The complexity of many faces in an arrangement of lines in the plane, related point-line incidence bounds, crossing numbers, probabilistic methods, and $\frac{1}{r}$-cuttings

Sudebkumar P. Pal
Computational Geometry, Spring 2023
Department of Computer Science and Engineering, IIT Kharagpur, 721302
February 26, 2023

1 The overall idea of partitioning with a random sample and using Canham's bound recursively for bounding the complexity of many faces

We consider the problem of estimating the number $K(m, n)$, the *many faces complexity* of edges of m faces in an arrangement of n lines. One way to visualize is to consider a set P of m points in the plane, and a set L of n lines in the plane. The (at most) m faces are determined by the m points in the arrangement $A(L)$ of lines in L. We get the inferior upper bound (known as the Canham bound) of $O(m\sqrt{n} + n)$ using the *forbidden subgraph* property of the *bipartite incidence graph* of lines and faces in an arrangement of lines. [The forbidden subgraph is $K_{2,5}$. Using the result by Kovari, Sos and Turan (Theorem 9.6 in [3]) for such forbidden component subgraphs, we get the above loose upper bound. See Pach and Agarwal [3], for a proof of the Kovari, Sos and Turan result.] We proceed to use a divide-and-conquer approach as follows, in order to derive a much better bound that also asymptotically matches the best known lower bounds (see Theorem 11.9 of [3]).

1.1 A divide-and-conquer approach using a grosser partition with a random sample subset of lines

Suppose we form an arrangement with a subset R of size r of the set L of n lines. The arrangement $A(L)$ is of our interest. However, we may first convert $A(R)$ into a trapezoidal map $A^*(R)$ with $k = s \leq 3r^2$ trapezoids/triangles as faces, by dropping plumbline vertical segments from vertices and intersection points of $A(R)$. It is nice if not too many lines from $L \setminus R$ intersect an arbitrary trapezoid Δ_j of $A^*(R)$, where the (fixed) point $p_j \in P$ lies in the (unique) trapezoid Δ_j. Even if this trapezoid is intersected by q_j lines, we wish to have the expectation $E(q_j) = O(\frac{n}{r})$, where the expectation is over all the $\binom{n}{r}$ random samples $R \subset L$. This is indeed possible and we show this later using combinatorial arguments in Section 1.3.2; this is a technical result of independent and deep import, which we will use crucially in Section 1.3.1.

Let the face Δ_i of $A^*(R)$ intersect n_i lines of $L \setminus R$ and contain m_i of the m points from the point set P. [Here, the set L_i of lines from $L \setminus R$ that intersect Δ_i, form an arrangement $A(L_i)$; the convex faces (cells) in $A(L_i)$ are just the faces of arrangements $A(L)$ or $A(R)$. In contrast, by the very definition of A^* $A^*(R)$, $A^*(L)$ and $A^*(L_i)$ have only trapezoids and triangles for faces (or cells).] Now, using recursion we write

$$K(m,n) \leq \sum_{i=1}^{s} K(m_i,n_i) + O(nr)$$

We will explain the $O(nr)$ term using the zone theorem and its non-trivial application in Section 1.2. for r. Using the Canham bound, can write

$$K(m,n) \leq \sum_{i=1}^{s} (m_i\sqrt{n_i} + n_i) + O(nr)$$

In Section 1.3, we use the existence of random sample R of size r to establish the upper bound

$$\sum_{i=1}^{s} m_i(n_i)^{\alpha} = O(m(\frac{n}{r})^{\alpha})$$

[This bound is established in part (ii) of Theorem 11.2 in [3]; part (i) of the same theorem claims that $\Sigma_{i=1}^{s} n_i \leq c_1 nr$, which holds for any $R \subset L$, where $|R| = r$.] So, we can write

$$K(m,n) \leq O(m(n/r)^{\frac{1}{2}}) + O(nr)$$

Now, by setting $r = min(n, \frac{m^{\frac{2}{3}}}{n^{\frac{1}{3}}})$ we get $nr = (mn)^{\frac{2}{3}}$ and therefore, $K(m,n) = O(m^{\frac{2}{3}}n^{\frac{2}{3}} + n)$.

[A quick and simple application of the zone theorem is instructive and useful now. Even if we consider an arbitrary $R \subset L$ so that $|R| = r$, the sum $\Sigma_{i=1}^{s} n_i = O(nr)$. In other words, the average number of lines of $L \setminus R$ intersecting the $s = O(r^2)$ trapezoids of $A^*(R)$ is $O(\frac{n}{r})$, a very good bound. Such a bound is certainly '*good on the average*'. This is essentially part (i) of Theorem 11.2 of [3]. In addition, if R is a random sample, then the quantity

$$\sum_{i=1}^{s} m_i(n_i)^{\alpha} = O(m(\frac{n}{r})^{\alpha})$$

and the random sample R of lines R works like a '$(\frac{1}{r}) -$ *cutting on the average*' for applications such as this very *many faces complexity* problem.]

1.2 The zone theorem and its non-trivial application in the establishing the $O(nr)$ term

See Theorem 11.7 from pages 175-176 [3].

1.3 Bounding the expectation of the (weighted) sum of αth moments over a space of random samples: Existence of a good random sample

As mentioned in Section 1, we outline the proof of Theorem 11.2 (ii) of [3]. In order to show that $\sum_{i=1}^{s} m_i(n_i)^\alpha = O(m(\frac{n}{r})^\alpha)$, for some random sample R of r lines of L, it is sufficient to show that the *expected value* or *expectation* of $\sum_{i=1}^{s} m_i(n_i)^\alpha$ over all possible (random samples) choices of R, is $O(m(\frac{n}{r})^\alpha)$.

We achieve this objective as follows, first by simplifying the expression whose expectation is being bounded in Section 1.3.1, and then using the expectations of certain simpler quantities in Section 1.3.2.

1.3.1 Simplifying the expectation of a sum over random samples

We define q_j as the number of lines of L intersecting the interior of the unique trapezoid in $A^*(R)$ that contains $p_j \in P$. We can now see that $E[\sum_{i=1}^{s} m_i n_i^\alpha] = E[\sum_{j=1}^{m} q_j^\alpha] = \sum_{j=1}^{m} E[q_j^\alpha] \leq \sum_{j=1}^{m}(E[q_j])^\alpha$. The last inequality is given as Exercise 11.5 in [3]. The property that the expectation of a sum is the sum of expectations is a very powerful and simplifying property used above so that now we may only concentrate on establishing

$$E[q_j] = O(\frac{n}{r})$$

So, we can now write $\sum_{j=1}^{m}(E[q_j])^\alpha = O(m(\frac{n}{r})^\alpha)$.

1.3.2 Bounding the expectation $E[q_j]$ over all random samples using counting and probabilistic arguments

See page 172 of [3] in the last part of the proof of Theorem 11.2 (ii) to see how an brilliant probabilistic argument is used to show that $E[q_j] = O(\frac{n}{r})$.

2 $\frac{1}{r}$-cuttings and 'cuttings on the average'

The *Cutting Lemma* in [1], Lemma 4.5.3 is a stronger version of the *Weaker Cutting Lemma*, Lemma 4.6.1. The number of lines (in the Weaker Cutting Lemma) in the random sample $S \subset L$ is $s = 6r \ln n$, whose $O(s^2)$-sized arrangement in the plane results in $t = O(s^2)$ *generalized triangles*, as opposed to only $O(r^2)$ *generalized triangles* in the (ambititious) Cutting Lemma. Clearly, the dependence on n, eventhough on a slowly growing $\ln(n)$ function, is a disadvantage of the Weaker Cutting Lemma; the advantage is a very short and elegant probabilistic argument in its proof (see page 66 of [1]). In Section 1, we have already seen such cutting-like structures, where the random sample R gives rise to $s = O(r^2)$ triangles, whose conflicts sets are of cardinality $O(\frac{n}{r})$ 'on the average' which is 'good', in contrast to the individual $O(\frac{n}{r})$ bound that we need for each triangle in a $\frac{1}{r} - cutting$.

3 Point-line incidences, many faces complexity, crossings and probabilistic methods

Pach et al. [3] (Lemma 10.6) demonstrates the close relationship between the number of maximum possible number of point-line incidences $I(m, n)$, and the many faces complexity $K(m, n)$ for a set of m cells in the arrangement of n lines, where $I(m, n) \leq \frac{1}{2}K(m, 2n) + m$. This upper bound on $I(m, n)$ along with the many faces upper bound on $K(m, n)$ of Section 1 leads to the upper bound on $I(m, n) = O(m^{\frac{2}{3}}n^{\frac{2}{3}} + m + n)$ (see Corollary 11.8 based on Lemma 10.6 and Theorem 11.7 of Pach et al. [3]). We state an alternative method for establishing this upper bound in Section 3.2, based on the lower bound for crossings in Section 3.1. Also, a matching, asymptotically tight lower bound for $I(m, n)$ in Theorem 11.9 in [3] holds for $K(m, n)$, because it is developed for $I(m, n)$ using some geometric construction that also uses elementary number theory.

3.1 A lower bound on crossing numbers: An elegant probabilistic argument

An *embedding* of a graph $G = (V, E)$ in the plane is a planar representation of it, where each vertex is represented by a point in the plane, and each edge $\{u, v\}$ is represented by a curve connecting the points corresponding to the vertices u and v. The *crossing number* of such an embedding is the number of pairs of intersecting curves that correspond to pairs of edges with no common endpoints. The crossing number $cr(G)$ of G is the minimum possible crossing number in an embedding of it in the plane. The only and trivial planar embedding of the graph K_3 has crossing number 0. Hence it is a planar graph. The complete graph K_4 of four vertices has crossing number o as well. In every planar embedding, the graph K_5 has at least one pair of edges crossing. Hence, it is a non-planar graph. $K_{3,3}$ also has crossing number 1. Hence, it is a non-planar graph. Kuratowski showed 1930 that a graph is planar if and only if it has no subgraph homeomorphic to K_5 or $K_{3,3}$. The following *Crossing Number Theorem* was proved by Ajtai, Chvatal, Newborn and Szemeredi in 1982, and independently, by Leighton. The crossing number of any simple graph (i.e., a graph with no multi-edges or no self-loops) with $|E| \geq 4|V|$ is at least $|E|^3/64|V|^2$. Let us describe a short probabilistic proof of this theorem.

We know Eulers formula for any spherical polyhedron, with $|V|$ vertices, $|E|$ edges and $|F|$ faces, $|V| - |E| + |F| = 2$. Any maximal planar graph (i.e., one to which no edge can be added without losing planarity) has triangular $|F|$ triangular faces implying $3|F| = 2|E|$. Hence, for any simple planar graph with $|V| = n \geq 3$ vertices, we have $|E| = |V| + |F| - 2 \leq |V| + (2/3)|E| - 2$ or $|E| \leq 3n - 6$, implying that it has at most $3n$ edges. Therefore, the crossing number of any simple graph with n vertices and m edges is at least $m - 3n$. Let $G = (V, E)$ be a graph with $|E| \geq 4|V|$ embedded in the plane with $t = cr(G)$ crossings. Let H be the random induced subgraph of G obtained by picking each vertex of G, randomly and independently, to be a vertex of H with probability p (whose value is to be chosen later). Then, the expected number of vertices in H is $p|V|$, the expected number of edges is $p^2|E|$, and the expected number of crossings (in its given embedding) is p^4t. Therefore, we have $p^4t \geq p^2|E| - 3p|V|$, implying

$t \geq |E|/p^2 - 3|V|/p^3$. Substituting $p = 4|V|/|E|$, which is less than one, we get the result.

3.2 The Szemeredi-Trotter point-line incidence upper bound and crossing numbers

Now we state the famous Szemeredi-Trotter Theorem. Let P be a set of n distinct points in the plane, and let L be a set of m distinct lines. Then the number of incidences I between the members of P and those of L (i.e., the number of pairs (p, l) with $p \in P$, $l \in L$, $p \in L$) is at most $c(m^{\frac{2}{3}}n^{\frac{2}{3}} + m + n)$, for some absolute positive constant c. We state a proof using probabilistic arguments. This proof is due to Szekely (1997).

We assume that every line in L is incident with at least one of the points of P and every point is on some line. Denote the number of such incidences by I. Form a graph $G = (V, E)$ with $V = P$, where for $p, q \in P, (p, q) \in E$ if and only if they are consecutive points of P on some line in L.

Clearly, $|V| = n$, and $|E| = \Sigma_{j=1}^{m}(k_j - 1) = (\Sigma_{j=1}^{m}k_j) - m = I - m$, where k_j is the number of points of P on line $j \in L$. Note that G is already embedded in the plane where the edges are represented by segments of the corresponding lines in L. In this embedding, every crossing is an intersection point of two members of L. So, $cr(G) \leq \binom{m}{2} \leq \frac{m^2}{2}$. By the Crossing Number Theorem, either $I - m = |E| < 4|V| = 4n$, that is,

$$I \leq m + 4n$$

or

$$\frac{m^2}{2} \geq cr(G) \geq \frac{(I-m)^3}{64n^2}$$

implying

$$I \leq (32)^{\frac{1}{3}} m^{\frac{2}{3}} n^{\frac{2}{3}} + m$$

In both cases,

$$I \leq 4(m^{\frac{2}{3}} n^{\frac{2}{3}} + m + n)$$

4 Random sampling for geometric/searching applications

The following discussion is based on Chapter 5 (pages 173-175) of [2]. We have n objects in the set N, and subsets of N can be the *defining elements* of *configurations*. Let $\Pi = \Pi(N)$ be the set of all configurations and $\sigma \in \Pi$ be one such configuration. For an example, imagine a one-dimensional space (the real line) with n distinct points and the $O(n^2)$ pairs of n points as configurations, which are actually linear intervals. If we fix a constant $r < n$, we may take a *random sample* R of r elements, selected out of the n elements in N. [Sampling is done without repetitions; each time an element is selected independently and randomly.] We can see that the set R immediately defines the set $\Pi(R)$ of $\binom{r}{2} = O(r^2)$ configurations or intervals. We say that the set $D(\sigma)$ of cardinality $d(\sigma)$, is the set of *triggers* or *defining elements* of any configuration $\sigma \in \Pi(R)$; $d(\sigma) \leq 2$ in this example of intervals on a line because each interval has at most two endpoints. Each configuration σ may contain (intersect) a set of elements from N. These are called *stoppers*. The set of stoppers is denoted by $L(\sigma)$ and its cardinality is denoted by $l(\sigma)$; $l(\sigma)$ is called the *conflict size* of the configuration $\sigma \in \Pi(R)$. We wish to bound the probability that we get a configuration with a large number of stoppers but no triggers (except for the two extreme defining triggers of the configuration). Such configurations are called *active configurations* of the random sample R. In other words, we say that a configuration $\sigma \in \Pi(N)$ is active over a subset (random sample) $R \subset N$ if it occurs as an interval in $\Pi(R)$. We show that probability that each active configuration of a random sample R of cardinality r would have conflict size $O(\frac{n}{r} \log r)$, is at least $\frac{1}{2}$.

The probability that σ has no point of R in conflict, given that its defining points are in R is

$$p(\sigma, r) \leq (1 - \frac{l(\sigma)}{n})^{r-d(\sigma)} \tag{1}$$

The intuitive justification is as follows. The interval being of conflict size $l(\sigma)$, the probability of choosing a conflicting point is at least $\frac{l(\sigma)}{n}$. Since we select $r - d(\sigma)$ points without conflicts, the probability required is upper bounded as in Inequality 1. [For a rigorous derivation, see page 175 in [2] or a simlar derivation in [4].] However, since $1-x \leq exp(-x)$ where $exp(x) = e^x$, we have

$$p(\sigma, r) \leq exp(-\frac{l(\sigma)}{n}(r - d(\sigma))) \tag{2}$$

Since $d(\sigma) \leq 2$, putting $l(\sigma) \geq c(n \ln s)/(r-2)$ for some $c > 1$ and $s \geq r$, we get

$$p(\sigma, r) \leq exp(-c \ln s) = \frac{1}{s^c} \tag{3}$$

Now an active configuration σ due to the random sample R must be such that all its defining points must be in R. In other words, $\sigma \in \Pi(R)$. Let this probability be $q(\sigma, r)$. The probability that σ is (i) an active configuration due to random sample R, with (ii) conflict set at least as big as $\frac{cn \ln s}{r-2}$, is therefore no more than

$$p(\sigma, r) q(\sigma, r)$$

The probability that some active configuration has such a "long" conflict set is no more than the sum of probabilities for all such "long" configurations $\sigma \in \Pi(R)$

$$\sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} p(\sigma, r) q(\sigma, r) \leq \sum_{\sigma \in \Pi : l(\sigma) > \frac{cn \ln s}{r-2}} q(\sigma, r)/s^c \leq \frac{1}{s^c} \sum_{\sigma \in \Pi} q(\sigma, r) \tag{4}$$

Now the last summation in the Inequality 4 is the expectation $E(\pi(R))$ and $\pi(R) = |\Pi(R)| = O(r^2)$. So, choosing $c > 2$ we can ensure that the probability of having a long active configuration in $\sigma \in \Pi(R)$ is less than $\frac{1}{2}$ for a random sample R.

For a reading exercise, study Section 5.1, pages 176-180, from [2] where generalizations to two and higher dimensions are considered. The configurations spaces considered have *bounded valence*. It is shown that a result similar to that derived above holds for such spaces (see Theorem 5.1.2 [2]). Similar results are also derived in [4] in Section 9.9 (Random Sampling), Lemma 9.11 (probability of getting a good cutting), Corollary 9.12 (Las Vegas algorithm for the partition/cutting), and Theorem 9.13 (a point-location data structure with query and preproscessing time complexities). See also the simpler Exercise 9.19 and 9.20 in [4].

5 The weak cutting lemma: Demonstration of a $\frac{1}{r}$-cutting using a probabilistic argument

With n lines in the plane, we can have at most $O(n^2)$ cells, and a lower bound of at least $\frac{n^2}{2}$ cells. Let L be this set of n lines. Any triangle that cuts k lines is divided into at most $2k^2$ parts. Suppose we have only t triangles partitioning the whole plane containing the arrangement of n lines, and each such triangle is cut by at most k of the n lines. Since each of these at least $\frac{n^2}{2}$ cells has be be covered by triangles, each of which has at most $2k^2$ cells as stated above, we need to have $\frac{n^2}{4k^2} = \Omega(r^2)$ cells provided we fix $k \leq \frac{n}{r}$.

We show that a set of $O(r^2 \log^2 n)$ triangles can be used to ensure that less than $\frac{n}{r}$ lines of the arrangement of n lines cross each such triangle. Let L be the set of n lines. Use a random sample $S \in L$, of size $s = 6r \log n$ to create $O(s^2)$ regions as follows. If there are non-triangular (convex) regions in the arrangement of these s lines, then we triangulate them. A *bad* triangle T (defined by three lines of the n lines in L) has at least $k = \frac{n}{r}$ lines intersecting T. Such a bad triangle is also called *interesting* if it appears in the triangulation of $S \subseteq L$ as one of the $O(s^2)$ triangles of the random sample S of size s as mentioned above.

There are at most n^6 interesting triangles; each triangle has three of the $\binom{n}{2}$ points of intersections as vertices. The probability that T is a bad triangle is less than n^{-6} if we choose $s = 6r \log n$. The upper bound on this probability is

$$(1 - \frac{k}{n})^s = (1 - \frac{1}{r})^{6r \ln n} \leq e^{-6 \ln n} = n^{-6}$$

So, the probability that some interesting triangle is bad is **strictly less** than unity. Therefore, there exists is a random sample S of size $s = 6r \log n$ such that the none of the $O(s^2)$ triangles induced by the triangulation of the arrangement of S meets more than $\frac{n}{r}$ lines of L. This is the *Weak Cutting Lemma* (Lemma 4.6.1) from [1].

References

[1] Jiri Matousek, *Lectures on Discrete Geometry*, Springer.

[2] Ketan Mulmuley, Computational Geometry: An Introduction Through Randomized Algorithms, Prentice Hall, 1994.

[3] Janos Pach and Pankaj Agarwal, *Combinatorial Geometry*, Wiley-Interscience Series in Discrete Mathematics and Optimization, 1995.

[4] R. Motwani and P. Raghavan, Randomized algorithms, Cambridge University Press, 1995.