Studia Scientiarum Mathematicarum Hungarica 52 (1), 52–64 (2015)
DOI: 10.1556/SScMath.52.2015.1.1301

EMPTY NON-CONVEX AND CONVEX FOUR-GONS IN RANDOM POINT SETS

RUY FABILA-MONROY[1], CLEMENS HUEMER[2] and DIETER MITSCHE[3]

[1] Departamento de Matemáticas, CINVESTAV-IPN, México
e-mail: ruyfabila@math.cinvestav.edu.mx

[2] Universitat Politècnica de Catalunya, BarcelonaTech
e-mail: clemens.huemer@upc.edu

[3] Laboratoire J.A. Dieudonné, Université de Nice Sophia-Antipolis
e-mail: dmitsche@unice.fr

Communicated by G. Fejes Tóth

(Received June 10, 2014; accepted September 28, 2014)

Abstract

Let $\mathcal{S}$ be a set of n points distributed uniformly and independently in a convex, bounded set in the plane. A four-gon is called empty if it contains no points of $\mathcal{S}$ in its interior. We show that the expected number of empty non-convex four-gons with vertices from $\mathcal{S}$ is $12n^2 \log n + o(n^2 \log n)$ and the expected number of empty convex four-gons with vertices from $\mathcal{S}$ is $\Theta(n^2)$.

1. Introduction

Sylvester's famous four-point problem asks for the probability that four points chosen at random inside a convex set $\mathcal{K}$ in the plane form the vertices of a non-convex (also called reentrant) quadrilateral. The answer to this question depends on the shape of $\mathcal{K}$, see the historical account of Pfiefer [16]. Almost a century ago, Blaschke [10] proved that this probability is maximal if $\mathcal{K}$ is a triangle, and minimal if $\mathcal{K}$ is an ellipse. Actually, this problem is equivalent to finding the expected triangle area $\mathbb{E}_\triangle$ of three points chosen at random from $\mathcal{K}$.

More precisely, the probability that a four-gon is non-convex is four times $\mathbb{E}_\triangle$ divided by the area of $\mathcal{K}$.

2010 *Mathematics Subject Classification.* Primary 60D05, 52A22.

Key words and phrases. random point set, empty four-gon, polygon, geometric probability.

The question studied here is related: A set $\mathcal{S}$ of n points is distributed at random inside $\mathcal{K}$. What is the expected number of empty four-gons with vertices from $\mathcal{S}$?

Figure 1 shows an empty non-convex and an empty convex four-gon in $\mathcal{S}$. The probability that a four-gon is empty depends on its area, and we will make use of the density function of the triangle area of three points chosen at random from $\mathcal{K}$, see Henze [13]. Also, the expected numbers of empty non-convex and empty convex four-gons are different. Further, we observe that the asymptotic order of magnitude of the number of empty four-gons does not depend on the shape of the convex set. Such a behaviour was also obtained in related works: Valtr [17] proved that the expected number of empty triangles in $\mathcal{S}$ is at most $2n^2 - 2n$ and at least $2n^2 - o(n^2)$, independent of the shape of $\mathcal{K}$.

Bárány et al. [8] proved that in expectation $\mathcal{S}$ contains two points which both appear in $\Omega(n/\log n)$ empty triangles.

Bárány and Füredi [7] proved that the number of empty simplices in random point sets of n points distributed uniformly and independently in a convex, bounded set in R^d is at most $K\binom{n}{d}$, for some constant K.

Balogh et al. [5] showed that the expected number of vertices of the largest empty convex polygon in $\mathcal{S}$ is $\Theta\left(\frac{\log n}{\log\log n}\right)$.

Sylvester's four-point problem was generalized by asking for the probability that n points chosen at random inside $\mathcal{K}$ are in convex position [6, 15, 18, 19]. In this case the answer depends on the shape of $\mathcal{K}$.

A lot of research has been done to determine the minimum number $f_k(n)$ of empty convex k-gons among all sets of n points in general position in the plane (not only random point sets).

For the case of empty triangles, Katchalski and Meir [14] showed that $f_3(n)$ is of order $\Theta(n^2)$. Later, this bound was refined [2, 7, 9, 11, 12, 17]; the currently best bounds are $n^2 - \frac{32n}{7} + \frac{22}{7} \leqq f_3(n) \leqq 1.6196n^2 + o(n^2)$.

Concerning empty convex four-gons, Bárány and Füredi [7] established that $f_4(n)$ is of order $\Theta(n^2)$, and the currently best bounds on $f_4(n)$ are $\frac{n^2}{2} - \frac{9}{4}n - o(n) \leqq f_4(n) \leqq 1.9397n^2 + o(n^2)$, see [2, 9].

Research mainly focussed on empty convex polygons. Only recently the number of convex and non-convex polygons in point sets has been studied [1, 3, 4].

In [1] it was shown that every set of n points in general position in the plane determines at least $\frac{5n^2}{2} - \Theta(n)$ empty four-gons and a point set with only $O(n^{5/2}\log n)$ empty four-gons is given. Our result improves the latter bound to $O(n^2\log n)$.

Let us state the results more formally. All logarithms in this paper are natural logarithms. Throughout this paper let $\mathcal{S}$ be a set of n points distributed uniformly and independently in a convex, bounded set $\mathcal{K}$ of unit area in the plane. Since with probability 1 the n points will be in general position (no three points are collinear), we may and will assume this throughout the paper. All asymptotics in this paper are w.r.t. the number of points n, that is, when $n \to \infty$. As our results are asymptotic, we may ignore also rounding issues throughout the paper, that is, if for some constant $c > 0$, cn is not an integer, depending on the context, we may and will consider $\lfloor cn \rfloor$ or $\lceil cn \rceil$ without changing the results.

A four-gon whose vertices are from $\mathcal{S}$ is empty if it contains no other point from $\mathcal{S}$ in its interior.

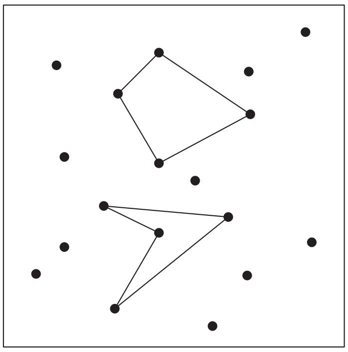

Fig. 1. An empty non-convex four-gon and an empty convex four-gon in $\mathcal{S}$

Denote by N_4 the random variable that counts the number of empty non-convex four-gons with vertices from $\mathcal{S}$.

Our main result is the following:

THEOREM 1.1. $\mathbb{E}[N_4] = 12n^2 \log n + o(n^2 \log n)$.

Denote by C_4 the random variable that counts the number of empty convex four-gons with vertices from $\mathcal{S}$.

Complementing Theorem 1.1, we prove the following result (the lower bound was already obtained in [7]. There, also another construction with $O(n^2)$ empty convex four-gons is given.)

THEOREM 1.2. $\mathbb{E}[C_4] = \Theta(n^2)$.

2. Proof of Theorem 1.1

The proof of Theorem 1.1 is implied by the following lemmas, for which we need some definitions. Fix three points p_a, p_b, p_c and denote by $\Delta(p_a, p_b, p_c)$ the triangle spanned by them. Let $\mathcal{P}$ be a set of $k \geqq 1$ points distributed uniformly and independently in $\Delta(p_a, p_b, p_c)$. Denote by $\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}$ the event that $\mathcal{P} \cup \{p_a, p_b, p_c\}$ contains an empty non-convex four-gon with $p_a p_b$ and $p_b p_c$ among its edges.

LEMMA 2.1. $\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_ap_b,p_bp_c}) = \frac{2}{k+1}$.

PROOF. First observe that the points $\{p_a, p_b, p_c\}$ together with a fourth point $p_d \in \mathcal{P}$ form an empty non-convex four-gon with p_ap_b and p_bp_c among its edges if and only if the triangle $\Delta(p_a, p_d, p_c)$ contains $\mathcal{P}\backslash\{p_d\}$ in its interior. We now determine the distribution of the height h_d, which is the distance from p_d to the segment p_ap_c. Denote by h the height of $\Delta(p_a, p_b, p_c)$ with respect to the edge p_ap_c. Let ℓ be the segment parallel to the edge p_ap_c, at distance h_d from this edge, and with endpoints on the edges p_ap_b and p_bp_c respectively. Assume w.l.o.g. that ℓ is a horizontal line with p_a and p_c below it. Define then $\Delta(\ell, p_b)$ to be the triangle coming from the intersection of $\Delta(p_a, p_b, p_c)$ and all points lying on or above ℓ, as shown in Figure 2.

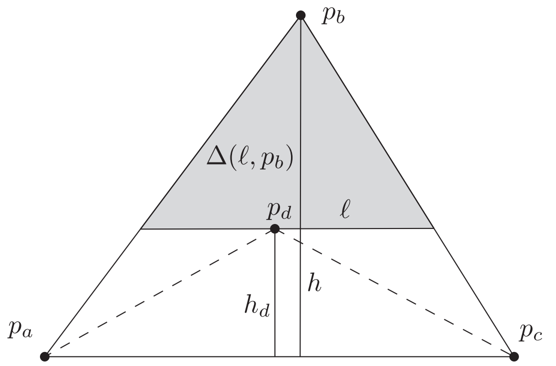

Fig. 2. The triangles $\Delta(p_a, p_b, p_c)$, $\Delta(p_a, p_d, p_c)$ and $\Delta(\ell, p_b)$

Since only relative heights between h and h_d matter, we may assume w.l.o.g. that $h = 1$.

By the intercept theorem we have $\frac{|p_ap_c|}{h} = \frac{|\ell|}{h-h_d}$, where $|p_ap_c|$ and $|\ell|$ are the lengths of the segments p_ap_c and ℓ.

It follows that

$$\frac{\text{area}\big(\Delta(\ell, p_b)\big)}{\text{area}(\Delta(p_a, p_b, p_c))} = \frac{\frac{|\ell|(h-h_d)}{2}}{\frac{|p_ap_c|h}{2}} = (1 - h_d)^2.$$

Hence, the distribution function F_{h_d} for the height h_d satisfies $F_{h_d}(x) = \mathbb{P}(h_d \leqq x) = 1 - (1 - x)^2$ and the density of the height h_d is $f_{h_d}(x) = 2 - 2x$ for $x \in [0, 1]$.

Fix any $p \in \mathcal{P} \setminus \{p_d\}$. Since the points are distributed uniformly at random inside $\Delta(p_a, p_b, p_c)$, we have

$$\mathbb{P}\big(p \in \Delta(p_a, p_d, p_c)\big) = \frac{\text{area}(\Delta(p_a, p_d, p_c))}{\text{area}(\Delta(p_a, p_b, p_c))} = h_d.$$

Therefore, integrating over all possible heights $0 \leqq h_d \leqq 1$,

$$\mathbb{P}\big(\mathcal{P}\backslash\{p_d\} \in \Delta(p_a, p_d, p_c)\big) = \int_0^1 x^{k-1}(2-2x)\,dx = \frac{2}{k(k+1)}.$$

As there are k choices for the point p_d, by a union bound, we have

$$\mathbb{P}\big(\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}\big) \leqq \frac{2}{k+1}.$$

On the other hand, there is at most one point $p_d \in \mathcal{P}$ such that $\mathcal{P}\backslash\{p_d\} \in \Delta(p_a, p_d, p_c)$: indeed, if this were true for another point $p_e \neq p_d$, then $p_d \notin \Delta(p_a, p_e, p_c)$, contradicting the assumption. Hence,

$$\mathbb{P}\Big(\big(\mathcal{P}\backslash\{p_d\} \in \Delta(p_a, p_d, p_c)\big) \wedge \big(\mathcal{P}\backslash\{p_e\} \in \Delta(p_a, p_e, p_c)\big)\Big) = 0,$$

and thus

$$\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}) = \bigcup_{p_d \in \mathcal{P}} \mathbb{P}\big(\mathcal{P}\backslash\{p_d\} \in \Delta(p_a, p_d, p_c)\big)$$

$$= \sum_{p_d \in \mathcal{P}} \mathbb{P}\big(\mathcal{P}\backslash\{p_d\} \in \Delta(p_a, p_d, p_c)\big) = \frac{2}{k+1}. \qquad \square$$

For the next lemma, we need one more definition. Let T_k denote the random variable that counts the number of triangles with vertices in $\mathcal{S}$ containing exactly $k \geqq 0$ points from $\mathcal{S}$ in its interior.

Lemma 2.2. *For any $k = k(n) \geqq 0$, $\mathbb{E}[T_k] \leqq 2n^2 - 2n$.*

Proof. Let $f_{\text{area}(\Delta)}(v)$ denote the density function for the area v of a triangle Δ formed by three points chosen uniformly and independently in the bounded, convex set $\mathcal{K}$. This density function has been studied in [13]; also see references therein. From the fact that $f_{\text{area}(\Delta)}(0) = 12$ and that $f_{\text{area}(\Delta)}(v)$ is a monotonically decreasing function, we have that for any area $v \geqq 0$, $f_{\text{area}(\Delta)}(v) \leqq 12$.

We remark that it also follows from Lemma 5.1 of [7] that $f_{\text{area}(\Delta)}(v)$ is bounded from above by a constant.

Fix now three points $p_a, p_b, p_c \in \mathcal{S}$, let as before $\Delta(p_a, p_b, p_c)$ be the triangle spanned by them, and let $\operatorname{int}\big(\Delta(p_a, p_b, p_c)\big)$ denote the interior of this triangle. Denote also for $x, y > 0$ by $\beta(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ the beta function of x and y.

Integrating over all possible areas v of the triangle $\Delta(p_a, p_b, p_c)$, we obtain

$$\mathbb{P}\Big(\big|\mathrm{int}\,\big(\Delta(p_a,p_b,p_c)\big)\cap\mathcal{S}\big| = k\Big) = \int\limits_0^1 \binom{n-3}{k} v^k (1-v)^{n-3-k} f_{\mathrm{area}(\Delta)}(v)\,dv$$

$$\leqq 12 \int\limits_0^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\,dv$$

$$= 12\binom{n-3}{k}\beta(k+1, n-k-2)$$

$$= 12\binom{n-3}{k}\frac{k!(n-k-3)!}{(n-2)!} = \frac{12}{n-2}.$$

Hence, by linearity of expectation, for any $k = k(n) \geqq 0$,

$$\mathbb{E}\,[T_k] \leqq \binom{n}{3}\frac{12}{n-2} = 2n^2 - 2n. \qquad \square$$

Remark. The special case $k = 0$ of Lemma 2.2 was also proved by Valtr in [17].

Lemma 2.3. *For every $\varepsilon > 0$ there exists some $\alpha = \alpha(\varepsilon) > 0$ such that $\mathbb{E}\,[T_k] \geqq (2-\varepsilon)n^2$ for any $k = 0, 1, \dots, \alpha n$.*

Proof. The density function $f_{\mathrm{area}(\Delta)}(v)$ for the area of the triangle $\Delta = \Delta(p_a, p_b, p_c)$, formed by three points p_a, p_b, p_c from $\mathcal{S}$, satisfies $f_{\mathrm{area}(\Delta)}(0) = 12$ and is then strictly monotonically decreasing. In particular, for every small $\varepsilon > 0$ there exists $v_0 > 0$ such that $f_{\mathrm{area}(\Delta)}(v_0) = 12 - \varepsilon$. We define $\alpha = 0.6v_0$.

$$\mathbb{P}\Big(\big|\mathrm{int}(\Delta)\cap\mathcal{S}\big| = k\Big) = \int\limits_0^1 \binom{n-3}{k} v^k (1-v)^{n-3-k} f_{\mathrm{area}(\Delta)}(v)\,dv$$

$$\geqq (12-\varepsilon)\int\limits_0^{v_0} \binom{n-3}{k} v^k (1-v)^{n-3-k}\,dv.$$

As in the proof of Lemma 2.2 we have

$$\int_0^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\, dv = \frac{1}{n-2}.$$

We will show that $\mathbb{P}\big(\big|\operatorname{int}(\Delta)\cap\mathcal{S}\big| = k\big) \geqq \frac{12-\varepsilon}{n-2} - o\big(\frac{1}{n}\big)$, for $k \leqq \alpha n$. To this end, it is sufficient to show that

$$\int_{v_0}^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\, dv = o\left(\frac{1}{n}\right). \tag{1}$$

Computing the derivative of the function $g(v) := v^k(1-v)^{n-3-k}$ in $[v_0, 1]$, we see that $g'(v) \leqq 0$, implying that in $[v_0, 1]$, $g(v)$ is maximized at $v = v_0$.

Thus,

$$\int_{v_0}^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\, dv \leqq \binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k}(1-v_0).$$

It is easily verified that

$$\binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k} < \binom{n-3}{k+1} {v_0}^{k+1} (1-v_0)^{n-3-k-1}$$

holds for $k < v_0(n-2) - 1$.

If we can show that

$$\binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k}(1-v_0) = o\left(\frac{1}{n}\right)$$

holds for $k = \alpha n$, then it holds for all smaller values of k as well, and then also (1) holds for all $k = 0, 1, \ldots, \alpha n$. Using $\binom{n}{k} \leqq \big(\frac{ne}{k}\big)^k$, we get for some constant $C > 0$

$$\binom{n-3}{\alpha n} {v_0}^{\alpha n} (1-v_0)^{n-3-\alpha n}(1-v_0)$$

$$\leqq C \left(\frac{e}{0.6 v_0}\right)^{0.6 v_0 n} v_0^{0.6 v_0 n} (1-v_0)^{(1-0.6v_0)n}$$

$$= C\big((e/0.6)^{0.6v_0}(1-v_0)^{1-0.6v_0}\big)^n$$
$$= o\left(\frac{1}{n}\right),$$

where the last line follows from the facts that

$$f(v_0) := \big(\frac{e}{0.6}\big)^{0.6v_0}(1-v_0)^{1-0.6v_0}$$

is monotone decreasing for $v_0 \in [0,1]$, that $v_0 > 0$, and that $f(0) = 1$. Thus,

$$\mathbb{P}\big(\big|\operatorname{int}(\Delta) \cap \mathcal{S}\big| = k\big) \geqq (12-\varepsilon) \int_0^{v_0} \binom{n-3}{k} v^k (1-v)^{n-3-k}\, dv$$
$$\geqq \frac{12-\varepsilon}{n-2} - o\left(\frac{1}{n}\right).$$

As before, by linearity of expectation,

$$\mathbb{E}\left[T_k\right] \geqq \binom{n}{3}\left(\frac{12-\varepsilon}{n-2} - o\left(\frac{1}{n}\right)\right) \geqq (2-\varepsilon)n^2$$

for any $k = 0, 1, \ldots, \alpha n$, thus concluding the proof. □

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. Note that each triangle $\Delta(p_a, p_b, p_c)$ with vertices $p_a, p_b, p_c \in \mathcal{S}$ determines at most three empty non-convex four-gons such that p_a, p_b, p_c are the vertices on the boundary of the convex hull of the quadrilateral: indeed, any pair of edges from $\{p_ap_b, p_bp_c, p_ap_c\}$ can be chosen and might possibly give rise to an empty non-convex four-gon. Let $\mathcal{P} \subsetneqq \mathcal{S}$ denote the set of points in the interior of a triangle $\Delta(p_a, p_b, p_c)$. We classify the empty non-convex four-gons in $\mathcal{S}$ according to the cardinality of $\mathcal{P}$. That is, we define $N_{4,k}$, for $1 \leqq k \leqq n-3$, as the random variable that counts the number of empty non-convex four-gons in all triangles $\Delta(p_a, p_b, p_c)$ with $|\mathcal{P}| = k$ points of $\mathcal{S}$ in its interior, such that the four-gons associated to a triangle $\Delta(p_a, p_b, p_c)$ have p_a, p_b, p_c among their vertices. Then $N_4 = \sum_{k=1}^{n-3} N_{4,k}$.

Let X_{k,p_ap_b,p_bp_c} be the indicator random variable for the event that a triangle $\Delta(p_a, p_b, p_c)$ with $|\mathcal{P}| = k$ points of $\mathcal{S}$ in its interior, contains an empty non-convex four-gon with p_ap_b and p_bp_c among its edges. Analogously, X_{k,p_bp_c,p_cp_a} and X_{k,p_cp_a,p_ap_b} are the indicator random variables for the events that $\Delta(p_a, p_b, p_c)$ contains an empty non-convex four-gon with $p_b p_c$ and $p_c p_a$, respectively $p_c p_a$ and $p_a p_b$, among its edges. For $X_k = X_{k,p_a p_b, p_b p_c} + X_{k,p_b p_c, p_c p_a} + X_{k,p_c p_a, p_a p_b}$ we get from Lemma 2.1 that $\mathbb{E}\left[X_k\right] = \frac{6}{k+1}$.

We have

$$\mathbb{E}\left[N_{4,k} \mid T_k = t\right] = t\mathbb{E}\left[X_k\right].$$

Then

$$\mathbb{E}\left[N_{4,k}\right] = \sum_{t=0}^{\binom{n}{3}} \mathbb{E}\left[N_{4,k} \mid T_k = t\right] \mathbb{P}(T_k = t)$$

$$= \sum_{t=0}^{\binom{n}{3}} t\mathbb{E}\left[X_k\right] \mathbb{P}(T_k = t) = \mathbb{E}\left[X_k\right] \sum_{t=0}^{\binom{n}{3}} t\mathbb{P}(T_k = t) = \mathbb{E}\left[X_k\right] \mathbb{E}\left[T_k\right]$$

and

$$\mathbb{E}\left[N_4\right] = \sum_{k=1}^{n-3} \mathbb{E}\left[N_{4,k}\right] = \sum_{k=1}^{n-3} \mathbb{E}\left[X_k\right] \mathbb{E}\left[T_k\right].$$

By Lemma 2.2,

$$\mathbb{E}\left[N_4\right] \leqq (2n^2 - 2n) \sum_{k=1}^{n-3} \frac{6}{k+1} \leqq (12n^2 - 12n) \log n.$$

By Lemma 2.3,

$$\mathbb{E}\left[N_4\right] \geqq (2-\varepsilon) n^2 \sum_{k=1}^{\alpha n} \frac{6}{k+1} \geqq 6(2-\varepsilon) n^2 \big((\log \alpha n) - 1\big).$$

Since the latter inequality holds for every $\varepsilon > 0$ and $\alpha(\varepsilon)$ we conclude that

$$\mathbb{E}\left[N_4\right] \geqq 12n^2 \log n - o(n^2 \log n).$$

This completes the proof of Theorem 1.1. □

3. Proof of Theorem 1.2

The following proof roughly follows the proof of Valtr [17] for the upper bound of $2n^2 - 2n$ empty triangles in $\mathcal{S}$.

PROOF. For each convex four-gon we focus on its largest (interior) diagonal.

Let $p_a, p_b \in \mathcal{S}$ and consider four-gons whose largest diagonal is $p_a p_b$ of length $|p_a p_b| = \ell$. Note that, since S is a set of points distributed in a bounded set, ℓ is bounded from above by a constant D, which is the diameter of $\mathcal{K}$. Suppose w.l.o.g. that for the coordinates of p_a and p_b we have $p_a = (\ell, \ell)$, $p_b = (2\ell, \ell)$. Consider the two axis-parallel rectangles R_1 and R_2 of width 3ℓ and height ℓ whose left lower cornerpoints are $(0, \ell)$, and $(0, 0)$, respectively, as shown in Figure 3.

Observe that if $p_a p_b$ is the largest diagonal of a convex four-gon with vertex set $\{p_a, p_b, p_c, p_d\}$, then it is necessary that the other two points p_c and p_d are in R_1 and R_2 (one in R_1, and one in R_2): indeed, since the four-gon is convex, both diagonals are inside, and hence if one of the points p_c or p_d were outside the rectangles, as the segment $p_c p_d$ has to cross the diagonal $p_a p_b$, its length were bigger than ℓ, contradicting the fact that $p_a p_b$ is the longest diagonal.

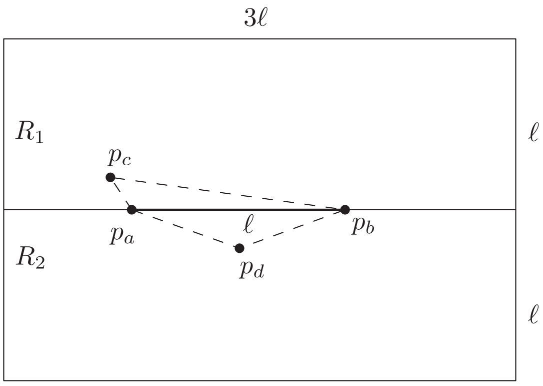

Fig. 3. The bounding box containing points p_c and p_d of a four-gon with largest diagonal $p_a p_b$ and $|p_a p_b| = \ell$

Now, fix 4 points $p_a, p_b, p_c, p_d \in \mathcal{S}$ (ordered, only the order between p_a and p_b does not matter) and define by $\mathcal{E}_{p_a p_b, p_c, p_d}$ the event that the 4 points p_a, p_b, p_c, p_d form an empty convex four-gon whose largest diagonal is formed by $p_a p_b$. Denote by $f_{\text{len}}(x)$ the probability density function that the length of the edgebetween two randomly chosen points in $\mathcal{K}$ is exactly x. Clearly,

$$f_{\text{len}}(x) \leqq 2x\pi,$$

since the position of the first point is arbitrary, and the second point then has to be on the circumference of a ball of radius x centered at the first point. Denote also by $F_{\text{hei}}(h)$ the probability that the y-coordinate of a randomly chosen point in $\mathcal{K}$ is in $[\ell, \ell+h]$,and simultaneously the x-coordinate is in $[0, 3\ell]$. Let $f_{\text{hei}}(h)$ be the corresponding probability density function of additionally having y-coordinate exactly $\ell+h$. We have

$$F_{\text{hei}}(h) \leqq 3\ell h, \qquad \text{and} \qquad f_{\text{hei}}(h) \leqq 3\ell.$$

Note that for $h \leqq \ell$, $f_{\text{hei}}(h)$ corresponds to the probability density that a randomly chosen point from $\mathcal{S} \setminus \{p_a, p_b\}$ is inside R_1, and at vertical distance h measured from $p_a p_b$ (in fact, it is an upper bound, since depending on the value of ℓ the rectangle might not be fully contained in $\mathcal{K}$). By symmetry, this is also an upper bound for the probability density function to be at vertical distance h from $p_a p_b$ and inside R_2. Call this vertical distance in both cases the *height* of a point.

Since we consider p_c and p_d to be an ordered pair, we consider for p_c only the probability density of all heights h_c in R_1, and for p_d only the probability density of all heights h_d in R_2. Assuming thatp_c and p_d are at heights h_c and h_d in the corresponding rectangles, let $\Delta(\ell, p_c)$ be the triangle with base edge $p_a p_b$ and height h_c, going through the point p_c, and analogously for $\Delta(\ell, p_d)$. Note that for $1 \leqq \ell \leqq D$, we have $h_c \leqq \frac{2}{\ell}$ and $h_d \leqq \frac{2}{\ell}$, as the areas of $\Delta(\ell, p_c)$ and $\Delta(\ell, p_d)$ are bounded by 1.

Also, by convexity, if all of p_a, p_b, p_c, p_d are inside $\mathcal{K}$,then also both triangles fall entirely into $\mathcal{K}$. We estimate the probability that there are no other points in $\mathcal{S} \setminus \{p_a, p_b, p_c, p_d\}$ that fall into the triangles (with disjoint interiors) $\Delta(\ell, p_c)$ and $\Delta(\ell, p_d)$ of total area $\frac{\ell(h_c+h_d)}{2}$. Distinguishing the two cases $0 \leqq \ell < 1$ and $1 \leqq \ell \leqq D$, we now integrate over all lengths $0 \leqq \ell \leqq D$ and all heights h_c, h_d with $0 \leqq h_c, h_d \leqq \ell$ with respect to their corresponding densities. This yields

$$\mathbb{P}(\mathcal{E}_{p_a p_b, p_c, p_d}) \leqq \int\limits_{\ell=0}^{1} \int\limits_{h_c=0}^{\ell} \int\limits_{h_d=0}^{\ell} 2\pi\ell(3\ell)^2 \left(1 - \frac{\ell(h_c+h_d)}{2}\right)^{n-4} dh_d\, dh_c\, d\ell$$

$$+ \int\limits_{\ell=1}^{D} \int\limits_{h_c=0}^{\frac{2}{\ell}} \int\limits_{h_d=0}^{\frac{2}{\ell}} 2\pi\ell(3\ell)^2 \left(1 - \frac{\ell(h_c+h_d)}{2}\right)^{n-4} dh_d\, dh_c\, d\ell,$$

and using $1 - x \leqq e^{-x}$, we obtain

$$\mathbb{P}(\mathcal{E}_{p_a p_b, p_c, p_d}) \leqq 18\pi \int\limits_{\ell=0}^{1} \int\limits_{h_c=0}^{\ell} \int\limits_{h_d=0}^{\ell} \ell^3 \exp\left(-\frac{\ell(h_c+h_d)}{2}(n-4)\right) dh_d\, dh_c\, d\ell$$

$$+ 18\pi \int\limits_{\ell=1}^{D} \int\limits_{h_c=0}^{\frac{2}{\ell}} \int\limits_{h_d=0}^{\frac{2}{\ell}} \ell^3 \exp\left(-\frac{\ell(h_c+h_d)}{2}(n-4)\right) dh_d\, dh_c\, d\ell.$$

Evaluating these two integrals separately, we get

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{p_a p_b, p_c, p_d}) &\leqq \frac{36\pi n}{(n-4)^3} + O\left(\frac{1}{n^3}\right) \\ &\quad + \frac{36(D^2-1)\pi}{(n-4)^2} + o\left(\frac{1}{n^2}\right) \\ &= \frac{36D^2\pi}{n^2} + o\left(\frac{1}{n^2}\right). \end{aligned}$$

Thus, since there are $\binom{n}{2}$ choices for the points p_a and p_b, and at most n^2 choices for the points p_c and p_d, we have

$$\mathbb{E}\left[C_4\right] \leqq \binom{n}{2} n^2 \mathbb{P}(\mathcal{E}_{p_a p_b, p_c, p_d}) \leqq 18D^2\pi n^2 + o(n^2),$$

proving the upper bound of the theorem. A lower quadratic bound is well known, see e.g. [7]. □

Acknowledgements. We thank Günter Rote for pointing out that our proof for $\mathbb{E}\left[N_4\right] = \Theta(n^2 \log n)$ of the conference version of this work actually gives the exact asymptotic growth of N_4, which is stated in Theorem 1.1. R.F. supported by CONACyT of Mexico grant 153984. C.H. supported by projects MEC MTM2012-30951 and Gen. Cat. DGR 2009SGR1040, 2014SGR46 and ESF EUROCORES programme EuroGIGA, CRP ComPoSe: grant EUI-EURC-2011-4306.

REFERENCES

[1] Aichholzer, O., Fabila-Monroy, R., González-Aguilar, H., Hackl, T., Heredia, M. A. Huemer, C., Urrutia, J. and Vogtenhuber, B., 4-holes in point sets, *Comput. Geom.,* **47(6)** (2014), 644–650.

[2] Aichholzer, O., Fabila-Monroy, R., Hackl, T., Huemer, C., Pilz, A. and Vogtenhuber, B., Lower bounds for the number of small convex k-holes, *Comput. Geom.,* **47(5)** (2014), 605–613.

[3] Aichholzer, O., Fabila-Monroy, R., González-Aguilar, H., Hackl, T., Heredia, M. A., Huemer, C., Urrutia, J., Valtr, P. and Vogtenhuber, B., On k-gons and k-holes in point sets, *23rd Annual Canadian Conference on Computational Geometry CCCG 2011,* 21–26. Toronto, Canada, 2011.

[4] Aichholzer, O., Hackl T. and Vogtenhuber, B., On 5-gons and 5-holes. *XIV Encuentros de Geometría Computacional ECG2011, Lecture Notes in Comput. Sci.,* vol. **7579**: 1–13. Springer, 2012.

[5] Balogh, J., González-Aguilar, H. and Salazar, G., Large convex holes in random point sets, *Comput. Geom.,* **46(6)** (2013), 725–733.

[6] Bárány, I., Sylvester's question: the probability that n points are in convex position, *The annals of probability,* **27** (1999), 2020–2034.

[7] Bárány, I. and Füredi, Z., Empty simplices in Euclidean space, *Canad. Math. Bull.,* **30(4)** (1987), 436–445.

[8] Bárány, I., Marckert, J.-F. and Reitzner, M., Many empty triangles have a common edge, *Discrete and Computational Geometry,* **50(1)** (2013), 244–252.

[9] Bárány, I. and Valtr, P., Planar point sets with a small number of empty convex polygons, *Stud. Sci. Math. Hung.,* **41** (2004), 243–266.

[10] Blaschke, W., Vorlesungen über Differentialgeometrie II. Affine Differentialge ometrie, Springer, Berlin, 1923.

[11] Dumitrescu, A., Planar sets with few empty convex polygons, *Stud. Sci. Math. Hung.,* **36** (2000), 93–109.

[12] García, A., A note on the number of empty triangles, *XIV Encuentros de Geometría Computacional ECG2011, Lecture Notes in Comput. Sci.,* vol. **7579**: 249–257. Springer, 2012.

[13] Henze, N., Random triangles in convex regions, *J. Appl. Prob.,* **20** (1983), 111–125.

[14] Katchalski, M. and Meir, A., On empty triangles determined by points in the plane, *Acta Math. Hung.,* **51** (1988), 323–328.

[15] Marckert, J.-F., Probability that n random points in a disk are in convex position, arXiv:1402.3512, 2014.

[16] Pfiefer, R. E., The Historical Development of J. J. Sylvester's Four Point Problem, *Math. Mag.,* **62** (1989), 309–317.

[17] Valtr, P., On the minimum number of empty polygons in planar point sets, *Stud. Sci. Math. Hung.,* **30** (1995), 155–163.

[18] Valtr, P., Probability that n random points are in convex position, *Discrete and Computational Geometry,* **13** (1995), 637–643.

[19] Valtr, P., The probability that n random points in a triangle are in convex position, *Combinatorica,* **16** (1996), 567–573.