= \int_{-\infty}^{\infty} |I_y \cap K| \cdot \text{Prob}(p_ip_jp_k \text{ is empty} \mid p_k \in I_y) \ dy =$$

$$= \int_{-\frac{2}{l}}^{\frac{2}{l}} |I_y \cap K| \cdot (1 - \frac{l \cdot |y|}{2})^{n-3} dy \leq \int_{-\frac{2}{l}}^{\frac{2}{l}} l \cdot (1 - \frac{l \cdot |y|}{2})^{n-3} dy = \frac{4}{n-2}.$$

Hence, for any pair $\{i,j\}$, the expected number of empty triangles $p_ip_jp_k$, where p_ip_j is the longest side, is at most $\frac{4}{n-2}(n-2) = 4$, and the overall expected number of empty triangles is at most $4\binom{n}{2} = 2n^2 - 2n$. □

The method from the proof of Theorem 2 can be extended to the higher dimension for the counting of the number of empty simplices.

Note that the estimations of the number of empty triangles for the above three configurations (Horton's construction, the random constructions from Theorems 1 and 2) are the best possible in the sense that the (expected) number of empty triangles in each of them is at least $2n^2 - o(n^2)$. In Section 3 we show a configuration with a smaller number of empty triangles.

3 Construction

We start with Horton's construction: For any positive integer n, we will define a point set $H(n)$ of n points. In $H(n)$ the set of the first coordinates is just $\{0, 1, ..., n-1\}$. First we define by induction a set $H(n)$ when n is a power of 2. Let $H(1) = \{(0, 0)\}$ and $H(2) = \{(0, 0), (1, 0)\}$. When $H(n)$ is defined, set

$$H(2n) = \{(2x, y) : (x, y) \in H(n)\} \cup \{(2x + 1, y + d_n) : (x, y) \in H(n)\}$$

where the numbers d_n are fastly growing, say $d_n = 3^n - 1$. These sets $H(n)$ are just the sets defined by Horton [Ho 83]. Now let n be a positive integer, and let n' be the least power of 2 which is not smaller than n. Set

$$H(n) = \{(x, y) \in H(n') : x < n\}.$$

All $y-$coordinates of points of $H(n)$ are smaller than 3^n. The building blocks of our construction are sets $Q(n)$ which are obtained from $H(n)$ by replacing each point (x, y) by $(x, (12 + n)^{-1}3^n y)$. Obviously, all points of $Q(n)$ lie in the $(12 + n)^{-1}-$neighborhood of the $x-$axis ($(12 + n)^{-1}$ is no specific number; it is only a sufficiently small positive number). Now let $m = 4n$ be a positive integer divisible by 4 (for simplicity). We construct an $m-$point set S_m in the following way:

$$S_m = Q_1 \cup Q_2 \cup Q_3 \cup Q_4,$$

where

$$Q_1 = Q(n), \quad Q_2 = Q(n) + (\frac{1}{4}, 1),$$

$$Q_3 = Q(n) + (0, 2), \quad Q_4 = Q(n) + (\frac{1}{4}, 3).$$

$Q(n) + (a, b)$ denotes the set $Q(n)$ shifted by the vector (a, b). So the points of S_m lie in the $(12+n)^{-1}-$neighborhoods of points of the set $\overline{S}_m = N \cup (N + (\frac{1}{4}, 1)) \cup (N + (0, 2)) \cup (N + (\frac{1}{4}, 3))$, where $N = \{(0, 0),\ (1, 0),\ ...,\ (n - 1, 0)\}$. Note now that the number $(12 + n)^{-1}$ is small enough in order that the set S_m is combinatorially equivalent to the set $\overline{S}_m$, except that the sets $Q_i, i = 1, 2, 3, 4$, do not lie on a line.

The shifts $(\frac{1}{4}, 1),\ (0, 2),\ (\frac{1}{4}, 3)$ in the definition of S_m were chosen to ensure that e.g. no triangle with one point in Q_4 and two points in Q_1 is empty. This, and some related properties are used in the proof of the Lemma below.

Define, for any $s \geq 3$, the following two sets:

$$G_s(3) = \{g : g \text{ is an empty } s - \text{gon in } Q_1 \cup Q_2 \cup Q_3, g \cap Q_1 \neq \emptyset, g \cap Q_3 \neq \emptyset\},$$

$$G_s(4) = \{g : g \text{ is an empty } s - \text{gon in } Q_1 \cup Q_2 \cup Q_3 \cup Q_4, g \cap Q_1 \neq \emptyset, g \cap Q_4 \neq \emptyset\}.$$

Lemma 3

$$\begin{array}{ll} |G_3(3)| < 3n^2, & |G_3(4)| \leq \frac{8}{3}n^2, \\ |G_4(3)| < 3n^2, & |G_4(4)| \leq \frac{8}{3}n^2, \\ |G_5(3)| < n^2, & |G_5(4)| \leq \frac{4}{3}n^2, \\ |G_6(3)| = 0, & |G_6(4)| \leq \frac{1}{3}n^2. \end{array}$$

Proof. For $i = 1,2,3,4$, denote the elements of Q_i by $q_{i,j}, j = 1,2,...,n$ in the order according to their $x-$coordinates. First we estimate the sizes of the sets $G_s(4)$. Each empty $s-$gon $g \in G_s(4)$ contains only one point of Q_1 and only one point of Q_4. For $i,j = 1,2,...,n$, we can easily count the number of empty polygons g such that $g \cap Q_1 = \{q_{1,i}\}$ and $g \cap Q_4 = \{q_{4,j}\}$.

If $i \equiv j (\mathrm{mod}\ 3)$, then $g \subseteq \{q_{1,i},\ q_{2,\frac{2i+j}{3}},\ q_{3,\frac{i+2j}{3}},\ q_{4,j}\}$ and g is one of the two empty triangles $q_{1,i}\ q_{2,\frac{2i+j}{3}}\ q_{4,j}$ and $q_{1,i}\ q_{3,\frac{i+2j}{3}}\ q_{4,j}$ or the empty quadrilateral $q_{1,i}\ q_{2,\frac{2i+j}{3}}\ q_{3,\frac{i+2j}{3}}\ q_{4,j}$.

If $i \equiv j - 1(\mathrm{mod}\ 3)$, then $g \subseteq \{q_{1,i},\ q_{2,\lfloor\frac{2i+j}{3}\rfloor},\ q_{3,\lceil\frac{i+2j}{3}\rceil},\ q_{4,j}\}$ and g is again one of two certain triangles or a certain quadrilateral.

If $i \equiv j + 1(\mathrm{mod}\ 3)$, then $g \subseteq \{q_{1,i},\ q_{2,\lfloor\frac{2i+j}{3}\rfloor},\ q_{2,\lceil\frac{2i+j}{3}\rceil},\ q_{3,\lfloor\frac{i+2j}{3}\rfloor},\ q_{3,\lceil\frac{i+2j}{3}\rceil},\ q_{4,j}\}$ and g is one of four triangles, six quadrilaterals, and four pentagons, or a hexagon.

There are $\left\lfloor\frac{n^2}{3}\right\rfloor$ pairs $\{i,j\}$ such that $i \equiv j + 1(\mathrm{mod}\ 3)$. Therefore

$$|G_3(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 4 + \left\lceil\frac{2n^2}{3}\right\rceil \cdot 2 \leq \frac{8}{3}n^2,$$

$$|G_4(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 6 + \left\lceil\frac{2n^2}{3}\right\rceil \cdot 1 \leq \frac{8}{3}n^2,$$

$$|G_5(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 4 \leq \frac{4}{3}n^2,$$

$$|G_6(4)| = \left\lfloor\frac{n^2}{3}\right\rfloor \cdot 1 \leq \frac{1}{3}n^2.$$

Now we estimate the sizes of the sets $G_s(3)$. Each empty $s-$gon $g \in G_s(3)$ contains either one or two consecutive points from Q_1. In the second case the points from g are from one of the $\frac{n(n-1)}{2}$ sets

$$\{q_{1,i},\ q_{1,i+1},\ q_{2,i+\Delta},\ q_{2,i+\Delta+1},\ q_{3,i+2\Delta+1}\},\ \ 1 \leq i \leq n-1,\ \ \left\lceil -\frac{i}{2}\right\rceil \leq \Delta \leq \left\lfloor\frac{n-i-1}{2}\right\rfloor.$$

Each of these sets contains one triangle ($q_{1,i}\ q_{1,i+1}\ q_{3,i+2\Delta+1}$) from $G_3(3)$, two quadrilaterals ($q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta}\ q_{3,i+2\Delta+1}$ and $q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta+1}\ q_{3,i+2\Delta+1}$) from $G_4(3)$, and one pentagon ($q_{1,i}\ q_{1,i+1}\ q_{2,i+\Delta}\ q_{2,i+\Delta+1}\ q_{3,i+2\Delta+1}$) from $G_5(3)$.

Consider now the empty $s-$gons $g \in G_s(3)$ containing only one point of the set Q_1. Most of these polygons are contained in one of the $\frac{n(n-1)}{2}$ sets

$$\{q_{1,i},\ q_{2,i+\Delta-1},\ q_{2,i+\Delta},\ q_{3,i+2\Delta-1},\ q_{3,i+2\Delta}\},\ \ 1 \leq i \leq n,\ \ \left\lceil\frac{2-i}{2}\right\rceil \leq \Delta \leq \left\lfloor\frac{n-i}{2}\right\rfloor.$$

Each of these sets contains 5 triangles from $G_3(3)$, 4 quadrilaterals from $G_4(3)$, and one pentagon from $G_5(3)$.

For odd $i > 1$, the points of g can be also from the set $\{q_{1,i},\ q_{2,\frac{i-1}{2}},\ q_{2,\frac{i+1}{2}},\ q_{3,1}\}$. There are $\lfloor\frac{n-1}{2}\rfloor$ such sets, each with two triangles from $G_3(3)$ and one quadrilateral from $G_4(3)$. For $i = 1$, we have to consider only the triangle $\{q_{1,1}\ q_{2,1}\ q_{3,1}\}$.

If $i \not\equiv n \pmod 2$, then the points of g can be still from the set $\{q_{1,i},\ q_{2,\frac{n+i-1}{2}},\ q_{2,\frac{n+i+1}{2}},\ q_{3,n}\}$. There are $\lfloor \frac{n}{2} \rfloor$ such sets, each with two triangles from $G_3(3)$ and one quadrilateral from $G_4(3)$.

The required bounds follow:

$$|G_3(3)| = \frac{n(n-1)}{2} + \frac{n(n-1)}{2} \cdot 5 + \left\lfloor \frac{n-1}{2} \right\rfloor \cdot 2 + 1 + \left\lfloor \frac{n}{2} \right\rfloor \cdot 2 < 3n^2,$$

$$|G_4(3)| = \frac{n(n-1)}{2} \cdot 2 + \frac{n(n-1)}{2} \cdot 4 + \left\lfloor \frac{n-1}{2} \right\rfloor + \left\lfloor \frac{n}{2} \right\rfloor < 3n^2,$$

$$|G_5(3)| = \frac{n(n-1)}{2} + \frac{n(n-1)}{2} < n^2,$$

$$|G_6(3)| = 0.$$

□

Theorem 4

$$f_3(S_m) < 1.8m^2, \qquad f_4(S_m) < 2.42m^2,$$

$$f_5(S_m) < 1.46m^2, \qquad f_6(S_m) < \frac{1}{3}m^2.$$

Proof. Let P be a point set in the plane and consider two points $u_1, u_2 \in P, u_1 = (x_1, y_1), u_2 = (x_2, y_2), x_1 < x_2$. We say that the line segment u_1u_2 is *open from below* if there is no point of P inside the strip $S = \{(x,y) : x_1 < x < x_2 \text{ and } (x,y) \text{ lies below the line } u_1u_2\}$. A subset X of P is called *open from below* if all the line segments connecting two points of X are open from below. Analogously, we define *open from above.*

For any positive integer r, denote by $h_r^-(P)$ and $h_r^+(P)$ the number of $r-$point subsets in P empty from below and above, respectively.

Bárány and Füredi [BF 87] showed

$$h_2^-(H(n)) < 2n, \qquad h_2^+(H(n)) < 2n,$$

and

$$h_3^-(H(n)) < n, \qquad h_3^+(H(n)) < n.$$

They proved the above inequalities when n is a power of 2. However, one can prove them for any positive integer n.

The construction of $H(n)$ is done so that, for any $r > 3$,

$$h_r^-(H(n)) = h_r^+(H(n)) = 0.$$

Obviously, all the above relations are satisfied for the set $H(n)$ as well as for the sets $Q(n)$ and $Q_i, i = 1,2,3,4$. For any $s \geq 3$, and any $r, 0 < r < s$, the number of empty $s-$gons G in $Q_1 \cup Q_2$ with $|G \cap Q_1| = r$ is equal to $h_r^+(Q_1) \cdot h_{s-r}^-(Q_2)$.

This is carried out by the construction (more precisely, by the fact that the set Q_2 lies entirely above any line containing two points of Q_1 and similarly the set Q_1 lies entirely below any line containing two points of Q_2). Thus

$$f_s(Q_1 \cup Q_2) = f_s(Q_1) + f_s(Q_2) + \sum_{r=1}^{s-1} h_r^+(Q_1) \cdot h_{s-r}^-(Q_2).$$

Since $f_s(Q_2 \cup Q_3) = f_s(Q_1 \cup Q_2)$ and $f_s(Q_2) = f_s(Q_1)$ we obtain

$$f_s(Q_1 \cup Q_2 \cup Q_3) = f_s(Q_1 \cup Q_2) + f_s(Q_2 \cup Q_3) - f_s(Q_2) + g_s(3) =$$

$$= 2f_s(Q_1 \cup Q_2) - f_s(Q_1) + g_s(3) = 3f_s(Q_1) + 2\sum_{r=1}^{s-1} h_r^+(Q_1) \cdot h_{s-r}^-(Q_2) + g_s(3).$$

Similarly

$$f_s(Q_1 \cup Q_2 \cup Q_3 \cup Q_4) = f_s(Q_1 \cup Q_2 \cup Q_3) + f_s(Q_2 \cup Q_3 \cup Q_4) - f_s(Q_2 \cup Q_3) + g_s(4) =$$

$$2(3f_s(Q_1) + 2\sum_{r=1}^{s-1} h_r^+(Q_1) \cdot h_{s-r}^-(Q_2) + g_s(3)) - (2f_s(Q_1) + \sum_{r=1}^{s-1} h_r^+(Q_1) \cdot h_{s-r}^-(Q_2)) + g_s(4) =$$

$$4f_s(Q_1) + 3\sum_{r=1}^{s-1} h_r^+(Q_1) \cdot h_{s-r}^-(Q_2) + 2g_s(3) + g_s(4).$$

Now the required bounds follow:

$$f_3(S_m) < 4 \cdot 2n^2 + 3 \cdot (n \cdot 2n + 2n \cdot n) + 2 \cdot 3n^2 + \frac{8}{3}n^2 = \frac{86}{3}n^2 = 1.791...m^2,$$

$$f_4(S_m) < 4 \cdot 3n^2 + 3 \cdot (n \cdot n + 2n \cdot 2n + n \cdot n) + 2 \cdot 3n^2 + \frac{8}{3}n^2 = \frac{116}{3}n^2 = 2.416...m^2,$$

$$f_5(S_m) < 4 \cdot 2n^2 + 3 \cdot (2n \cdot n + n \cdot 2n) + 2 \cdot n^2 + \frac{4}{3}n^2 = \frac{70}{3}n^2 = 1.458...m^2,$$

$$f_6(S_m) < 4 \cdot \frac{1}{2}n^2 + 3 \cdot (n \cdot n) + 2 \cdot 0 + \frac{1}{3}n^2 = \frac{16}{3}n^2 = \frac{1}{3}m^2.$$

□

The proof that for any positive integer m (not necessarily divisible by 4) there is a set S_m satisfying Theorem 3 requires only more computation.

Remark. The author [Va 91] constructed, for any n, a set A_n of n points in general position in the plane with the following unrelated property: The ratio between the maximum and minimum distance is at most $\Theta(\sqrt{n})$, and the set A_n does not contain more than $\mathcal{O}(n^{1/3})$ vertices of a convex polygon. This is essentially the best possible result. Imre Bárány suggested that the sets A_n might be used to improve the best known upper bound of $f_3(n)$. Indeed the set A_n contains less than $1.68n^2$ empty triangles, for any large n. However, the proof of this fact which we know is involved and so we considered the simpler construction which gives slightly worse results.

Acknowledgements. The author thanks Imre Bárány and Emmerich Welzl for fruitful discussions.

References

[BF 87] I. Bárány and Z. Füredi, Empty simplices in Euclidean space, Canadian Math. Bull. 30 (1987), 436-445.

[Er 75] P. Erdős, On some problems of elementary and combinatorial geometry, Ann. Mat. Pura. Appl. (4) 103 (1975), 99-108.

[ES 35] P. Erdős and G. Szekeres, A combinatorial problem in geometry, Compositio Math. 2 (1935), 463-470.

[Ha 78] H. Harborth, Konvexe Fünfecke in ebenen Punktmengen, Elem. Math. 33 (1978), 116-118.

[Ho 83] J. D. Horton, Sets with no empty convex 7-gons, Canadian Math. Bull. 26 (1983), 482-484.

[KM 87] M. Katchalski and A. Meir, On empty triangles determined by points in the plane, Acta. Math. Hungar. (1987).

[Va 91] P. Valtr, Convex independent sets and 7-holes in restricted planar point sets, to appear in Discrete Comput. Geom. 6 (1991).