Intersecting families are approximately contained in juntas

Yuval Filmus

April 6, 2019

Our goal in this sketch is to give an exposition of *Independent Sets in Graph Powers are Almost Contained in Juntas* by Dinur, Friedgut and Regev, as applied to intersecting families by Dinur and Friedgut in *Intersecting families are essentially contained in juntas.*

1 Constant p

Let $\mathcal{F}$ be an intersecting family on n points, and let $p \leq 1/2$. We define the distribution $\mu_{p,p}$ on pairs of subsets A, B as follows: for each element i, with probability p put it in A, with probability p put it in B, and with probability $1-2p$ put it in neither.

We would like to prove the following result:

Theorem 1. *If $\mathcal{F}$ is an intersecting family then for every $p < 1/2$ and $\epsilon > 0$ there exists a junta $\mathcal{H}$ such that:*

(a) $\mathcal{F}$ is almost contained in $\mathcal{H}$: $\mu_p(\mathcal{F} \setminus \mathcal{H}) \leq \epsilon$.

(b) $\mathcal{H}$ is almost intersecting: if $A, B \sim \mu_{p,p}$ then $\Pr[A, B \in \mathcal{H}] \leq \epsilon$.

We comment that this has recently been improved by Friedgut and Regev, *Kneser graphs are like Swiss cheese*, who showed that $\mathcal{H}$ can be assumed to be intersecting. This has been generalized further by Lifshitz, *Hypergraph removal lemmas via robust sharp threshold theorems.*

Let $q = 1 - p$. The starting point is the observation that

$$\langle f, T_{-p/q} g \rangle = f' \begin{bmatrix} 1 - \frac{p}{q} & \frac{p}{q} \\ 1 & 0 \end{bmatrix}^{\otimes n} g.$$

(Here the inner product is with respect to μ_p.) To see this, it suffices to check that

$$\begin{bmatrix} 1 - \frac{p}{q} & \frac{p}{q} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 - \frac{p}{q} & \frac{p}{q} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -p \\ q \end{bmatrix} = -\frac{p}{q} \begin{bmatrix} -p \\ q \end{bmatrix}.$$

In particular, this shows that if $\mathcal{F}$ is intersecting and $f = 1_{\mathcal{F}}$ then $\langle f, T_{-p/q} f \rangle = 0$.

How would we get a junta out of f? Suppose that f was obtained from a junta by taking a random constant fraction of it. If we looked at $T_{1-\eta} f$ then we essentially get back the junta, multiplied by the constant fraction.

For an appropriate parameter η, we look at $g = T_{1-\eta} f$. It is easy to check that $\mathrm{Inf}[g] = O_\eta(1)$, and so we can extract the junta variables out of g by considering the set J of all variables with influence at least γ, for some constant γ.

In order to guarantee property (a) in the theorem, it is natural to choose the family $\mathcal{H}$ by taking all slices on which the expected value of g is at least ϵ. Indeed, let $\tilde{f}, \tilde{g}$ result from averaging over all non-junta coordinates; note $\tilde{g} = T_{1-\eta}\tilde{f}$. Note also that $\mu_p(\mathcal{F} \setminus \mathcal{H}) = \mathbb{E}[\tilde{f} 1_{\overline{\mathcal{H}}}]$. Therefore

$$\mu_p(\mathcal{F} \setminus \mathcal{H})^2 = \mathbb{E}[\tilde{f} 1_{\overline{\mathcal{H}}}]^2 \leq \langle \tilde{f} 1_{\overline{\mathcal{H}}}, T_{1-\eta}\tilde{f} 1_{\overline{\mathcal{H}}} \rangle \leq \langle \tilde{f} 1_{\overline{\mathcal{H}}}, \tilde{g} \rangle \leq \epsilon\, \mathbb{E}[\tilde{f} 1_{\overline{\mathcal{H}}}] = \epsilon \mu_p(\mathcal{F} \setminus \mathcal{H}).$$

The more challenging property to prove is property (b). While g spans few edges, it could be that $\mathcal{H}$ spans many more edges, since we chose all fibers with density at least ϵ (rather than just with at least constant density). The proof combines a version of Majority is Stablest with small-set expansion. We will show that if $\mu_{p,p}(\mathcal{H}) \geq \epsilon$, then g must contain an influential coordinate outside of J, contradicting the definition of J.

For an appropriate parameter τ, for every $a \in \{0,1\}^J$ we let $L(a)$ consist of all variables of influence at least τ in the a-fiber of g. As above, these lists have constant size.

Consider any two $a, b \in \{0,1\}^J$. By construction, $\mathbb{E}[g(a,\cdot)], \mathbb{E}[g(b,\cdot)] \geq \epsilon$. Majority is Stablest shows that either $\langle g(a,\cdot), T_{-p/q} g(b,\cdot) \rangle \geq \delta$ (for some constant δ), or $g(a,\cdot), g(b,\cdot)$ must contain a common influential variable, that is, $L(a)$ and $L(b)$ intersect.

Since $\langle f, T_{-p/q} f \rangle = 0$, calculation shows that $\langle g, T_{-p/q} g \rangle$ is small, hence $\langle g(a,\cdot), T_{-p/q} g(b,\cdot) \rangle$ is small on average. By choosing parameters appropriately, we can ensure that for $a, b \sim \mu_{p,p}(\{0,1\}^J)$, the probability that $L(a) \cap L(b) \neq \emptyset$ is at least $\epsilon/2$. We would like to conclude that there exists some label ℓ which appears in many sets $L(a)$; for an appropriate choice of the parameters, this will imply that g must contain an influential coordinate outside of J, contradicting its definition.

A random choice of labels $\ell(a) \in L(a)$ will satisfy $\Pr_{a,b \sim \mu_{p,p}}[\ell(a) = \ell(b)] = \Omega(\epsilon)$. In particular, we can find a set of vertices V with the same label ℓ such that $\Pr_{a,b \sim \mu_{p,p}}[a, b \in V] = \Omega(\epsilon)\mu_p(V)^2$. Since there are $\Omega(\epsilon)$ edges inside V, small-set expansion implies that $\mu_p(V)$ is large.

2 Small p

For small p, Dinur and Friedgut proved the following result.

Theorem 2. *If $\mathcal{F}$ is an intersecting family then there is an element i such that $\mu_p(\mathcal{F} \setminus S_i) = O(p^2)$ for all $p \leq 1/2$, where $S_i = \{S : S \ni i\}$.*

The proof uses Friedgut's junta theorem. Dinur and Friedgut also proved the following generalization: for every d there is an intersecting $O_d(1)$-junta $\mathcal{G}$ such that $\mu_p(\mathcal{F} \setminus \mathcal{G}) = O(p^d)$.

We start by observing that $\mathcal{F}$ can be assumed to be monotone, since the upset of $\mathcal{F}$ is also intersecting. We can also assume that $p \leq 1/3$, since otherwise the statement is trivial.

Russo's lemma together with Friedgut's junta theorem shows that for some $q \in (1/3, 1/2)$, the family $\mathcal{F}$ is ϵ-approximated by a junta $\mathcal{J}$ depending on a set J of $O_\epsilon(1)$ coordinates; we will choose ϵ later on.

We start by bounding the number of sets in $\mathcal{F}$ which are disjoint from J. Let $\mathcal{F}_0 \subseteq \{0,1\}^{\overline{J}}$ consist of those sets $S \subseteq \overline{J}$ such that $S \in \mathcal{F}$. We have

$$\epsilon \geq \mu_q(\mathcal{F} \setminus \mathcal{J}) \geq \mu_q^J(\overline{\mathcal{J}}) \mu_q^{\overline{J}}(\mathcal{F}_0),$$

showing that

$$q \geq \mu_q(\mathcal{F}) \geq \mu_q(\mathcal{J}) - \epsilon \geq 1 - \frac{\epsilon}{\mu_q(\mathcal{F}_0)} - \epsilon,$$

and so

$$\mu_q(\mathcal{F}_0) \leq \frac{\epsilon}{1-q-\epsilon}.$$

For small enough (but constant) $\epsilon > 0$, this implies that $\mu_q(\mathcal{F}_0) \leq q^2$, and so the Kahn–Kalai isoperimetric theorem shows that $\mu_p(\mathcal{F}_0) \leq p^2$.

Let us now define a J-junta $\mathcal{I}$ by taking all fibers $A \subseteq J$ such that $\Pr_{S\sim\mu_p}[S \in \mathcal{F} \wedge S \cap J = A] > p^2$. The argument above shows that the empty fiber is not in $\mathcal{I}$. By construction, $\mu_p(\mathcal{F}\backslash\mathcal{I}) \leq 2^{|J|}p^2$, and it remains to show that $\mathcal{I}$ is intersecting.

Note first that if $|A| \geq 2$ then since $\Pr[S \cap J = A] \leq p^2$, none of these fibers belong to $\mathcal{I}$. If the $\{i\}$-fiber belongs to $\mathcal{I}$ then $\Pr[S \cap J = \{i\}] \leq p$ implies that $\Pr_{S\sim\mu_p}[S \in \mathcal{F} \mid S \cap J = \{i\}] > p$. If two singleton fibers belong to $\mathcal{I}$ then the corresponding fibers of $\mathcal{F}$ have geometric mean larger than p, contradicting the cross-intersecting Erdős–Ko–Rado theorem. We conclude that all sets in $\mathcal{I}$ (if any) contain some element i.

3 Forbidden intersection

Finally, let us briefly describe, in general terms, the proof in Ellis, Keller and Lifshitz, *Stability for the Complete Intersection Theorem, and the Forbidden Intersection Problem of Erdős and Sós*.

Our goal is to prove that the Ahlswede–Khachatrian theorem holds even if we relax the assumption from t-intersecting to not-exactly-$(t-1)$-intersecting.

The first step is to prove a regularity lemma, showing that *every* family can be approximated by a J-junta $\mathcal{G}$ such that:

(a) $\mu_p(\mathcal{F} \setminus \mathcal{G}) \leq \epsilon$.

(b) For every $A \subseteq J$, if $\mathcal{G}$ contains the A-fiber then the A-fiber of $\mathcal{F}$ has measure at least $\epsilon/2$, and is pseudorandom: setting the value of h variables can only affect the measure by up to δ.

Here ϵ, δ, h are parameters, and p needs to be bounded away from $0, 1$.

Now let $\mathcal{F}$ be a not-exactly-$(t-1)$-intersecting family of large measure. Applying the regularity lemma, we approximate $\mathcal{F}$ by some J-junta $\mathcal{G}$. The pseudorandomness condition implies (via an argument similar to the one in Section 2) that $\mathcal{G}$ is in fact t-intersecting: roughly speaking, if some A, B have intersection smaller than t, then we can find in the corresponding fibers of $\mathcal{F}$ two sets with intersection *exactly* $t-1$, using pseudorandomness to reduce to the case $|A \cap B| = t-1$, and then using bounds on cross-intersecting families.

Since $\mathcal{G}$ is t-intersecting and has large measure, it has to be close to a family optimal for the Ahlswede–Khachatrian theorem (this follows from a stability version of the theorem, proved by a modification of the standard shifting proof). A further bootstrapping argument shows that the measure of $\mathcal{F}$ is at most the measure of an optimal family for the Ahlswede–Khachatrian theorem.

The same kind of regularity lemma is also used by Friedgut and Regev and by Lifshitz to prove their strengthened versions of Dinur–Friedgut–Regev.