Graph regularity and incidence phenomena in distal structures

Artem Chernikov

(IIMJ-PRG)

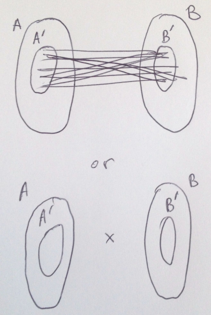
Luminy,

April 7, 2015

- Joint with Sergei Starchenko, University of Notre Dame.

Homogeneous subsets

- Let (R, A, B) be a bipartite graph, i.e. A and B are two disjoint sets of vertices and $R \subseteq A \times B$.
- We say that a pair of sets $A' \subseteq A, B' \subseteq B$ is *R-homogeneous* if either $A' \times B' \subseteq R$ or $(A' \times B') \cap R = \emptyset$.



- [Kövári, Sós, Turán, Erdős] If $|A|, |B| \geq n$, then there is a homogeneous pair (A', B') with $|A'|, |B'| \geq c \log n$.

Semialgebraic graphs

- Optimal in general. But what if we restrict to some geometrically motivated graphs?
- A set $A \subseteq \mathbb{R}^d$ is *semialgebraic* if it can be defined by a finite boolean combination of polynomial equalities and inequalities.
- Examples: hyperplanes, balls, boxes, tubes, etc. in $\mathbb{R}^d$.
- We say that the *description complexity* of a semialgebraic set $A \subseteq \mathbb{R}^d$ is $\leq t$ if $d \leq t$ and A can be defined by a boolean combination involving at most t polynomial inequalities, each of degree at most t.
- Examples of semialgebraic graphs and hypergraphs:
  - the incidence relation between points and lines on the plane,
  - pairs of circles in $\mathbb{R}^3$ that are linked,
  - two parametrized families of semialgebraic varieties having a non-empty intersection,
  - multi-dimensional analogues, etc.

Semialgebraic Ramsey

- [N. Alon, J. Pach, R. Pinchasi, R. Radoičić, M. Sharir, "Crossing patterns of semi-algebraic sets", 1995]:

Theorem

For every $t \in \mathbb{N}$ there is some $\varepsilon > 0$ such that: if $R \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is semialgebraic of complexity bounded by t, then for any finite sets $A \subseteq \mathbb{R}^{d_1}, B \subseteq \mathbb{R}^{d_2}$ there are some $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq \varepsilon |A|, |B'| \geq \varepsilon |B|$ and (A', B') is R-homogeneous. Moreover, $A' = A \cap S_1$ and $B' = B \cap S_2$, where S_1, S_2 are certain semialgebraic sets of complexity bounded in terms of t.

- This result has many applications: semialgebraic regularity lemma, incidence questions, unit distance problem, higher dimensional Ramsey, etc.

Motivation for our work

- Some natural questions:
  - Can we allow more complicated graphs, e.g. if we want to define the edge relation via some conditions expressed in terms of e^x or some analytic functions? What about graphs coming from p-adic geometry?
  - Can we prove similar results for more general measures (other than just counting points, e.g. Lebesgue, Haar)?
- Model theory provides both context and methods for such generalizations.

Back to the Ramsey statement

- The previous result can be reformulated by saying that $M = (\mathbb{R}, +, \times, 0, 1)$ satisfies the following property.
- $(*)$ For every *definable* relation $R \subseteq M^{d_1} \times M^{d_2}$ there is some $\varepsilon > 0$ such that: for every finite $A \subseteq M^{d_1}, B \subseteq M^{d_2}$ there are some $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq \varepsilon |A|, |B'| \geq \varepsilon |B|$ and (A', B') is R-homogeneous.
  Moreover, $A' = A \cap S_1$ and $B' = B \cap S_2$, where S_1, S_2 are definable by a certain formula depending just on the formula defining R (and not on its parameters).
- Which other structures satisfy $(*)$?

o-minimal structures satisfy $(*)$

- [Basu, 2007] Topologically closed graphs in *o*-minimal expansions of real closed fields satisfy $(*)$.
- E.g., $M = \left(\mathbb{R}, +, \times, e^x, f \restriction_{[0,1]} \text{ for } f \text{ restricted analytic}\right)$.
- As the logarithmic bound on the size of homogeneous subsets is optimal for general graphs, it follows that $(*)$ implies NIP (i.e. all uniformly definable families of sets have finite VC-dimension).

$(*)$ fails in algebraically closed fields of positive characteristic

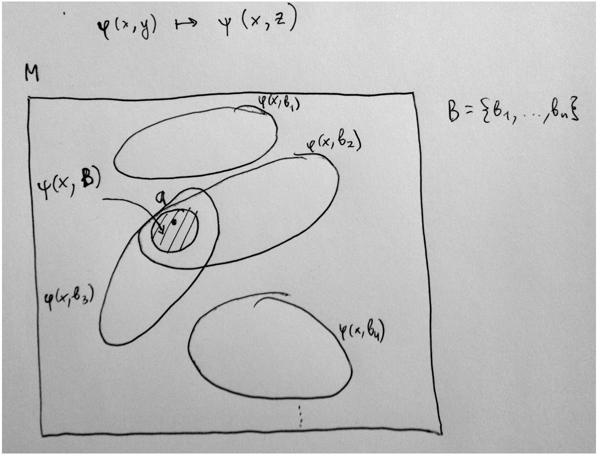
- Without requiring definability of the homogeneous sets $(*)$ holds in algebraically closed fields of char 0 — as $(\mathbb{C}, \times, +)$ is interpreted in $\left(\mathbb{R}^2, \times, +\right)$.
- For a finite field $\mathbb{F}_q$, let P_q be the set of all points in $\mathbb{F}_q^2$ and let L_q be the set of all lines in $\mathbb{F}_q^2$.
- Let $I \subseteq P_q \times L_q$ be the incidence relation. Using the fact that the lazy Szemerédi-Trotter bound $|I(P_q, L_q)| \leq |L_q| \, |P_q|^{\frac{1}{2}} + |P_q|$ is optimal in finite fields one can check:
- **Claim**. For any fixed $\delta > 0$, for all large enough q if $L_0 \subseteq L_q$ and $P_0 \subseteq P_q$ with $|P_0| \geq \delta q^2$ and $|L_0| \geq \delta q^2$ then $I(P_0, L_0) \neq \emptyset$.
- As every finite field of char p can be embedded into $\bar{\mathbb{F}}_p$, it follows that $(*)$ fails in $\bar{\mathbb{F}}_p$ (even without requiring definability of the homogeneous pieces) for I the incidence relation.

Towards the right setting

- The class of *distal structures* was introduced and studied by [P. Simon, 2011] in order to capture the class of "purely unstable" NIP theories.
- The original definition is in terms of certain properties of indiscernible sequences.
- [C., Simon, 2012] gives a combinatorial characterization of distality:

Distal structures

- **Theorem/Definition** An NIP structure M is *distal* if and only if for every definable family $\{\phi(x,b) : b \in M^d\}$ of subsets of M there is a definable family $\{\psi(x,c) : c \in M^{kd}\}$ such that for every $a \in M$ and every finite set $B \subset M^d$ there is some $c \in B^k$ such that $a \in \psi(x,c)$ and for every $a' \in \psi(x,c)$ we have $a' \in \phi(x,b) \Leftrightarrow a \in \phi(x,b)$, for all $b \in B$.



Examples of distal structures

- All (weakly) *o*-minimal structures are distal, e.g. RCVF.
- Any *p*-minimal theory with Skolem functions is distal. E.g.$(\mathbb{Q}_p, +, \times)$ for each prime *p*, with analytic exapnsion, is distal (due to the *p*-adic cell decomposition of Denef).
- Certain topological differential (valued) fields (see Point's talk) and the ordered differential field of transseries (via recent work of Aschenbrenner, van den Dries, van der Hoeven) are distal.
- Nice pairs of distal structures are distal.

Keisler measures

- A *(Keisler) measure* μ over a structure M is a finitely additive probability measure on the boolean algebra $\mathrm{Def}_x(M)$ of definable subsets of M.
- Let $S_x(M)$ be the compact space of types over M, i.e. the Stone dual of $\mathrm{Def}_x(M)$. Every Keisler measure over M can be viewed as a measure defined on all clopen subsets $S_x(M)$, and then it admits a unique extension to a regular Borel probability measure on $S_x(M)$.
- Let $\mathbb{M} \succ M$ be a saturated elementary extension of M (a "universal domain", in the case of real closed fields we in particular throw in some infinitesimals, infinitesimals with respect to those infinitesimals, etc.)

Generically stable measures, 1

- A measure μ over an NIP struture M is *generically stable* if there is a unique $\mathrm{Aut}(\mathbb{M}/M)$-invariant Keisler measure over $\mathbb{M}$ extending μ.
- [Vapnik–Chervonenkis, 1971] + [Hrushovski, Pillay, Simon, 2010]: Generically stable measures are uniformly approximable by frequency measures: for every $\phi(x, y) \in L$ and $\varepsilon > 0$ there is some $n \in \mathbb{N}$ such that for every generically stable measure μ over M there are some $a_0, \ldots, a_{n-1} \in M^{|x|}$ such that for any $b \in M^{|y|}$ we have $\left|\mu(\phi(x, b)) - \frac{|\{i<n : \models \phi(a_i, b)\}|}{n}\right| \leq \varepsilon$.

Generically stable measures, 2

- Examples of generically stable measures:
  - A counting measure concentrated on a finite set (in any structure).
  - Lebesgue measure on $[0, 1]$ (over reals, restricted to definable sets).
  - Haar measure on a compact ball over p-adics.
  - Let G be a (definably) compact group in an o-minimal theory or over p-adics. Then it admits a unique G-invariant measure, which is generically stable.

Main results: Distal Ramsey

Theorem

[C., Starchenko] Let M be a distal structure. Then it satisfies:

1. *Strong $(*)$: For every definable relation $R(x,y)$ there is some $\varepsilon > 0$ such that: for all generically stable measures μ on $M^{|x|}$ and ν on $M^{|y|}$ there are some sets $S_1 \subseteq M^{|x|}, S_2 \subseteq M^{|y|}$ uniformly definable depending just on R, such that $\mu(S_1) \geq \varepsilon, \nu(S_2) \geq \varepsilon$ and (S_1, S_2) is R-homogeneous.*
2. *Moreover, if M satisfies $(*)$ then M is distal.*

- Of course, strong $(*)$ implies $(*)$ by taking μ, ν to be counting measures concentrated on finite sets.
- In the case of p-adics, not uniform in p: the problem with $\mathbb{F}_p$ is treated by increasing the constant.
- Density version, version for hypergraphs, etc.

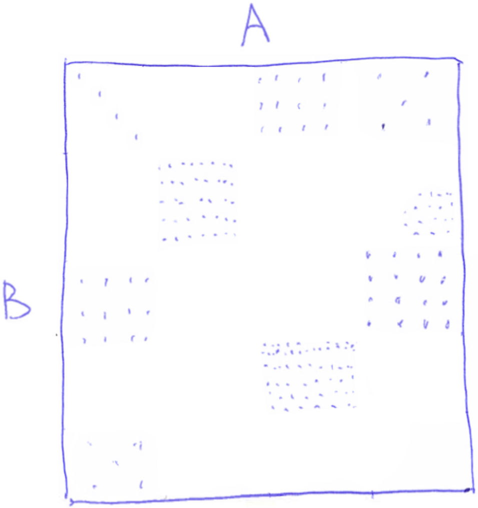
Szemerédi regularity lemma

Theorem

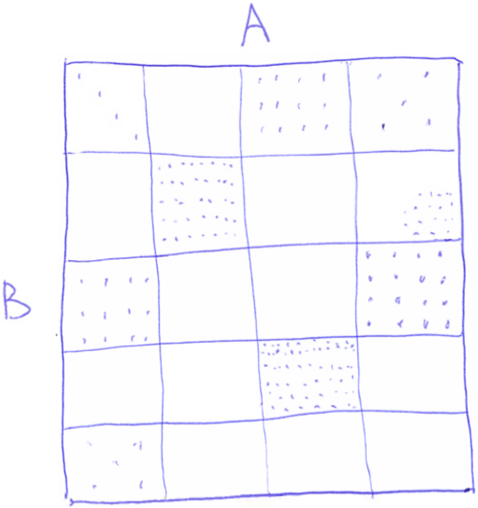
[E. Szemerédi, 1975] If $\varepsilon > 0$, *then there exists* $K = K(\varepsilon)$ *such that:*
for any finite bipartite graph $R \subseteq A \times B$, *there exist partitions* $A = A_0 \cup \ldots \cup A_k$ *and* $B = B_0 \cup \ldots \cup B_k$ *into non-empty sets, and a set* $\Sigma \subseteq \{1, \ldots, k\} \times \{1, \ldots, k\}$ *with the following properties.*

1. *Bounded size of the partition:* $k \leq K$.
2. *Few exceptions:* $\left|\bigcup_{(i,j)\in\Sigma} A_i \times B_j\right| \geq (1-\varepsilon)|A \times B|$.
3. ε*-regularity: for all* $(i,j) \in \Sigma$, *and all* $A' \subseteq A_i$, $B' \subseteq B_j$, *one has*
$$\left|\frac{|R \cap (A' \times B')|}{|A' \times B'|} - \frac{|R \cap (A_i \times B_j)|}{|A_i \times B_j|}\right| \leq \varepsilon.$$

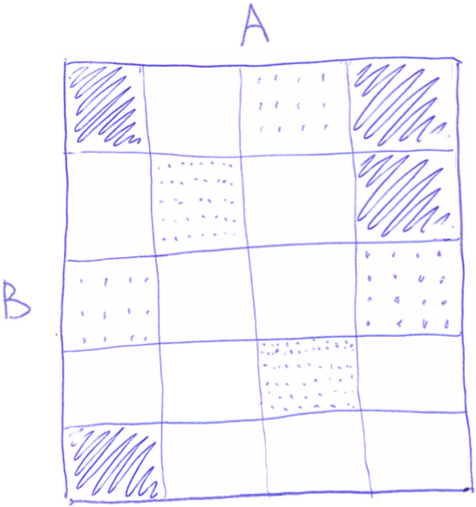
Szemerédi regularity lemma



Szemerédi regularity lemma



Szemerédi regularity lemma



Szemerédi regularity lemma: bounds and applications

- Exist various versions for weaker and stronger partitions, for hypergraphs, etc. Increasing the error a little one may assume that sets in the partition are of (approximately) equal size.
- Has many applications in extreme graph combinatorics, additive number theory, computer science, etc.
- [T. Gowers, 1997] The size of the partition $K(\varepsilon)$ grows as a tower of twos $2^{2^{\cdots}}$ of height $\left(1/\varepsilon^{16}\right)$.
- What about restricted families of graphs?

Classification of regularity lemmas

1. [T. Tao, 2012] Algebraic graphs of bounded complexity in large finite fields (pieces of the partition are algebraic, no exceptional pairs, stronger regularity), based on the work of [Chatzidakis, van den Dries, Macintyre].
   1.1 + some generalizations by Hrushovski; Pillay, Starchenko; Macpherson, Steinhorn.
2. [L. Lovász, B. Szegedi, 2010] Graphs of bounded VC-dimension, i.e. NIP graphs (density arbitrarily close to 0 or 1, the size of the partition is bounded by a polynomial in $\left(\frac{1}{\varepsilon}\right)$).
   2.1 [M. Malliaris, S. Shelah, 2011]: graphs without arbitrary large half-graphs, i.e. stable graphs (no exceptional pairs).
   2.2 [J.Fox, M. Gromov, V. Lafforgue, A. Naor, and J. Pach, "Overlap properties of geometric expanders", 2010], [J. Fox, J. Pach, A. Suk, "A polynomial regularity lemma for semi-algebraic hypergraphs and its applications in geometry and property testing", 2015] Semialgebraic graphs of bounded complexity.

Application: Distal regularity lemma

Theorem

[C., Starchenko] Let M be distal. For every definable $R(x,y)$ and every $\varepsilon > 0$ there is some $K = K(\varepsilon, R)$ such that: for any generically stable measures μ on $M^{|x|}$ and ν on $M^{|y|}$, there are $A_0, \ldots, A_k \subseteq M^{|x|}$ and $B_0, \ldots, B_k \subseteq M^{|y|}$ uniformly definable depending just on R and ε, and a set $\Sigma \subseteq \{1, \ldots, k\}^2$ such that:

1. $k \leq K$,
2. $\omega\left(\bigcup_{(i,j)\in\Sigma} A_i \times B_j\right) \geq 1 - \varepsilon$, *where ω is the (unique, generically stable) product measure of μ and ν,*
3. *for all $(i,j) \in \Sigma$, the pair (A_i, B_j) is R-homogeneous.*

Moreover, K is bounded by a polynomial in $\left(\frac{1}{\varepsilon}\right)$.

Application: Erdős-Hajnal property

- Let (G, V) be an undirected graph. A subset $V_0 \subseteq V$ is *homogeneous* if either $(v, v') \in E$ for all $v \neq v' \in V_0$ or $(v, v') \notin E$ for all $v \neq v' \in V_0$.
- A class of finite graphs $\mathcal{G}$ has the *Erdős-Hajnal property* if there is $\delta > 0$ such that every $G \in \mathcal{G}$ has a homogeneous subset of size $\geq |V(G)|^{\delta}$.
- **Erdős-Hajnal conjecture.** For every finite graph H, the class of all H-free graphs has the Erdős-Hajnal property.
- **Fact.** If $\mathcal{G}$ is a class of finite graphs closed under subgraphs and $\mathcal{G}$ satisfies $(*)$ (without requiring definability of pieces), then $\mathcal{G}$ has the Erdős-Hajnal property.
- Thus, we obtain many new families of graphs satisfying the Erdős-Hajnal conjecture (e.g. quantifier-free definable graphs in arbitrary valued fields of characteristic 0).