Model theory and hypergraph regularity

Artem Chernikov

UCLA

AMS Special Session on Recent Advances in Regularity Lemmas

Baltimore, US, Jen 15, 2019

Model theory and combinatorics

- Infinitary combinatorics is one of the essential ingredients of the classification program in model theory.
- A well investigated theme: close connection of the combinatorial properties of a family of finite structures with the model theory of its infinite limit (smoothly approximable structures, homogeneous structures, etc.).
- More recent trend: applications of (generalized) stability-theoretic techniques for extremal combinatorics of "tame" finite structures.
- Parallel developments in combinatorics, surprisingly well aligned with the model-theoretic approach and dividing lines in Shelah's classification.
- We survey some of these results (group-theoretic regularity lemmas, again closely intertwined with the study of definable groups in model theory, will be discussed in the other talks).

Szemerédi's regularity lemma, standard version

- By a graph $G = (V, E)$ we mean a set G with a symmetric subset $E \subseteq V^2$. For $A, B \subseteq V$ we denote by $E(A, B)$ the set of edges between A and B.
- [Szemerédi regularity lemma] Let $G = (V, E)$ be a finite graph and $\varepsilon > 0$. There is a partition $V = V_1 \cup \cdots \cup V_M$ into disjoint sets for some $M < M(\varepsilon)$, where the constant $M(\varepsilon)$ depends on ε only, real numbers $\delta_{ij}, i, j \in [M]$, and an exceptional set of pairs $\Sigma \subseteq [M] \times [M]$ such that

$$\sum_{(i,j)\in\Sigma} |V_i||V_j| \leq \varepsilon |V|^2$$

and for each $(i, j) \in [M] \times [M] \setminus \Sigma$ we have

$$\big| |E(A, B)| - \delta_{ij}|A||B| \big| < \varepsilon |V_i||V_j|$$

for all $A \subseteq V_i$, $B \subseteq V_j$.
- Regularity lemma can naturally be viewed as a more general measure theoretic statement.

Context: ultraproducts of finite graphs with Loeb measure

- For each $i \in \mathbb{N}$, let $G_i = (V_i, E_i)$ be a graph with $|V_i|$ finite and $\lim_{i\to\infty} |V_i| = \infty$.
- Given a non-principal ultrafilter $\mathcal{U}$ on $\mathbb{N}$, the ultraproduct

$$(V, E) = \prod_{i\in\mathbb{N}} (V_i, E_i)$$

  is a graph on the set V of size continuum.
- Given $k \in \mathbb{N}$ and an *internal* set $X \subseteq V^k$ (i.e. $X = \prod_{\mathcal{U}} X_i$ for some $X_i \subseteq V_i^k$), we define $\mu^k(X) := \lim_{\mathcal{U}} \frac{|X_i|}{|V_i|^k}$. Then:
  - μ^k is a finitely additive probability measure on the Boolean algebra of internal subsets of V^k,
  - extends uniquely to a countably additive measure on the σ-algebra $\mathcal{B}_k$ generated by the internals subsets of V^k(using saturation).
- Then $\left(V, \mathcal{B}_k, \mu^k\right)$ is a *graded probability space*, in the sense of Keisler (satisfies Fubini, etc.).
- Many other examples, with $V = M$ some first-orders structure and $\mathcal{B}_k$ the definable subsets of M^k.

Szemerédi's regularity lemma as a measure-theoretic statement: Elek-Szegedy, Tao, Towsner, ...

- Via orthogonal projection in L^2 onto the subspace of $\mathcal{B}_1 \times \mathcal{B}_1 \subsetneq \mathcal{B}_2$-measurable functions (conditional expectation) we have:
- [Regularity lemma] Given a graded probability space $\left(V, \mathcal{B}_k, \mu^k\right)$, $E \in \mathcal{B}_2$ and $\varepsilon > 0$, there is a decomposition of the form

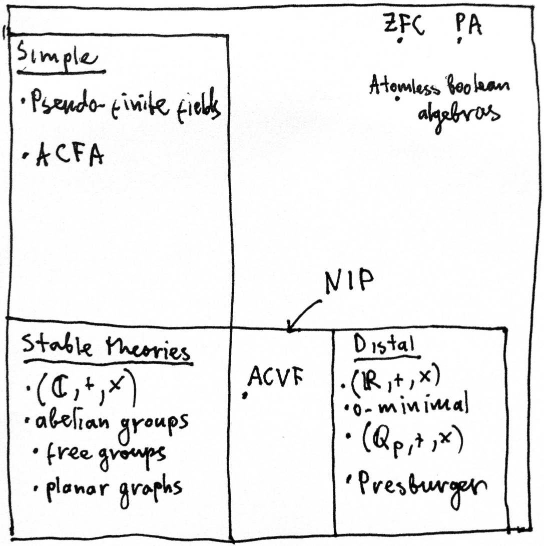
$$1_E = f_{\mathrm{str}} + f_{\mathrm{qr}} + f_{\mathrm{err}},$$

  where:
  - $f_{\mathrm{str}} = \sum_{i \leq n} d_i 1_{A_i}(x) 1_{B_i}(y)$ for some $M = M(\varepsilon) \in \mathbb{N}$, $A_i, B_i \in \mathcal{B}_1$ and $d_i \in [0,1]$ (so f_{str} is $\mathcal{B}_1 \times \mathcal{B}_1$-simple),
  - $f_{\mathrm{err}} : V^2 \to [-1,1]$ and $\int_{V^2} \left|f_{\mathrm{err}}\right|^2 d\mu^2 < \varepsilon$,
  - f_{qr} is *quasi-random:* for any $A, B \in \mathcal{B}_1$ we have $\int_{V^2} 1_A(x) 1_B(y) f_{\mathrm{qr}}(x,y) \, d\mu^2 = 0$.
- Hypergraph regularity lemma: via a sequence of conditional expectations on nested algebras.

Better regularity lemmas for tame structures

- Some features for general graphs:
  - [Gowers] $M(\varepsilon)$ grows as an exponential tower of 2's of height polynomial in $\frac{1}{\varepsilon}$.
  - Bad pairs are unavoidable in general (half-graphs).
  - Quasi-randomness (f_{qr}) is unavoidable in general.
- Turns our that these issues are closely connected to certain properties of first-order theories from Shelah's classification (we'll try to present them in the most "finitary" way possible).

Classification



VC-dimension and NIP

- Given $E \subseteq V^2$ and $x \in V$, let $E_x = \{y \in V : (x, y) \in E\}$ be the x-fiber of E.
- A graph $E \subseteq V^2$ has *VC-dimension* $\geq d$ if there are some $y_1, \ldots, y_d \in V$ such that, for every $S \subseteq \{y_1, \ldots, y_d\}$ there is $x \in V$ so that $E_x \cap \{y_1, \ldots, y_d\} = S$.
- **Example.** If E_i is a random graph on V_i and $(V, E) = \prod_{\mathcal{U}} (V_i, E_i)$, then $\mathrm{VC}(E) = \infty$.
- **Example.** If E is definable in an NIP theory (e.g. E is semialgebraic, definable in $\mathbb{Q}_p$, ACVF, etc.), then $\mathrm{VC}(E) < \infty$.
- [Sauer-Shelah] If $\mathrm{VC}(E) \leq d$, then for any $Y \subseteq V, |Y| = n$ we have $|\{S \subseteq Y : \exists x \in V,\ S = Y \cap E_x\}| = O\left(n^d\right)$.

Regularity lemma for graphs of finite VC-dimension

- [Lovasz, Szegedy] Let $\left(V, \mathcal{B}_k, \mu^k\right)$ be given by an ultraproduct of finite graphs. If $E \in \mathcal{B}_2$ and $\mathrm{VC}(E) = d < \infty$, then:
  - for any $\varepsilon > 0$, there is some $E' \in \mathcal{B}_1 \times \mathcal{B}_1$ such that $\mu^2(E \Delta E') < \varepsilon$,
  - the number of rectangles in E' is bounded by a polynomial in $\frac{1}{\varepsilon}$ of degree $O\left(d^2\right)$.
- So the quasi-random term disappears from the decomposition, and density on each regular pair is 0 or 1.
- Proof sketch:
  - given $\varepsilon > 0$, by the *VC-theorem* can find $x_1, \ldots, x_n \in V$ such that: for every $y, y' \in V$, $\mu(E_y \Delta E_{y'}) > \varepsilon \implies x_i \in E_y \Delta E_{y'}$ for some i;
  - for each $S \subseteq \{x_1, \ldots, x_n\}$, let $B_S := \left\{y \in V : \bigwedge_{i \leq n} (x_i, y) \in E \leftrightarrow x_i \in S\right\}$;
  - then $\forall y_1, y_2 \in B_S$, $\mu(E_{y_1} \Delta E_{y_2}) < \varepsilon$;
  - for each S, pick some $b_S \in B_S$, and let $E' := \bigcup E_{b_S} \times B_S \in \mathcal{B}_1 \times \mathcal{B}_1$.
  - Then $\mu(E \Delta E') < \varepsilon$.
  - The number of different sets B_S is polynomial by Sauer-Shelah.

For hypergraphs and other measures

- We say that $E \subseteq V^k$ satisfies $\mathrm{VC}(E) < \infty$ if viewing E as a binary relation on $V \times V^{k-1}$, for any permutation of the variables, has finite VC-dimension.
- [C., Starchenko] Let $\left(V, \mathcal{B}_k, \mu^k\right)$ be a graded probability space, $E \in \mathcal{B}_k$ with μ a *finitely approximable* measure and μ^k given by its free product, and $\mathrm{VC}(E) \leq d$. Then for any $\varepsilon > 0$ there is some $E' \in B_1 \times \ldots \times B_1$ such that $\mu^k(E \Delta E') < \varepsilon$ and the number of rectangles needed to define E' is a poly in $1/\varepsilon$ of degree $4(k-1)d^2$.
- Examples of fap measures on definable subsets, apart from the ultraproduct of finite ones: Lebesgue measure on $[0,1]$ in $\mathbb{R}^n$; the Haar measure in $\mathbb{Q}_p$ normalized on a compact ball.
- [Fox, Pach, Suk] improved bound to $O(d)$.

Stable regularity lemma

- Turns out that half-graphs is the only reason for irregular pairs.
- A relation $E \subseteq V \times V$ is d-stable if there are no $a_i, b_i \in V$, $i = 1, \ldots, d$, such that $(a_i, b_j) \in E \iff i \leq j$.
- A relation $E \subseteq V^k$ is d-stable if it is d-stable viewed as a binary relation $V \times V^{k-1}$ for every partition of the variables.
- [Malliaris, Shelah] Regularity lemma for finite k-stable graphs.
- [Malliaris, Pillay] A new proof for graphs and arbitrary Keisler measures. However, their argument doesn't give a polynomial bound on the number of pieces.
- Elaborating on these results, we have:

Stable regularity lemma

Theorem

[C., Starchenko] Let $\left(V, \mathcal{B}_k, \mu^k\right)$ *be a graded probability space, and let* $E \in \mathcal{B}_k$ *be d-stable. Then there is some* $c = c(d)$ *such that: for any* $\varepsilon > 0$ *there are partitions* $\mathcal{P}_i \subseteq \mathcal{B}_1, i = 1, \ldots, k$ *with* $\mathcal{P}_i = \{A_{1,i}, \ldots, A_{M,i}\}$ *satisfying*

1. $M \leq \left(\frac{1}{\varepsilon}\right)^c$;
2. ***for all*** $(i_1, \ldots, i_k) \in \{1, \ldots, M\}^k$ *and* $A'_1 \subseteq A_{1,i_1}, \ldots, A'_k \subseteq A_{k,i_k}$ *from* $\mathcal{B}_1$*we have either* $d_E(A'_1, \ldots, A'_k) < \varepsilon$ *or* $d_E(A'_1, \ldots, A'_k) > 1 - \varepsilon$.

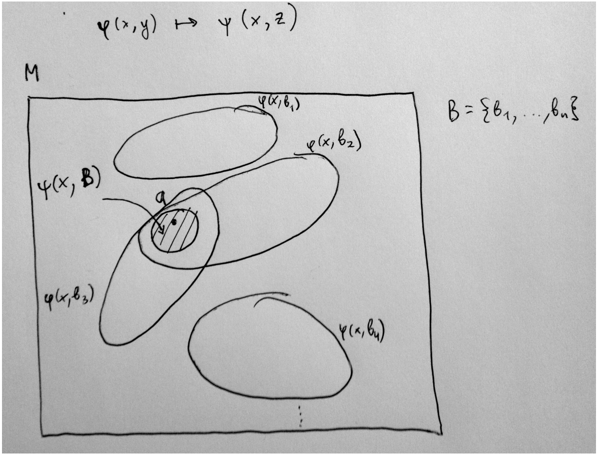
- So, there are no irregular tuples!
- Independently, Ackerman-Freer-Patel proved a variant of this for finite hypergraphs (and more generally, structures in finite relational languages).

Distal case, 1

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

Distal structures

- **Theorem/Definition** A 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, e.g. $M = (\mathbb{R}, +, \times, e^x)$.
- Presburger arithmetic.
- Any *p*-minimal theory with Skolem functions is distal. E.g. $(\mathbb{Q}_p, +, \times)$ for each prime p is distal (e.g. due to the *p*-adic cell decomposition of Denef).
- The differential field of transseries.

Distal regularity lemma

Theorem

[C., Starchenko] Let $\left(V, \mathcal{B}_k, \mu^k\right)$ be a graded probability space with $\mathcal{B}_k$ given by the definable sets in a distal structure M. For every definable $E\left(x_1, \ldots, x_k\right)$ there is some $c = c\left(E\right)$ such that: for any $\varepsilon > 0$ and any finitely approximable measure μ there are partitions $V = \bigcup_{j<K} A_{i,j}$ with sets from $\mathcal{B}_1$ and a set $\Sigma \subseteq \left\{1, \ldots, M\right\}^k$ such that

1. $M \leq \left(\frac{1}{\varepsilon}\right)^c$;
2. $\mu^k\left(\bigcup_{(i_1,\ldots,i_k)\in\Sigma} A_{1,i_1} \times \ldots \times A_{k,i_k}\right) \geq 1 - \varepsilon$;
3. *for all* $(i_1, \ldots, i_k) \in \Sigma$, *either* $(A_{1,i_1} \times \ldots \times A_{k,i_k}) \cap E = \emptyset$ *or* $A_{1,i_1} \times \ldots \times A_{k,i_k} \subseteq E$.

- We can formulate this for general graded probability spaces, but this would require some additional definitions.
- Without the definability of the partition clause passes to reducts, so is satisfied by many stable graphs.

Semialgebraic case

- This generalizes the very important semialgebraic case due to [Fox, Gromov, Lafforgue, Naor, Pach, 2012] and [Fox, Pach, Suk, 2015].
- But also applies e.g. to graphs definable in the p-adics, with respect to the Haar measure.
- Many questions about the optimality of the bounds remain, in the o-minimal and the p-adic cases in particular.

2-dependence

- In the hypergraph regularity lemma, we would like to characterize the arity at which the quasi-random components of the decomposition become trivial.
- The following generalization of VC-dimension is implicit in Shelah's definition of 2-*dependent theories.*
- $E \subseteq V^3$ has *VC_2-dimension* $\geq d$ if there is a rectangle $y_1, \ldots, y_d, z_1, \ldots, z_d \in V$ such that: for every $S \subseteq \{y_1, \ldots, y_d\} \times \{z_1, \ldots, z_d\}$ there is some $x \in V$ so that $E_x \cap (\{y_1, \ldots, y_d\} \times \{z_1, \ldots, z_d\}) = S$.
- **Example:** if E is an ultraproduct of random finite 3-hypergraphs, then $\mathrm{VC}_2(E) = \infty$.
- **Example.** Let $F, G, H \subseteq V^2$ be ultraproducts of random finite graphs and let E consist of those (x, y, z) for which the odd number of pairs $(x, y), (x, z), (y, z)$ belongs to F, G, H, respectively. Then $\mathrm{VC}_2(E) < \infty$.
- **Example.** For any relation $E(x, y, z)$ definable in a smoothly approximable structure, $\mathrm{VC}_2(E) < \infty$.

Towards a regularity lemma

- An analogue of Sauer-Shelah lemma:
- [C., Palacin, Takeuchi] If $\mathrm{VC}_2(E) \leq d$ then $\exists \varepsilon(d) > 0$ such that for any $Y, Z \subseteq V, |Y| = |Z| = n$ we have $|\{S \subseteq Y \times Z : \exists x \in V, S = (Y \times Z) \cap E_x\}| \leq 2^{n^{2-\varepsilon}}$ (close to optimal).
- A generalization of the VC-theorem? Not so clear what it should mean...

Regularity for k-dependent hypergraphs

- Let $\mathcal{B}_{3,2} \subseteq \mathcal{B}_3$ be the algebra generated by "cylindrical" sets of the form

$$\left\{(x, y, z) \in V^3 : (x, y) \in A \wedge (x, z) \in B \wedge (y, z) \in C\right\}$$

for some $A, B, C \in \mathcal{B}_2$. Again, $\mathcal{B}_{3,2} \subsetneq \mathcal{B}_3$.

Theorem

[C., Towsner] Let $\left(V, \mathcal{B}_k, \mu^k\right)$ be a graded probability space given by an ultraproduct of finite sets. If $E \in \mathcal{B}_3$ has finite VC_2-dimension, then for any $\varepsilon > 0$ there is some $E' \in \mathcal{B}_{3,2}$ such that $\mu^3 \left(E \Delta E'\right) < \varepsilon$.

- More generally, we have: for any $n > k$ and any $E \in \mathcal{B}_n$ with finite VC_k-dimension (under any partition of the variables into $k + 1$ groups), E belongs to $\mathcal{B}_{n,k}$.