Szemerédi-Trotter: Applications and Limitations

Maxwell Fishelson

Abstract

In this paper, we investigate the Szemerédi-Trotter theorem and its various applications in combinatorics. We prove that given a set of points P in plane, the number of triangles with unit perimeter having vertices in P is at most $c \cdot |P|^{7/3}$, for a positive constant c, independent of |P|. Furthermore, we look at the original statement of the Szemerédi-Trotter theorem and investigate some of the recent theorems that generalize the result (Szemerédi-Trotter for general curves, and in higher dimensions). In addition, we investigate Szemerédi-Trotter's applications in algebra, considering points incident to polynomial curves, and prove a general result that bounds the sizes of sets determined by polynomials. Last but not least, we discuss the limitations of Szemerédi-Trotter and compare it to other techniques.

1 Preliminaries

Definition 1 We let O signify an asymptotic upper bound. Say f(n) and g(n) are two functions with codomain \mathbb{R} . To assert that f(n) = O(g(n)) is to assert that there exists a positive constant c such that, for all n, $f(n) < c \cdot g(n)$.

Definition 2 Similarly, we let Ω signify an asymptotic lower bound. To assert that $f(n) = \Omega(g(n))$ is to assert that there exists a positive constant c such that, for all $n, f(n) > c \cdot g(n)$.

Definition 3 For a point p and a curve c, we say p is "incident" to c if p lies on c. For a set of points P and a set of curves C, we define the incidence graph of $P \times C$ to be the following bipartite graph: a vertex for each point in P on one side, a vertex for each curve in C on the other, and an edge between a point and a curve if that point is incident to that curve.

2 The Unit Perimeter Triangles Problem

We introduce the type of theorem that can be proven with the help of Szemerédi-Trotter with the following original result:

Let $P = \{P_1, P_2, \dots, P_n\}$ be a set of n points that we place in the real plane. We situate the points in such a way that we maximize the number of unordered triples of distinct points $\{P_i, P_j, P_k\}$ where triangle $P_i P_j P_k$ has unit perimeter (perimeter of 1). Let us call this maximal number of unit perimeter triangles T_n . For all integers i from 1 to n, we define $T_{i,n}$ to be the number of unit perimeter triangles with P_i as one of the three vertices.

Theorem 1.
$$T_n = O(n^{7/3})$$

Proof. We claim that $T_{1,n} = O(n^{4/3}) \Rightarrow T_n = O(n^{7/3})$. If we show $T_{1,n} = O(n^{4/3})$, we will have $T_{i,n} = O(n^{4/3})$ for all i by symmetry. Therefore, $\sum_{i=1}^n T_{i,n} = O(n^{7/3})$

as it is a sum of n terms each individually upper bounded by $O(n^{4/3})$. Note that $3 \cdot T_n = \sum_{i=1}^n T_{i,n}$ since each unit perimeter triangle is counted exactly 3 times on both sides of the equation. Since constants do not affect asymptotic bounds, $\frac{\sum_{i=1}^n T_{i,n}}{3} = T_n = O(n^{7/3})$. So it is sufficient to show $T_{1,n} = O(n^{4/3})$.

We construct a series of ellipses in the following manner. For all \mathbb{Z} i from 2 to n, if the Euclidian distance between P_1 and P_i is less than $\frac{1}{2}$ (say $d(P_1, P_i) = x$), we draw the locus of points p such that P_1 , P_i , and p form a unit perimeter triangle. This will be the set of points p such that $d(p, P_1) + d(p, P_i) = 1 - x$. It is well known that this locus of points is an ellipse, since 1 - x > x. Assume there are exactly m elements of $P \setminus \{P_1\}$ within a distance of $\frac{1}{2}$ from P_1 . This means that this process will form m ellipses: E_1, E_2, \dots, E_m with m < n. For each ellipse E_k , we let e_k be the number of points in $P \setminus \{P_1\}$ that are incident to E_k . We note that $2 \cdot T_{1,n} = 1$ the total number of incidences between our ellipses and points in $P \setminus \{P_1\}$, since each incidence marks a completed unit perimeter triangle. We see that each triangle is counted twice, e.g. if P_1, P_2, P_3 forms a unit perimeter triangle, P_3 would be incident to the ellipse formed with P_2 and vice versa, which gives us the factor of 2, but this factor of 2 is irrelevant, as we are working asymptotically. Therefore, showing that the total number of incidences $= \sum_{k=1}^m e_k = O(n^{4/3})$ is sufficient.

Consider an unordered pair of two distinct elements of P: (P_i, P_j) and one of our ellipses E_k . We say (P_i, P_j) is a "segment" of E_k if P_i and P_j meet the following two conditions. Firstly, P_i and P_j must both be incident to E_k . Secondly, either the path starting at P_i , going clockwise around the perimeter of E_k , and ending at P_j goes through no other points in P, or the path starting at P_j , going clockwise around the perimeter of E_k , and ending at P_i goes through no other points in P. If the clockwise path starts at P_i and ends at P_j , we say that P_i is the "start" of the segment, and vice versa (note that both P_i and P_j can be the start, if $e_k = 2$). If $e_k > 2$, we claim that each segment has a unique start. Assume for the sake of contradiction that (P_i, P_j) is a segment of E_k with start P_i and P_j and with $e_k > 2$. Therefore, there are no points in P on the clockwise path from P_i to P_j or the clockwise path from P_j to P_i , which is a contradiction since this forces P_i , P_j to be the only elements of P incident to E_k ,

meaning $e_k = 2$. If $e_k > 1$, for all $p \in P$ incident to E_k , we claim that p is the start of exactly one segment. Starting at p, travel clockwise around the ellipse until you first encounter another element of P, call it q. (p,q) is a segment, by definition, and for any other $r \in P$ incident to E_k , the path from p clockwise to r is not a segment, since this path must contain q. Therefore, for an ellipse E_k , if $e_k > 2$, E_k has exactly e_k segments, since each $p \in P$ incident to E_k is the start of exactly one segment, and each segment has exactly one start. However, if $e_k = 0, 1$, there will be no segments, and if $e_k = 2$, there will be only 1 distinct segment (even though this segment will have 2 starts). This means, in general, that if there are x points in P incident to an ellipse, that ellipse has at least x - 1 segments. Consequently, $\sum_{k=1}^{m} e_k \leq \sum_{i=1}^{m} (the number of segments of <math>E_i + 1$) $\leq \sum_{i=1}^{m} (the number of segments of <math>E_i) + n$ since m < n. Therefore, showing $\sum_{i=1}^{m} (the number of segments of <math>E_i) = O(n^{4/3})$ is sufficient.

We now map our diagram of points and ellipses to graph theory. We consider a graph G where the n points in P are the vertices. Let the set of edges of E be G. We say two distinct vertices P_i, P_j have an edge between them if there exists an ellipse E_k such that (P_i, P_j) is a segment of E_k . For the most part, one segment on one ellipse adds one edge to G. However, we have to consider when two points, say P_a, P_b , form a segment on multiple different ellipses, despite the fact that the edge is only added to the graph once. Nevertheless, we claim that the points P_a, P_b can only both be incident to at most 4 ellipses, meaning (P_i, P_j) can only be a segment on at most 4 ellipses. Assume for the sake of contradiction that there exists a set S of 5 points in P (none of which are P_1 , P_a , or P_b) such that, for all $p \in S$, $d(P_1,p)<\frac{1}{2}$ (meaning there was an ellipse formed with foci P_1 and p) and P_a,P_b lie on that ellipse. If this were true, triangles P_1P_ap and P_1P_bp would have unit perimeters. Thus, $A = d(P_1, P_a) < \frac{1}{2}$ and $B = d(P_1, P_b) < \frac{1}{2}$ since the longest edge of a triangle is less than half its perimeter by the triangle inequality. Therefore, $p \in$ the set of points q such that $d(q, P_1) + d(q, P_a) = 1 - A$ and $p \in \text{the set of points } q$ such that $d(q, P_1) + d(q, P_b) = 1 - B$. Since these two sets are distinct ellipses (since they have different foci), and |S| = 5, we must have 5 distinct points incident to 2 distinct ellipses, which is a contradiction since two distinct ellipses can intersect in at most 4 points. Consequently, each element of E is formed from at most 4 segments on distinct ellipses. Thus, $\sum_{i=1}^{m}$ (the number of segments of E_i) $\leq 4|E|$. Therefore, showing $|E| = O(n^{4/3})$ is sufficient.

Given that (P_i, P_j) is a segment of E_k , with start (P_i) , we define the "curve" of a segment to be the continuous path of points starting at P_i , going clockwise around E_k , and ending at P_j . If $e_k > 2$, we can determine the curve of segment (P_i, P_j) without being given the start, since the segment has a unique start. We iterate through our list of ellipses E_1, E_2, \dots, E_m from E_1 to E_m , and for each E_k , we create a set F_k of curves of segments of E_k as follows. If $e_k = 0, 1$, we set $F_k = \emptyset$. If $e_k > 2$, we start by setting F_k to be the set of the curves of every segment of E_k , but then we remove all the curves of segments (P_i, P_j) if (P_i, P_j) was also a segment on E_j for some j < k. If $e_k = 2$ (say P_i, P_j incident to E_k with i < j), we set $F_k = \{$ the curve of (P_i, P_j) with start P_i , unless (P_i, P_j) was also a segment on E_j for some j < k, in which case $F_k = \emptyset$. Finally, define $F = \bigcup_{k=1}^m F_k$. Note that, in the way we constructed F, if P_i and P_j are connected in G, there is exactly 1 curve in F between P_i and P_j , and if P_i and P_j are not connected in G, there are no curves in F between them. Also note that for each P_i and P_j connected in G, the curve between them in F is a valid way to draw the edge between them in G, since each curve is continuous and passes through no other vertices. Therefore, setting the curves in F to the edges in G is one valid way to draw G's edges. Let's call this specific edge drawing of G, G^* . Let $cr(G^*)$ be the number of edge crossings in G^* , and let cr(G) (aka the crossing number of G) be the minimum number of edge crossings over all the ways to draw the edges of G. Since we may not have an optimally low number of edge crossings in G^* , we say $cr(G) \leq cr(G^*)$. $cr(G^*)$ will be the number of pairs of curves in F that cross (intersect at a point not in P). Two curves that come from the same ellipse cannot cross, since a point $p \notin P$ on an ellipse lies on only one path between consecutive points in P incident to the ellipse and therefore cannot belong to two distinct curves. Let f_k be the set of all points on all the curves in F_k . Since we know curves from the same ellipse do not cross, we can just consider intersections between different f_k 's. Thus, $cr(G^*) = \sum_{1 \le j \le k \le m} |(f_j \cap f_k) \setminus P|$. Note $(f_j \cap f_k) \setminus P \subset E_j \cap E_k$ since $f_j \subset E_j$ and $f_k \subset E_k$. Also note $|E_j \cap E_k| \leq 4$ since two distinct ellipses have at most 4 common points. Therefore, $cr(G^*) = \sum_{1 \le j < k \le m} 4 < 4m^2 < 4n^2$.

We finish by using a well known graph theory inequality:

(Ajtai, Chvátal, Newborn, Szemerédi [4]) Let G be a simple graph with e edges and n vertices. If $e \ge 4n$, then $cr(G) \ge \frac{e^3}{64n^2}$

We have n vertices and |E| edges. If |E| < 4n, clearly we have $|E| = O(n^{4/3})$. So, assuming $|E| \ge 4n$, we get $4n^2 > cr(G^*) \ge cr(G) \ge \frac{|E|^3}{64n^2}$, and we obtain the desired $|E| = O(n^{4/3})$.

3 The Szemerédi-Trotter Theorem and Its Extensions

The Szemerédi-Trotter Theorem states:

(Szemerédi, Trotter [9]) Let P be a set of m points and let L be a set of n lines, both in \mathbb{R}^2 . Then, the number of incidences between points in P and lines in $L = I(P, L) = O(m^{2/3}n^{2/3} + m + n)$

The original proof provided by Szemerédi and Trotter was complex. It used a technique known as cell decomposition and the arguments used were difficult to generalize. Later, Székely came up with a simple combinatorial proof of Szemerédi-Trotter [1]. His argument converted the points, lines, and incidences into a graph theory diagram, and then used the same crossing number inequality [4] that I did. It is important to note that many of the arguments I used in the Unit Perimeter Triangle proof are inspired by the arguments used by Székely in this paper.

With the help of Székely's simpler proof, Pach and Sharir were able to generalize

his arguments using a multigraph. They derived a theorem that bounded incidences between points and algebraic curves in general, not just incidences between points and lines:

(Pach, Sharir [2]) Let P be a set of m points and let C be a set of n constant-degree algebraic curves, both in \mathbb{R}^2 , such that the incidence graph of $P \times C$ does not contain a copy of $K_{s,t}$. Then, $I(P,C) = O(m^{\frac{s}{2s-1}}n^{\frac{2s-2}{2s-1}} + m + n)$.

Recently, the Szemerédi-Trotter Theorem has had its biggest generalization since Pach and Sharir: extending it to higher dimensions. A new powerful tool, known as polynomial partitioning (Guth, Katz [8]), was the key. Polynomial partitioning and the cell decomposition technique used in the original Szemerédi-Trotter proof both serve to split the space into a number of cells, but in polynomial partitioning the cells are determined algebraically by roots of a polynomial as opposed to the combinatorial way to divide cells used by Szemerédi and Trotter. Using polynomial partitioning and the aforementioned result of Pach and Sharir, Zahl proved an extension of Szemerédi-Trotter bounding incidences between points and curves in 3-space:

(Zahl [3]) Let P be a set of m points and let V be a set of n smooth constant-degree algebraic varieties, both in \mathbb{R}^3 , such that the incidence graph of $P \times V$ does not contain a copy of $K_{s,t}$. Then, $I(P,V) = O(m^{\frac{2s}{3s-1}}n^{\frac{3s-3}{3s-1}} + m + n)$.

4 Using Szemerédi-Trotter with Polynomials

Let A and B be sets of real numbers. We define the set $A + B = \{a + b : a \in A, b \in B\}$. We define the set $A \cdot B = \{ab : a \in A, b \in B\}$. Let f(x) be a polynomial. We define the set $f(A) = \{f(a) : a \in A\}$.

Say we want to lower bound |A + A| in terms of |A|. Unfortunately, the tightest bound we have is $|A + A| = \Omega(|A|)$, which comes from the trivial fact that |A + A| > |A| since $2a \in A + A$ for every $a \in A$. To show that this is the strongest

bound we have, we look at the case when the terms of |A| form an arithmetic sequence (say $A = \{x, 2x, \dots, nx\}$). For all A of this form, we have |A + A| = 2|A| - 1. However, we can get non-trivial bounds when we try to minimize the number of elements in 2 different sets at the same time, such as max(|A+A|, |A*A|). A famous unproven conjecture of Erdős and Szemerédi states:

(Erdős, Szemerédi [6]) For every set of reals A, $max(|A+A|, |A*A|) = \Omega(|A|^{2-\epsilon})$ for every positive ϵ arbitrarily close to 0.

Elekes [10] was able to prove the bound $max(|A+A|, |A*A|) = \Omega(|A|^{5/4})$ using an argument involving Szemerédi-Trotter. The following result also serves to bound two sets at the same time, except we will be considering sets constructed with polynomials. We use a Szemerédi-Trotter argument similar to that of Elekes.

Theorem 2. Let A be a set of real numbers and let f(x), g(x) be polynomials with $d = deg(f) \le deg(g)$. Then, $max(|A + A|, |f(A) + g(A)|) = \Omega(|A|^{\frac{2d+1}{2d}})$.

Proof. Let P be the set of points in \mathbb{R}^2 : $\{(x,y): x\in A+A, y\in f(A)+g(A)\}$. Therefore, $|P|=|A+A|\cdot|f(A)+g(A)|$. Let C be the set of polynomial curves in \mathbb{R}^2 : $\{y=f(x-u)+g(v): u,v\in A\}$. $|C|=O(|A|^2)$ since there are $|A|^2$ choices for u,v. We may have duplicate polynomials in C if g(v) takes on the same value for different v in A. However, for any real r, there are at most deg(g) reals v such that g(v)=r. Therefore, each polynomial in $\{y=f(x-u)+g(v): u,v\in A\}$ is duplicated at most deg(g) times over the $|A|^2$ choices for u,v. Thus, $|C|\geq \frac{|A|^2}{deg(g)}\Rightarrow |C|=\Omega(|A|^2)$.

We want to bound the number of incidences I(P,C) between points in P and curves in C. For each curve y = f(x - u) + g(v) in C, the point (k + u, f(k) + g(v)) is incident to it for every real k. For every $k \in A$, the point $(k + u, f(k) + g(v)) \in P$. Therefore, each curve in C must have at least |A| points in P incident to it: $(k + u, f(k) + g(v)) \in y = f(x - u) + g(v)$ for each of the |A| ways to pick k. Consequently, $I(P,C) > |A| \cdot |C| = \Omega(|A|^3)$.

We finish by using the generalized version of Szemerédi-Trotter for curves:

(Pach, Sharir [2]) Let P be a set of m points and let C be a set of n constant-degree algebraic curves, both in \mathbb{R}^2 , such that the incidence graph of $P \times C$ does not contain a copy of $K_{s,t}$. Then, $I(P,C) = O(m^{\frac{s}{2s-1}}n^{\frac{2s-2}{2s-1}} + m + n)$.

All of the polynomials in C have degree d and the same leading coefficient. Therefore, any two distinct curves in C intersect in at most d-1 points. Thus, the incidence graph of $P \times C$ does not contain a copy of $K_{d,2}$. Consequently, $I(P,C) = O(|P|^{\frac{d}{2d-1}} \cdot |C|^{\frac{2s-2}{2s-1}} + |P| + |C|) = O(|P|^{\frac{d}{2d-1}} \cdot |A|^{\frac{4s-4}{2s-1}} + |P| + |A|^2)$. Since I(P,C) is asymptotically lower bounded by $|A|^3$ and asymptotically upper bounded by $|P|^{\frac{d}{2d-1}} \cdot |A|^{\frac{4s-4}{2s-1}} + |P| + |A|^2 = \Omega(|A|^3)$. Therefore, $|P| = \Omega(|A|^{\frac{2d+1}{d}})$ since one of the terms in the sum $|P|^{\frac{d}{2d-1}} \cdot |A|^{\frac{4s-4}{2s-1}} + |P| + |A|^2$ must be asymptotically greater than $|A|^3$. And since $|P| = |A+A| \cdot |f(A)+g(A)| = \Omega(|A|^{\frac{2d+1}{d}})$, we must have $\max(|A+A|,|f(A)+g(A)|) = \Omega(\sqrt{|A|^{\frac{2d+1}{d}}}) = \Omega(|A|^{\frac{2d+1}{2d}})$.

□ How-

Note: The bound $\Omega(|A|^{\frac{2d+1}{2d}})$ seems relatively weak, as it is close to $\Omega(|A|)$. However, we realize that d is solely determined by the minimum degree between our two polynomials, meaning we can get decent bounds from this generalization as long as one of the two polynomials has relatively small degree.

5 Limitations

Szemerédi-Trotter usually will not give the strongest bounds possible. For example, we look at the famous "Unit Area Triangle" problem. The statement of the problem is identical to that of the "Unit Perimeter Triangle" problem from earlier in this paper, except we are looking for triangles with area of 1, not perimeter of 1. For two points, P, Q, the locus of points R in plane such that triangle PQR has area 1 is two parallel lines, instead of the ellipses we get with unit perimeter trian-

gles. Considering these parallel lines to be our curves, a similar Szemerédi-Trotter argument can be used to get the same $O(n^{7/3})$ bound on the number of unit area triangles. However, this bound is not tight. A recent paper published by Raz, Sharir [7] improves this bound to $O(n^{20/9})$.

Despite often not giving exact bounds, bounds derived from Szemerédi-Trotter are stronger than those derived from the theorems that proceeded it. Prior to Szemerédi-Trotter, the best known way to derive upper bounds in combinatorics problems like these was the following:

(Kővári, Sós, Turán [5]) For a simple graph G with n vertices containing no $K_{a,b}$ subgraph, the number of edges of $G = e(n, K_{a,b}) \le \frac{1}{2} \sqrt[a]{b-1} \cdot n^{2-1/a} + \frac{a-1}{2}n$

We contrast the strength of these two theorems by using both of them to get two different bounds on the same well-known problem. We place n circles $C_1, C_2, \dots C_n$ in a plane, freely choosing the locations of their centers and radii, with the condition that no three circles are all tangent to each other at the same point. Our goal is to maximize the number of unordered pairs of circles $\{C_i, C_j\}$ with C_i externally tangent to C_j . We will see that Szemerédi-Trotter gives a stronger upper bound on this maximal number of pairs of tangent circles than Kővári-Sós-Turán gives.

We look at the graph with n vertices $v_1, v_2, \dots v_n$ where v_i and v_j have an edge between them if C_i is externally tangent to C_j . A well-known folklore result of Apollonius states that we cannot have three circles A, B, C each externally tangent to three other circles P, Q, R without having A, B, C all internally tangent to each other at the same point. We asserted, however, that no three of our n circles are all tangent to each other at the same point. We conclude that our graph contains no $K_{3,3}$. Using Kővári-Sós-Turán, we get that the number of pairs of externally tangent circles = the number of edges in our graph = $O(n^{5/3})$.

If we impose an \mathbb{R}^2 coordinate system on the plane containing our n circles, we can represent each circle C_i as the ordered triple (x_i, y_i, r_i) where (x_i, y_i) are the coordinate system on the plane containing our n circles, we can

nates of the center of C_i and r_i is its radius. Two circles are externally tangent if and only if the distance between their centers is equal to the sum of their radii. So, a circle (x,y,r) is externally tangent to C_i if (x,y,r) is a solution to $(x-x_i)^2+(y-y_i)^2=(r+r_i)^2$. Now, in \mathbb{R}^3 , consider the set of n points $P: \{P_1, P_2, \cdots P_n\}$ where P_i has coordinates (x_i, y_i, r_i) . Also consider the set of n smooth constant-degree algebraic curves $V: \{V_1, V_2, \cdots V_n\}$ where V_i is $(x-x_i)^2+(y-y_i)^2-(z+r_i)^2=0$. The number of incidences between these points and these curves will be exactly twice the number of pairs of externally tangent circles: if C_i is externally tangent to C_j , P_i is incident to V_j and P_j is incident to V_i . This factor of two changes nothing asymptotically, so we only need to bound the number of incidences I(P, V). If we again look at the result of Apollonius, we see that the incidence graph of $P \times V$ cannot contain a copy of $K_{3,3}$. The extension of Szemerédi-Trotter to 3 dimensions states:

(Zahl [3]) Let P be a set of m points and let V be a set of n smooth constant-degree algebraic varieties, both in \mathbb{R}^3 , such that the incidence graph of $P \times V$ does not contain a copy of $K_{s,t}$. Then, $I(P,V) = O(m^{\frac{2s}{3s-1}}n^{\frac{3s-3}{3s-1}} + m + n)$.

Plugging in m = n, s = t = 3 gives $I(P, V) = O(n^{3/2})$.

6 References

- [1] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry, Combinat. Probab. Comput. 6 (1997), 353–358.
- [2] J. Pach and M. Sharir, On the number of incidences between points and curves, Combinat. Probab. Comput. 7 (1998), 121–127.
- [3] J. Zahl, An improved bound on the number of point-surface incidences in three dimensions, *Contrib. Discrete Math.* 8 (2013), 100–121.
- [4] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi, Crossing-free subgraphs, *Annals of Discrete Mathematics* **12** (1982), 9–12.
- [5] T. Kővári, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, Colloq. Math. 3 (1954), 50–57.
- [6] P. Erdős and E. Szemerédi, On sums and products of integers, *Studies in Pure Mathematics* **2** (1983), 213–218.
- [7] O. Raz and M. Sharir, The number of unit-area triangles in the plane: Theme and variations, arXiv:1501.00379 (2015).
- [8] L. Guth and N. H. Katz, On the Erdős distinct distances problem in the plane, arXiv:1011.4105 (2010).
- [9] E. Szemerdi and W. Trotter, Extremal problems in discrete geometry, *Combinatorica* **3** (1983), 381–392.
- [10] G. Elekes, On the number of sums and products, *Acta Arithmetica* **LXXXI.4** (1997), 365–367.