On a question of Gowers concerning isosceles right-angle triangles

Gábor N. Sárközy, Stanley Selkow
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
gsarkozy@cs.wpi.edu, sms@cs.wpi.edu

February 23, 2003

Abstract

We give a simple quantitative proof that for every natural number $p \geq 3$ and real number $\delta > 0$, there is a natural number $N_0 = N_0(p, \delta)$ such that for $N \geq N_0$, every set of at least δN^2 points of $[N]^2$ contains a set of p points that determine at least $p - \lceil \log_2 p \rceil$ isosceles right-angle triangles; i.e. triples in the form

$$\{(a,b),(a+d,b),(a,b+d)\}.$$

1 Introduction

1.1 Notation and definitions

For basic graph concepts see the monograph of Bollobás [2]. $V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G. (A,B) or (A,B,E) denote a bipartite graph $G = (V,E)$, where $V = A \cup B$, and $E \subset A \times B$. In general, given any graph G and two disjoint subsets A, B of $V(G)$, the pair (A,B) is the graph restricted to $A \times B$. $N(v)$ is the set of neighbors of $v \in V$. Hence the size of $N(v)$ is $|N(v)| = deg(v) = deg_G(v)$, the degree of v. For a vertex $v \in V$ and set $U \subset V - \{v\}$, we write $deg(v,U)$ for the number of edges from v to U. We denote by $e(A,B)$ the number of edges of G with one endpoint in A and the other in B. For non-empty A and B,

$$d(A,B) = \frac{e(A,B)}{|A||B|}$$

is the *density* of the graph between A and B.

Definition 1. *The pair* (A,B) *is* ε-**regular** *if*

$$X \subset A,\ Y \subset B,\ |X| > \varepsilon|A|,\ |Y| > \varepsilon|B|$$

imply

$$|d(X,Y) - d(A,B)| < \varepsilon,$$

otherwise it is ε-irregular.

A triangle of the form $\{(a,b),(a+d,b),(a,b+d)\}$, where a,b and d are integers, is called *perfect*; i.e. a perfect triangle is an isosceles right-angle triangle whose sides are parallel to the axes.

1.2 Perfect triangles

Using Szemerédi's celebrated theorem [12] about the existence of long arithmetic progressions in dense sets of integers, Ajtai and Szemerédi [1] proved the following result: For sufficiently large N, every subset of $[N]^2$ of size at least δN^2 contains a perfect triangle (here $[N] = \{1,2,\ldots,N\}$). Later Fürstenberg and Katznelson [3] proved a much more general theorem, but their proof does not give an explicit bound on N as it uses ergodic theory. After giving a new analytic proof for Szemerédi's Theorem in [4], Gowers posed the question of finding better quantitative proofs for the Ajtai-Szemerédi Theorem (see the last section of [4] or Problem 1 in [5]). Very recently Solymosi [10] gave an elegant, quantitative proof for the Ajtai-Szemerédi Theorem using the Ruzsa-Szemerédi Theorem [7], which in turned used Szemerédi's Regularity Lemma [13]. Then Vu [14] improved on this result by using Gowers' quantitative version of the Szemerédi's Theorem and he showed that for sufficiently large N, every subset of $[N]^2$ of size at least $\frac{100N^2}{(\log^* N)^{1/4}}$ contains a perfect triangle. Let us also mention that recently Solymosi [11] showed that the conditions of the Ajtai-Szemerédi Theorem guarantee not just a perfect triangle, but a square as well.

In this paper we strengthen the Ajtai-Szemerédi Theorem in a different direction, we show that there is a set of points which contain many perfect triangles.

More precisely, our main result is the following.

Theorem 1. *For every natural number $p \geq 3$ and real number $\delta > 0$, there is a natural number $N_0 = N_0(p,\delta)$ such that for $N \geq N_0$, every set of at least δN^2 points of $[N]^2$ contains a set of p points that determine at least $p - \lceil \log_2 p \rceil$ perfect triangles.*

In the proof we also apply the Regularity Lemma. In the next section we provide the tools including the Regularity Lemma. Then in Section 3 we prove the theorem.

2 Tools

2.1 The Regularity Lemma

In the proof the Regularity Lemma of Szemerédi ([13]) plays a central role. Here we will use the following variation of the lemma.

Lemma 1 (Regularity Lemma – Degree form). *For every $\varepsilon > 0$ there is an $M = M(\varepsilon)$ such that if $G = (V, E)$ is any graph and $d \in [0, 1]$ is any real number, then there is a partition of the vertex-set V into $l+1$ sets (so-called clusters) $C_0, C_1, ..., C_l$, and there is a subgraph $G' = (V, E')$ with the following properties:*

- $l \leq M$,
- $|C_0| \leq \varepsilon |V|$,
- *all clusters C_i, $i \geq 1$, are of the same size,*
- $deg_{G'}(v) > deg_G(v) - (d+\varepsilon)|V|$ *for all* $v \in V$,
- $G'|_{C_i} = \emptyset$ *(C_i are independent in G'),*
- *all pairs $G'|_{C_i \times C_j}$, $1 \leq i < j \leq l$, are ε-regular, each with a density 0 or exceeding d.*

This form (see [6]) can easily be obtained by applying the original Regularity Lemma (with a smaller value of ε), adding to the exceptional set C_0 all clusters incident to many irregular pairs, and then deleting all edges between any other clusters where the edges either do not form a regular pair or they do but with a density at most d.

2.2 Applying the Regularity Lemma

We will prove the following lemma by applying the Regularity Lemma.

Lemma 2. *For every $c_1 > 0$, $c_2 \geq 1$ there are positive constants η, n_0 with the following properties. Let G be a graph on $n \geq n_0$ vertices with $|E(G)| \geq c_1 n^2$ that is the edge disjoint union of matchings $M_1, M_2, \ldots, M_m$ where $m \leq c_2 n$. Then there exist an $1 \leq i \leq m$ and $A, B \subset V(M_i)$ such that*

- $(A \times B) \cap M_i = \emptyset$,
- $|A| = |B| \geq \eta n$,
- $|E(G|_{A\times B})| \geq \frac{c_1}{4}|A||B|$.

Proof: We note that the proof of this lemma already appeared in [8] (see also [9]), but for the sake of completeness we give the proof here as well.

Let us apply the degree form of the Regularity Lemma (Lemma 1) with

$$d = \frac{c_1}{2} \quad \text{and} \quad \varepsilon = \frac{c_1}{6c_2}. \tag{1}$$

Let $G'' = G' \setminus C_0$. Then we have

$$deg_{G''}(v) > deg_G(v) - (d+\varepsilon)n - |C_0| \geq deg_G(v) - (d+2\varepsilon)n \quad \text{for all} \quad v \in V(G'').$$

Thus using (1)

$$|E(G'')| = \frac{1}{2}\sum_{v\in V(G'')} deg_{G''}(v) > \frac{1}{2}\sum_{v\in V(G'')} deg_G(v) - \frac{d+2\varepsilon}{2}n^2 =$$

$$= \frac{1}{2}\sum_{v\in V(G)} deg_G(v) - \frac{1}{2}\sum_{v\in C_0} deg_G(v) - \frac{d+2\varepsilon}{2}n^2 \geq |E(G)| - \frac{d+3\varepsilon}{2}n^2 \geq \frac{c_1}{2}n^2.$$

Hence there is an $1 \leq i \leq m$ such that

$$|M_i|_{G''}| > \frac{c_1}{2c_2}n = 3\varepsilon n. \tag{2}$$

Write $U = V(M_i|_{G''})$ for the vertex set of $M_i|_{G''}$. (2) implies that $|U| > 6\varepsilon n$. Write also $U_i = U \cap C_i$. Define $I = \{i \mid |U_i| > 3\varepsilon|C_i|\}$, and set $U' = \cup_{i\in I} U_i$ and $U'' = U \setminus U'$. Clearly $|U''| \leq 3\varepsilon n$. Since $|U| > 6\varepsilon n$, we have two vertices $u, v \in U'$ adjacent in $M_i|_{G''}$. Let $u \in C_i$ and $v \in C_j$. In G'' we have at least one edge between C_i and C_j, and hence we must have a density more than $d = \frac{c_1}{2}$ between them. Consider U_i and U_j. A is an arbitrary subset of U_i with $|A| = \lfloor \varepsilon|C_i|\rfloor + 1 > \varepsilon|C_i|$. B is an arbitrary subset of U_j with $|B| = \lfloor \varepsilon|C_j|\rfloor + 1 > \varepsilon|C_j|$ and $(A \times B) \cap M_i = \emptyset$. This is possible since

$$|U_j| > 3\varepsilon|C_j| > 2\lfloor \varepsilon|C_j|\rfloor + 2,$$

if $n \geq n_0$. Then the first property of A, B in the lemma is clearly satisfied. For the second property we can choose $\eta = \frac{\varepsilon(1-\varepsilon)}{M(\varepsilon)}$. Finally for the third property, ε-regularity of the pair (C_i, C_j) implies that the density between A and B is more than $d - \varepsilon \geq \frac{c_1}{4}$. This means

$$|E(G|_{A\times B})| \geq \frac{c_1}{4}|A||B|,$$

and thus completing the proof of the lemma. □

3 Proof of Theorem 1

Let $p \geq 3$ be an integer and $\delta > 0$ a real number. Let $l = \lceil \log_2 p \rceil$. Assume that N is sufficiently large.

Let S be a set of at least δN^2 points of the grid $[N]^2$. The points of the grid will be represented by their coordinates (i, j) for $i, j \in \{1, \ldots, N\}$.

Following Solymosi [10], we define the bipartite graph $G_b = G_b(A, B)$ with vertex sets $A = \{u_1, \ldots, u_N\}$ and $B = \{v_1, \ldots, v_N\}$, where we have an edge $(u_i, v_j) \in E(G_b)$ iff $(i, j) \in S$. Thus $E(G_b) \geq \delta N^2$. We partition the edges of G_b into $2N - 1$ matchings M_i. The matching M_i contains those edges $(u_k, v_l) \in E(G_b)$ for which $k + l = i$.

Next by applying Lemma 2 iteratively in G_b, we will find a sequence of submatchings $\overline{M}_1, \ldots, \overline{M}_l$. To obtain $\overline{M}_1$ we apply Lemma 2 in G_b. We can choose

$$c_1 = c_1^1 = \frac{\delta}{4} \quad \text{and} \quad c_2 = c_2^1 = 1.$$

$\overline{M}_1$ is the M_i guaranteed in the lemma. Denote $\overline{M}_1 = (A_1, B_1)$ where $A_1 \subset A, B_1 \subset B$. Lemma 2 also guarantees that there are $A_1', B_1' \subset V(\overline{M}_1)$ such that

- $(A_1' \times B_1') \cap \overline{M}_1 = \emptyset$,
- $|A_1'| = |B_1'| \geq \eta_1 n$,
- $\left|E\left(G_b|_{A_1' \times B_1'}\right)\right| \geq \frac{c_1}{4}|A_1'||B_1'|$.

To obtain $\overline{M}_2$ we apply Lemma 2 again, now for $G_b|_{A_1' \times B_1'}$. Here we can choose

$$c_1 = c_1^2 = \frac{c_1^1}{16} \quad \text{and} \quad c_2 = c_2^2 = \frac{c_2^1}{2\eta_1}.$$

$\overline{M}_2$ is the M_i guaranteed in the lemma. Note that technically this $\overline{M}_2$ is not the whole M_i in G_b, but it is restricted to $G_b|_{A_1' \times B_1'}$. Denote $\overline{M}_2 = (A_2, B_2)$ where $A_2 \subset A_1, B_2 \subset B_1$.

We continue in this fashion. Assume that $\overline{M}_j = (A_j, B_j)$ is already defined where $A_j \subset A_{j-1}, B_j \subset B_{j-1}$. Futhermore, we have $A_j', B_j' \subset V(\overline{M}_j)$ such that

- $(A_j' \times B_j') \cap \overline{M}_j = \emptyset$,
- $|A_j'| = |B_j'| \geq \eta_j(|A_{j-1}'| + |B_{j-1}'|)$,
- $\left|E\left(G_b|_{A_j' \times B_j'}\right)\right| \geq \frac{c_1^j}{4}|A_j'||B_j'|$.

To obtain $\overline{M}_{j+1}$ we apply Lemma 2 for $G_b|_{A_j' \times B_j'}$. We can choose

$$c_1 = c_1^{j+1} = \frac{c_1^j}{16} \quad \text{and} \quad c_2 = c_2^{j+1} = \frac{c_2^j}{2\eta_j}.$$

$\overline{M}_{j+1}$ is the M_i guaranteed in the lemma. Denote $\overline{M}_{j+1} = (A_{j+1}, B_{j+1})$. We continue until $\overline{M}_1, \ldots, \overline{M}_l$ are selected.

Next using these matchings $\overline{M}_j$ we will select a set of p points containing at least $p - \lceil \log_2 p \rceil$ perfect triangles.

Lemma 3. *For any $1 \leq i \leq l = \lceil \log_2 p \rceil$, let G_i be the graph obtained from bipartite graph $\left(A, B, \cup_{j=1}^{i} \overline{M}_j\right)$ by removing all components which do not contain a vertex of $A_i \cup B_i$. The vertices of G_i are partitioned into $|\overline{M}_i|$ trees, each with $2^i - 1$ edges.*

Proof: We use induction on i. For $i = 1$, G_1 is just $\overline{M}_1$, and each tree of G_1 has one edge. We assume the lemma to hold for $i - 1$. Each endpoint of each edge $e \in \overline{M}_i$ is in $A_{i-1} \cup B_{i-1}$ and thus by the inductive hypothesis belongs to exactly one tree of G_{i-1}, and each of these trees has $2^{i-1} - 1$ edges. Edge e, along with the two trees it joins, comprise a new tree with $2^i - 1$ edges.□

In G_i we will call the edges $(u_1, v_1), (u_2, v_1), (u_2, v_2)$ a *triple*, if $(u_1, v_1), (u_2, v_2) \in \overline{M}_j$ and $(u_2, v_1) \in \overline{M}_{j'}$ for some $1 \leq j < j' \leq i$. Furthermore, the (u_2, v_1) edge is called the *center* of the triple, and the *index* of the triple is j'. For an edge e in G_i, the index of e is the maximum number m such that there is a triple with index m containing e. Note that each triple corresponds to a perfect triangle in S.

Lemma 4. *There exist p points in S containing at least $p - \lceil \log_2 p \rceil$ perfect triangles.*

Proof: In case $p = 2^l - 1$, we consider a tree τ in G_l, and then the p points of S are the points corresponding to the p edges of τ. Then the number of triples in τ (and thus the number of perfect triangles) is at least:

$$2^{l-2} + 22^{l-3} + \ldots + (l-1)2^0 = 2^l - 1 - l = p - \lceil \log p \rceil.$$

This is because each $\overline{M}_2$ edge of τ is a center for one triple in τ, each $\overline{M}_3$ edge of τ is a center for two triples, etc. finally the one $\overline{M}_l$ edge of τ is a center for $(l-1)$ triples.

In case $2^{l-1} - 1 < p < 2^l - 1$, again consider a tree τ in G_l, as above. τ consists of two trees τ_1 and τ_2 in G_{l-1} joined by an $\overline{M}_l$ edge. We will start removing leaves from τ_1 until we have exactly p edges left in the remaining tree, and then these p edges will correspond to the p points of S in the lemma. Furthermore, we will remove the leaves in such a way, that with each removal we destroy only one new triple (and thus perfect triangle), and thus on the resulting p points we will have the $p - \lceil \log_2 p \rceil$ perfect triangles, as desired. In order to achieve this goal, from τ_1 we always remove the leaf with the minimum index, where if we have several leaves with the minimum index, we remove one of them arbitrarily. From the construction, it is not hard to see that with each removal we destroy only one new perfect triangle. This implies that when we finish and we have only p edges left, we have at least $p - \lceil \log_2 p \rceil$ perfect triangles on the corresponding points in S.

This completes the proof of Theorem 1. □

References

[1] M. Ajtai, E. Szemerédi, Sets of lattice points that form no squares, *Studia Scient. Math. Hung.*, 9 (1974), 9-11.

[2] B. Bollobás, *Extremal Graph Theory*, Academic Press, London (1978).

[3] H. Fürstenberg, Y. Katznelson, A density version of the Hales-Jewett Theorem, *J. d'Analyse Math.*, 57 (1991), 64-119.

[4] T. Gowers, A new proof of Szemerédi's Theorem, *Geom. Funct. Anal.*, 11(3) (2001), 465-588.

[5] T. Gowers, Some unsolved problems in additive/combinatorial number theory, preprint.

[6] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory, in *Combinatorics, Paul Erdős is Eighty* (D. Miklós, V.T. Sós, and T. Szőnyi, Eds.), pp. 295-352, Bolyai Society Mathematical Studies, Vol. 2, Budapest, 1996.

[7] I.Z. Ruzsa, E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II.* 939-945.

[8] G.N. Sárközy, S.M. Selkow, An extension of the Ruzsa-Szemerédi Theorem, submitted for publication, WPI Technical Report WPICS-CS-TR-03-02, http://www.cs.wpi.edu/Resources/techreports/

[9] G.N. Sárközy, S.M. Selkow, On a hypergraph problem of Brown, Erdős and T. Sós, submitted for publication, WPI Technical Report WPICS-CS-TR-03-05, http://www.cs.wpi.edu/Resources/techreports/

[10] J. Solymosi, Note on a generalization of Roth's Theorem, to appear.

[11] J. Solymosi, Note on a question of Erdős and Graham, to appear.

[12] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.*, 27 (1975), 299-355.

[13] E. Szemerédi, Regular partitions of graphs, Colloques Internationaux C.N.R.S. N$^{\underline{o}}$ 260 - Problèmes Combinatoires et Théorie des Graphes, Orsay (1976), 399-401.

[14] V.H. Vu, On a question of Gowers, *Annals of Combinatorics*, 6 (2002), 229-233.