The Complexity of Proving Ramsey Principles

Nicola Galesi

Dept. of Computer, Control and Management Engineering "A. Ruberti" (DIAG)
Sapienza Università Roma

Pisa - July 9-11 2023

Workshop Logical methods in Ramsey theory and related topics

Proof systems

Definition (Proof system for L)

Polynomial time onto mapping $F : \{0,1\}^* \rightarrow L$

Our Settings

- $L = TAUT(resp.UNSAT)$
- $F(x) = A$ means: x is a proof (resp. refutation) of A
- F thought as a polynomial time verifier $V(x, A)$ that x is a correct proof of A

Towards $NP \neq coNP$ [Cook Reckhow 74]

Definition (Proof System)

A polynomial time Verifier $V(,)$ s.t.

$$A \in TAUT \equiv \exists x \in \{0,1\}^* : V(x, A)$$

Definition (Polynomially bounded proof system)

A polynomial time Verifier $V(,)$ s.t.

$$A \in TAUT \equiv \exists x \in \{0,1\}^*, |x| \leq |A|^{O(1)} : V(x, A)$$

Theorem (Cook-Reckhow)

There exists a polynomially bounded proof system iff $NP = coNP$

Resolution

$F(x_1 \ldots, x_n)$ an UNSAT CNF formula.
Refutations of F are sequences $A_1, \ldots, A_m$ of clauses, concluding with $A_m = \Box$, formed according to:

Axioms

$$A_i \in F$$

Rule

$$\frac{A \vee x \quad \bar{x} \vee B}{A \vee B}$$

Resolution over k-DNF

Rules

1. The $\wedge$-*introduction rule*

$$\frac{\mathcal{D}_1 \vee \bigwedge_{j\in J_1} l_j \quad \mathcal{D}_2 \vee \bigwedge_{j\in J_2} l_j}{\mathcal{D}_1 \vee \mathcal{D}_2 \vee \bigwedge_{j\in J_1\cup J_2} l_j},$$

   provided that $|J_1 \cup J_2| \leq k$.

2. The *cut (or resolution) rule*

$$\frac{\mathcal{D}_1 \vee \bigvee_{j\in J} l_j \quad \mathcal{D}_2 \vee \bigwedge_{j\in J} \neg l_j}{\mathcal{D}_1 \vee \mathcal{D}_2},$$

Sizes

Let us given an UNSAT CNF $F(x_1, \ldots, x_n)$.
Let $\pi = A_1, \ldots, A_m$ be a resolution refutation of $F(\vec{x})$.

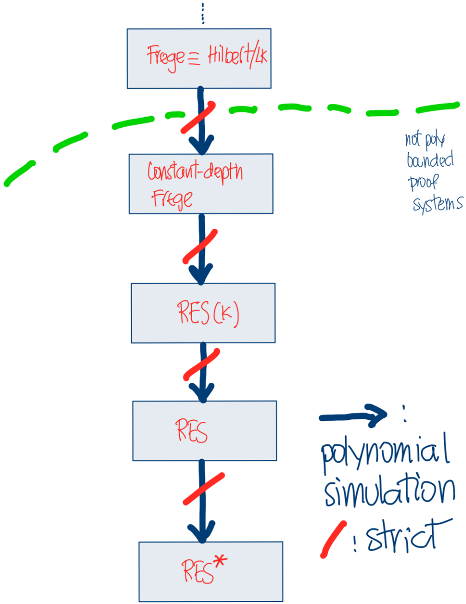
$$Sz(\pi) = m$$

$$Sz(F \vdash) = \min_{F \vdash_\pi \Box} Sz(\pi)$$

Question (Res is not poly bounded)

Exhibit a family of UNSAT CNFs $(F_n)_{n \in \mathbb{N}}$ and prove that $Sz(F_n \vdash) = \Omega(\exp(|F_n|))$ (a superpolynomial suffices)

In search for hard-to-prove formulas



Ramsey Theorem and its propositional formulation

Theorem (Ramsey Theorem)

There exists a number $r(k, s)$ that is the smallest number such that any graph with at least $r(k, s)$ vertices contains either a clique of size k or an independent set of size s.

[Krishnamurty Moll 81]]We are interested in propositional formulation of valid Ramsey statements

$$n \longrightarrow (k)^2_2$$

which expresses Ramsey theorem for $s = k$ and $r_k = r(k, k)$.

Ramsey Theorem and its propositional formulation

$X \subseteq [n]$

$$Cli(X) := \bigwedge_{(ij)\in\binom{X}{2}} E_{ij} \qquad X \text{ is a clique}$$

$$Ind(X) := \bigwedge_{(ij)\in\binom{X}{2}} \neg E_{ij} \qquad X \text{ is an independent set}$$

$$\mathrm{RAM}(n,k) := \bigvee_{X\subseteq[n],|X|=k} Cli(X) \vee \bigvee_{X\subseteq[n],|X|=k} Ind(X) \quad \text{is TAUT for } n \geq r_k$$

$|\mathrm{RAM}(n,k)| = O(n^k)$ it has $\binom{n}{k}$ disjuncts each of size $\binom{k}{2}$

Proof complexity of RAM(n, k) formulas

Theorem (Erdös ...)

$$2^{k/2} < r_k < 4^k$$

What is the complexity of proving RAM(r_k, k) ?

1. Evidence that RAM(r_k, k) is hard for RES (the width is at least $r_k/2$) is and is proved hard (an exponential lower bound for the size required) in RES*. [Krishnamurty Moll 81]
2. Hard (it requires exponential size proofs) to prove in constant depth-Frege [Krajicek 11].

Proof complexity of RAM(n, k) formulas

The problem with RAM(r_k, k) is that we do not know the exact value of r_k, so that we cannot prove upper bounds on proofs of RAM(r_k, k)) to compare the lower bounds with.

Therefore researchers start to study the complexity of proofs of RAM(4^k, k) which is the same as RAM(n, $\frac{\log n}{2}$)

1. RAM(n, $\frac{\log n}{2}$) can be proved with quasipolynomial size proofs in constant-depth Frege [Pudlák 91]
2. RAM(n, $\frac{\log n}{2}$) requires exponential size proofs in RES [Pudlák 12]
3. RAM(n, $\frac{\log n}{2}$) requires exponential size proofs in RES*(log) [Krajicek 01]

Complexity of certifying Ramsey graphs

RAM(n, $\frac{\log n}{2}$) suggests the following definition

Definition ([Lauria Rödl Pudlák Thapen 17])

A graph over n vertices G is c-Ramsey if it has no clique or independent set of size $c \log n$.

Question (Complexity theory point of view)

1. *Efficiency of construction: can these c-Ramsey graphs be constructed in polynomial time ?*
2. *Verification: How hard is to certify that a graph with n vertices is c-Ramsey ?*

Natural certificates that a given graph G is c-Ramsey are proofs/refutations that G is/is not c-Ramsey

k-clique principle

$G = (V, E)$. We want to define a formula

$\mathrm{Clique_k}(G)$ satisfiable iff G contains a k-clique.

$x_{iv} \equiv$ "v is the i-th node in the clique"

$$\mathrm{Clique_k}(G) = \begin{cases} \bigvee_{v \in V} x_{i,v} & i \in [k] & \text{a node in each position} \\ \neg x_{i,v} \vee \neg x_{i,u} & u \neq v \in V, i \in [k] & \text{no two nodes in one position} \\ \neg x_{i,u} \vee \neg x_{j,v} & (u,v) \notin E, i \neq j \in [k] & \text{"no-edges" are not in the clique} \end{cases}$$

Fact

$\mathrm{Clique_k}(G)$ *UNSAT iff* G *does not have a* k*-clique*

Motivation for k-clique: Parameterized Resolution

[Dantchev Martin Szeider 11]: a parameterized Resolution system where assignments are restricted to have weight at most k.

Let $F(x_1, \ldots, x_n)$ be an UNSAT CNF and let $Enc_{n,k}(\vec{x}, \vec{y})$ be a CNF encoding that assignments on $\vec{x}$ with weight more than k are forbidden.

Problem (Proof complexity in ParaRes)

Minimal size of Resolution refutations for $F(\vec{x}) \wedge Enc_{n,k}(\vec{x}, \vec{y})$.
(counting clauses in $Enc_{n,k}(\vec{x}, \vec{y})$ only if used)

First Encoding

$$Enc^1_{n,k}(\vec{x}) := \bigwedge_{i_1,\ldots,i_{k+1}\in[n]} (\bar{x}_{i_1} \vee \ldots \vee \bar{x}_{i_{k+1}})$$

- $F(\vec{x}) + Enc^1_{n,k}(\vec{x})$ have size bounded by $n^{O(k)}$.

Question

- *Does $F(\vec{x}) + Enc^1_{n,k}(\vec{x})$ require refutations of size $n^{\Omega(k)}$?*
- *Or $F(\vec{x}) + Enc^1_{n,k}(\vec{x})$ can be refuted using size $f(k)n^{O(1)}$, for some f ?*

[Beyersdorff Galesi Lauria Razborov 12]: $PHP_n + Enc^1_{n,k}(\vec{x})$ requires RES refutations of size $n^{\Omega(k)}$.

$$\mathrm{PHP}^m_n : \begin{array}{ll} \bigvee_{j=1}^n p_{i,j} & i \in [m] \\ \overline{p}_{i,j} \vee \overline{p}_{i',j} & i, \neq i' \in [m], j \in [n] \end{array}$$

Second Enconding

Uses variable s_{ij}, for $i \in [k], j \in [n]$ and encode an injective mapping from $[k]$ to $[n]$

$$Enc^2_{n,k}(\vec{x}, \vec{s}) := \begin{cases} \bar{x}_i \vee \bigvee_{j\in[k]} p_{ij} & i \in [n] \\ \bar{p}_{ij} \vee \bar{p}_{i'j} & i \neq i' \in [n], j \in [k] \end{cases}$$

[Dantchev Martin Szeider 11]: $PHP_n + Enc^2_{n,k}(\vec{x})$ has proof of size $O(kn^2)2^k$.

Problem

Prove $n^{\Omega(k)}$ lower bounds in $Res{+}Enc^2_{n,k}(\vec{x})$

Problem

$Enc^2(\vec{x}, \vec{p})$ is built-in for $\mathrm{Clique}^n_k(G)$. *Prove there are no* RES *proofs of size $n^{O(1)} f(k)$ when G does not contain a k-clique*

k-Clique

Given a graph $G = (V, E)$ and a parameter k, $\mathrm{Clique}^n_k(G)$ is:

$$\begin{array}{ll} \bigvee_{v \in V} x_{i,v} & i \in [k] \\ \neg x_{i,u} \vee \neg x_{j,v} & i, j \in [k],\ i \neq j \text{ and } \{u, v\} \notin E \\ \neg x_{i,u} \vee \neg x_{i,v} & u \neq v \in V. \end{array}$$

$x_{i,v}$ means vertex v is the ith member of the clique.

Property

$\mathrm{Clique}^n_k(G)$ *is satisfiable if and only if the graph G has a clique of size k.*

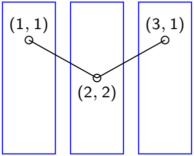
Problem (Open)

$Enc^2(\vec{x}, \vec{p})$ is built-in for $\mathrm{Clique}^n_k(G, k)$. *Prove there are no* RES *proofs of size $n^{O(1)} f(k)$ when G does not contain a k-clique*

k-Clique Principle: Simplified version

- G formed from k blocks V_b of n nodes each: $G = (\bigcup_{b \in [k]} V_b, E)$
- Variables $v_{i,q}$ with $i \in [k], a \in [n]$, with clauses

$$\mathrm{Clique}^{\mathrm{n}}_{\mathrm{k}}(G) = \begin{cases} \neg v_{i,a} \vee \neg v_{j,b} & ((i,a),(j,b)) \notin E \\ \bigvee_{a \in [n]} v_{i,a} & i \in [k] \end{cases}$$



Fact

$\mathrm{Clique}^{\mathrm{n}}_{\mathrm{k}}(G)$ *UNSAT iff G does not have a k-clique*

The case of the complete $(k-1)$-partite graph

The canonical graph without a k-clique is C_n the complete $(k-1)$-partite graph.

Theorem (Beyersdorff Galesi Lauria 12)

$\mathrm{Clique}^{\mathrm{n}}_{\mathrm{k}}(C_n)$ *requires treelike* RES* *of size* $n^{\Omega(k)}$ *but have* $O(2^k k^2 n^2)$ RES *refutations.*

Upper Bound Proof Idea. In $O(k^2 n^2)$ proof steps reduce to PHP^k_{k-1} using the fact that proofs are trying to exclude the presence of a k-clique into the complete $(k-1)$-partite graph. Use the mapping

$$p_{i,h} \longleftrightarrow \bigvee_{v \in V_h} x_{i,v}.$$

Then use that PHP^k_{k-1} has Resolution refutations of size $O(2^k)$

Prover Delayer Games

Problem (Search(F, α))

Given UNSAT CNF $F(x_1, \dots x_n)$ and a assignment $\vec{\alpha} \mapsto \vec{x}$, find the clause $C \in F$ such that C false under α.

[Pudlák Impagliazzo 00, Beyersdorff Galesi Lauria 12]: Two persons (Prover, Delayer) game solving Search(F, α).

Game: In each round, Prover places a variable x_i, and Delayer either chooses a value 0 or 1 for x_i or leaves decision to the Prover. In this last case the Delayer gets 1 points. The assignment is recorded in α.

Stop: first round α falsifies a clause in F

Cost: number of points earned by Delayer

The Asymmetric Case

Game: In each round, the number of points Delayer earns depends on the variable x_i, the assignment α constructed so far in the game, and two functions c_0 and c_1.

$$\begin{array}{ll} 0 & \text{if Delayer chooses the value,} \\ \log c_0(x_i, \alpha) & \text{if Prover sets } x_i \text{ to 0, and} \\ \log c_1(x_i, \alpha) & \text{if Prover sets } x_i \text{ to 1.} \end{array}$$

c_0 and c_1 are non negative and are chosen in such a way that for each variable x and assignment α

$$\frac{1}{c_0(x, \alpha)} + \frac{1}{c_1(x, \alpha)} = 1 \tag{1}$$

Delayer Strategies give Lower Bounds

Theorem (Pudlák Impagliazzo 00, Beyersdorff Galesi Lauria 12)

If $(F_n)_{n \in \mathbb{N}}$ have treelike Resolution refutations of size S, then for each (c_0, c_1)-game played on (F_n) there is a Prover strategy leaving at most $\log S$ points to the Delayer.

Theorem (Beyersdorff Galesi Lauria 12)

There are c_0 and c_1 s.t. in any APD-game on $\mathrm{Clique}(C_n, k)$, *Delayer earns $(k-1)\log n$ points.*

proof

The set of vertices of the graph C_n is partitioned into the sets $V_1, \ldots, V_{k-1}$ of size n each.

Delayer strategy objective: at the end of the game the partial assignment always has $k-1$ indexes assigned to specific vertices in different blocks.

Score function: on each block Delayer scores exactly $\log n$ points.

Conclusion:Delayer always wins $\geq (k-1)\log n$ points

Delayer info: keeps $k-1$ sets $Z_j \subseteq V_j, j \in [k-1]$ which represent the excluded vertices in each block.

Delayer Strategy: Let α current ass and $x_{i,v}$ for $v \in V_j$ the variable queried.

Then Delayer sets x_{iv} to:

1. 0 if $\alpha(x_{iw}) = 1$ for some $w \neq v$;
2. 0 if $\alpha(x_{lw}) = 1$ for some $l \in [k] \setminus \{i\}$ and some $w \in V_j$;
3. 0 if $v \in Z_j$;
4. 1 if $v \notin Z_j$ and $Z_j = V_j \setminus \{v\}$;
5. and leave decision to Prover otherwise.

Delayer Update of Z_j's :

- If Delayer sets x_{iv}, then Z_j remains unaltered.
- if Prover decides 0 then $Z_j := Z_j \cup \{v\}$.
- If Prover decides 1, then $Z_j := V_j \setminus \{v\}$.

Score Function: Measure the information of the degree of freedom of Delayer to answer 0 to the variable queried in the block j.

- $c_1 = |V_j| - |Z_j|$.
- $c_0 = \frac{|V_j|-|Z_j|}{|V_j|-|Z_j|-1}$

$(k-1)$ indices at the end: by contradiction assume no index in V_j. Consider the last moment in the game in which $x_{iv} = 0$ has been assigned for some $v \in V_j$. All variables x_{iu} for $u \in V_j \setminus \{v\}$ have been queried before and set to 0. According to the Delayer strategy, either $x_{iu} = 0$ was set by Delayer by rule 3, or $x_{i,u} = 0$ was decided by Prover. In both cases $u \in Z_j$ and therefore $Z_j = V_j \setminus \{v\}$. But then Delayer would assign x_{iv} to 1 according to item 4 of her strategy, a contradiction.

Number of points in each block: Fix a block i. Exactly one variable x_{iv} is set to one. Let us say that $|Z_i| = z$ right before that decision. Until that moment $|Z_i|$ increases one by one every time Delayer scores some point on Prover deciding for some x_{iu} to be zero. Delayer scores

$$\sum_{t=0}^{z-1} \log \frac{|V_i| - t}{|V_i| - t - 1} = \log |V_i| - \log(|V_i| - z).$$

Delayer chooses to set $x_{iv} = 1$ if and only if $z = |V_i| - 1$, otherwise the Prover chooses which gives $\log(|V_i| - z)$ points to Delayer. In both cases Delayer scores $\log |V_i|$ points on block i. Thus in the end, Delayer gets exactly $(k-1) \log n$ points.

Finding graphs hard to certify to be c-Ramsey in RES

Distribution of graphs $\mathcal{G}_{k,\epsilon}$:

Consider $V = kn$ vertices divided into k blocks of n vertices: $V_1, V_2, \ldots, V_k$. $0 < \epsilon < 1$.

- $(u, v) \in E$ with $u \in V_i$, $v \in V_j$ and $i < j$, the edge $\{u, v\}$ is present with probability $p = n^{-(1+\epsilon)\frac{2}{k-1}}$.

Slight variation of the Erdős-Rényi model $G(n, p)$.

Fact

It is known that k-cliques appear at the threshold probability $p^* = n^{-\frac{2}{k-1}}$. If $p < p^*$, then with high probability in $G \sim \mathcal{G}_{k,\epsilon}$ there is no k-clique;

All graphs in $\mathcal{G}_{k,\epsilon}$ are properly colorable with k colors.

Random graphs make hard $\text{Clique}^n_k(G)$ for RES*

Simplified $\text{Clique}^n_k(G)$: In a k-colorable graph G with color classes $V_1, \ldots, V_k$ a k-clique contains exactly one vertex per color class. In this case we can simplify formula $\text{Clique}^n_k(G)$ by setting $x_{i,v} = 0$ for every $i \in [k]$ and $v \in V_j$ such that $i \neq j$. Essentially we are forcing the ith vertex in the clique to be in the ith block.

$$\text{Clique}^n_k(G) := \begin{cases} \bigvee_{v \in V_i} x_v & i \in [k] \\ \neg x_u \vee \neg x_v & \{u, v\} \notin E(G). \end{cases}$$

Theorem (Beyersdorff Galesi Lauria 12)

Let $0 < \epsilon < 1$. For a random graph $G \sim \mathcal{G}_{k,\epsilon}$, then w.h.p. the smallest RES* *refutations of* $\text{Clique}^n_k(G)$ *has size* $n^{\Omega(k(1-\epsilon))}$.

Complexity of $\text{Clique}^n_k(G)$ in RES: a challenge

Problem (Difficult Open Problem)

Prove significative lower bounds for refutations of $\text{Clique}^n_k(G)$ *in* RES *when* $G \sim \mathcal{G}_{k,\epsilon}$.

RES ⇢ r.o.RES ⇢ RES*

Theorem ([Atserias Bonacina de Rezende Lauria Nördstrom Razborov 21])

If $G \sim \mathcal{G}_{k,\epsilon}$, *then with high probability* $\text{Clique}^n_k(G)$ *require r.o.*RES *refutations of size* $n^{\Omega(k)}$.

The Binary Clique Principle: $\text{Bin-Clique}^n_k(G)$

- (Bit-)Variables: $\omega_{i,j}$, for $i \in [k], j \in [\log n]$
- Notation:

$$\omega_{i,j}^{a_j} = \begin{cases} \omega_{i,j} & \text{if } a_j = 1 \\ \neg\omega_{i,j} & \text{if } a_j = 0 \end{cases}$$

$$v_{i,j} \equiv (\omega_{i,1}^{a_1} \wedge \ldots \wedge \omega_{i,\log n}^{a_{\log n}}), \text{ where } (j)_2 = \vec{a}$$

$$\text{Bin-Clique}^n_k(G) = \bigwedge_{((i,a),(j,b)) \notin E} \left((\omega_{i,1}^{1-a_1} \vee \ldots \vee \omega_{i,\log n}^{1-a_{\log n}}) \vee (\omega_{j,1}^{1-b_1} \vee \ldots \vee \omega_{j,\log n}^{1-b_{\log n}}) \right)$$

The complexity of Bin-Clique$^n_k(G)$ in RES

Binary versions of combinatorial principles:

- preserve the combinatorial hardness of the unary principle;
- are less exposed to details of the encoding when attacked with a lower bound technique;
- give significative lower bounds.

Theorem ([Lauria Pudlák Rödl Thapen 17])

If $G \sim \mathcal{G}_{k,\epsilon}$, *then with high probability* Bin-Clique$^n_k(G)$ *requires* RES *refutations of size* $n^{\Omega(k)}$.

$Res(k)$: Resolution with k-conjunctions

A refutation system for k- DNFs. Disjunctions of k-terms.

Rules

1. $\wedge$-*introduction* is

$$\frac{\mathcal{D}_1 \vee \bigwedge_{j \in J_1} l_j \quad \mathcal{D}_2 \vee \bigwedge_{j \in J_2} l_j}{\mathcal{D}_1 \vee \mathcal{D}_2 \vee \bigwedge_{j \in J_1 \cup J_2} l_j},$$

   provided that $|J_1 \cup J_2| \leq s$.

2. *cut* is

$$\frac{\mathcal{D}_1 \vee \bigvee_{j \in J} l_j \quad \mathcal{D}_2 \vee \bigwedge_{j \in J} \neg l_j}{\mathcal{D}_1 \vee \mathcal{D}_2},$$

3. *weakening* are

$$\frac{\mathcal{D}}{\mathcal{D} \vee \bigwedge_{j \in J} l_j} \quad \text{and} \quad \frac{\mathcal{D} \vee \bigwedge_{j \in J_1 \cup J_2} l_j}{\mathcal{D} \vee \bigwedge_{j \in J_1} l_j},$$

   provided that $|J| \leq s$.

Unifying Unary and Binary case for the clique principle

Lemma ([Dantchev Galesi Martin 18])

Let $G \sim \mathcal{G}^{k,\epsilon}$ and suppose there are RES *refutations of* $\mathsf{Clique}^{\mathsf{n}}_{\mathsf{k}}(G)$ *of size S . Then there are* RES$(\log n)$ *refutations of* $\mathsf{Bin\text{-}Clique}^{\mathsf{n}}_{\mathsf{k}}(G)$ *of size S.*

Corollary

Prove $n^{\Omega(k)}$ lower bounds in RES$(\log n)$ *for* $\mathsf{Bin\text{-}Clique}^{\mathsf{n}}_{\mathsf{k}}(G)$ *to catch $n^{\Omega(k)}$ lower bounds in* RES *for* $\mathsf{Clique}^{\mathsf{n}}_{\mathsf{k}}(G)$

Theorem ([Dantchev Galesi Ghani Martin *To appear*])

If $G \sim \mathcal{G}_{k,\epsilon}$, then $\mathsf{Bin\text{-}Clique}^{\mathsf{n}}_{\mathsf{k}}(G)$ *require* RES$(\sqrt{\log \log n})$ *refutations of size $n^{\Omega(k)}$.*

Lower Bound Proof for RES($\log\log n$)

Main Tools (for Binary Principles):

1. *Covering Number* on s-DNFs [1]
   - RES(s) proofs with small CN efficiently simulated in RES($s-1$)
   - *Bottlenecks*
2. *(Random) restrictions* for binary principles
3. *Hardness properties* of Bin-Clique$^n_k(G)$, when $G \sim \mathcal{G}(n, p)$ [2]
4. Induction on s.
   - Base Case: known hardness on RES(1) [3].

[1]=[Segerlind Buss Impagliazzo 04]
[2]=[Beyersdorff Galesi Lauria 13]
[3]=[Lauria Pudlák Rödl Thapen 17]

Covering number of a RES(s) proof

A *covering set* for a s-DNF $\mathcal{F}$ is a set of literals L such that each term of $\mathcal{F}$ has at least a literal in L.

The *covering number* $cv(\mathcal{F})$ of a s-DNF $\mathcal{F}$ is the minimal size of a covering set for $\mathcal{D}$.

$$CN(\pi) = \max_{\mathcal{F}\in\pi} c(\mathcal{F})$$

Small covering number vs simulations

Lemma (Simulation Lemma)

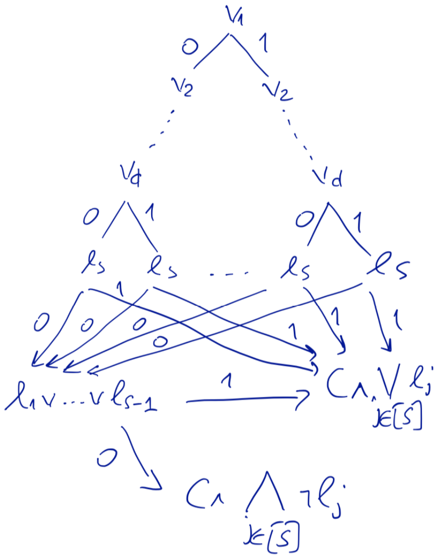
If F has a refutation π in $\mathrm{RES}(s)$ *with* $CN(\pi) < d$, *then F has a* $\mathrm{RES}(s-1)$ *refutation of size at most* $2^{d+2}N$.

Put π upside-down. Get a restricted branching s-program whose nodes are labelled by s-CNFs and at each node some s-disjunction $\bigvee_{j\in[s]} l_j$ is queried.

<u>Example</u>

$$\begin{array}{ccc} & \vdots & \\ & \mathcal{C} & \\ & ?\bigvee_{j\in[s]} l_j & \\ 1 \swarrow & & \searrow 0 \\ \mathcal{C}\wedge\bigvee_{j\in[s]} l_j & & \mathcal{C}\wedge\bigwedge_{j\in[s]}\neg l_j \end{array} \qquad (2)$$

Let $cv(\mathcal{C}) < d$, witnessed by variable set $\{v_1, \ldots, v_d\}$.



Bottlenecks in RES(s)

A c-bottleneck in a RES(s) proof is a s-DNF F whose $cv(F) \geq c$.
$c(s)$ is the *bottleneck number* at RES(s).

Fact (Independence)

If $c = rs$, $r \geq 1$ and $cv(F) \geq c$, then in F it is always possible to find r pairwise disjoint s-tuples of literals $T_1 = (\ell_1^1, \ldots, \ell_1^s), \ldots, T_r = (\ell_r^1, \ldots, \ell_r^s)$ such that the $\bigwedge T_i$'s are terms of F.

Restrictions

A *s-restriction* assigns $\lfloor \frac{\log n}{2^{s+1}} \rfloor$ bit-variables $\omega_{i,j}$ in each block $i \in [k]$.

Fact

if σ and τ are (disjoint) s-restrictions, then $\sigma\tau$ is a $(s-1)$-restriction

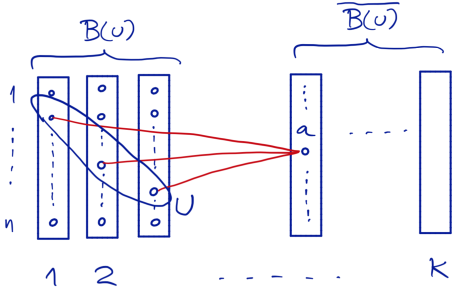
A *random s-restriction* for $\text{Bin-Clique}^{\text{n}}_{\text{k}}(G)$ is an s-restriction obtained by choosing independently in each block i, $\lfloor \frac{\log n}{2^{s+1}} \rfloor$ variables among $\omega_{i,1}, \ldots, \omega_{i,\log n}$, and setting these uniformly at random to 0 or 1.

$G = (\bigcup_{b \in [k]} V_b, E)$ and $0 < \alpha < 1$. U is *α-transversal* if:

1. $|U| \leq \alpha k$, and
2. for all $b \in [k]$, $|V_b \cap U| \leq 1$.

Let $B(U) \subseteq [k]$ be the set of blocks mentioned in U, and $\overline{B(U)} = [k] \setminus B(U)$.

U is *extendible* in a block $b \in \overline{B(U)}$ if there exists a vertex $a \in V_b$ which is a *common neighbour of all nodes in U*.



A restriction σ is *consistent* with $v = (i, a)$ if for all $j \in [\log n]$, $\sigma(\omega_{i,j})$ is either a_j or not assigned (i.e. assigns the right bit or can do it in the future)

Definition

Let $0 < \alpha, \beta < 1$. A α-transversal U is β-extendible, if for all β-restriction σ, there is a node v^b in each block $b \in \overline{B(U)}$, such that σ is consistent with v^b.

Lemma (Extension Lemma, similar to [1])

Let $0 < \epsilon < 1$, let $k \leq \log n$. Let $1 > \alpha > 0$ and $1 > \beta > 0$ such that $1 - \beta > \alpha(2 + \epsilon)$. Let $G \sim \mathcal{G}(n, p)$. With high probability both properties hold:

1. *all α-transversal sets U are β-extendible;*
2. *G does not have a k-clique.*

[1]=[Beyersodrff Galesi Lauria 13]

Idea of the proof

Property (Clique(G, s, k))

For any s-restriction ρ, there are no Res(s) refutations of $\text{Bin-Clique}^{\text{n}}_{\text{k}}(G)\restriction_\rho$ *of size less than* $n^{\frac{\delta(k-1)}{\mathrm{d}(s)}}$.

Theorem

If Clique(G, s, k) *holds, then there are no* RES(s) *proofs of* $\text{Bin-Clique}^{\text{n}}_{\text{k}}(G)$ *with size* $n^{\frac{\delta(k-1)}{\mathrm{d}(s)}}$.

Corollary

Let $1 < s = o(\sqrt{\log\log n})$. *There exists a graph* G *such that* RES(s) *refutations of* $\text{Bin-Clique}^{\text{n}}_{\text{k}}(G)$ *are* $n^{\Omega(k)}$.

By Extension Lemma there exists a $G \sim \mathcal{G}_{k,\epsilon}$ with the extension properties.

Lemma

Clique($G, 1, k$) *holds. (use [1])*

[1]=[Lauria Pudlák Rödl Thapen 17]

Steps of the proof

Lemma

$\text{Clique}(G, s-1, k) \Rightarrow \text{Clique}(G, s, k)$ *as long as* $s = o(\sqrt{\log\log n})$.

We prove that $\neg\,\text{Clique}(G, s, k) \Rightarrow \neg\,\text{Clique}(G, s-1, k)$. Let $L(s) = n^{\frac{\delta(k-1)}{\mathrm{d}(s)}}$.

1. Since $\neg\,\text{Clique}(G, s, k)$, then $\exists$ a s-restriction ρ and π a proof of $\text{Bin-Clique}^{\mathrm{n}}_{\mathrm{k}}(G)\restriction_\rho$, such that $|\pi| < L(s)$.
2. Let $c = c(s)$ be the bottleneck number and $r = cs$
3. σ be a s-random restriction on $\text{Bin-Clique}^{\mathrm{n}}_{\mathrm{k}}(G)\restriction_\rho$.
4. $\Pr[\text{bottleneck } F \text{ survives in } \pi\restriction_\sigma] \leq e^{-\frac{r}{\mathrm{p}(s)}}$. Use *Independence Property*.
5. $\Pr[CN(\pi\restriction_\sigma) \geq c] < 1$. *Union bound*.
6. Define $\tau = \sigma\rho$ and apply *Simulation Lemma* to $\pi\restriction_\sigma$. We get a (s-1)-restriction τ and a $\leq L(s)2^{c+2}$ size proof in $Res(s-1)$ of $\text{Bin-Clique}^{\mathrm{n}}_{\mathrm{k}}(G)\restriction_\tau$. If $L(s)2^{c+2} < L(s-1)$, this is $\neg\,\text{Clique}(G, s-1, k)$.
7. knowing $\mathrm{p}(s)$, define $\mathrm{d}(s)$ and $c(s)$ in such a way to force $L(s)2^{c+2} < L(s-1)$ and union bound to work.