Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

# Pseudorandom Generators in Proof Complexity

Based on the 2001 article of the same name by Alekhnovich, Ben-Sasson, Razborov, & Wigderson

> N. Hayek & B. Kyle

April 1st 2025

## Pseudorandom Generators

Pseudorandom Generators in Proof Complexity

> N. Hayek & B. Kyle

Definition (Generator)

A mapping  $G_n: \{0,1\}^n \to \{0,1\}^m$  is called a generator.

## Pseudorandom Generators

Pseudorandom Generators in Proof Complexity

N. Hayek a B. Kyle

#### Definition (Generator)

A mapping  $G_n: \{0,1\}^n \to \{0,1\}^m$  is called a generator.

#### Definition (Pseudorandomness)

A deterministic generator is *pseudorandom* if no efficient algorithm can differentiate between the probability distributions of  $G_n(\vec{x})$  and  $\vec{y}$ , where  $\vec{x}$  and  $\vec{y}$  are truly random.

## **Pseudorandom Generators**

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

#### Definition (Generator)

A mapping  $G_n: \{0,1\}^n \to \{0,1\}^m$  is called a generator.

#### Definition (Pseudorandomness)

A deterministic generator is *pseudorandom* if no efficient algorithm can differentiate between the probability distributions of  $G_n(\vec{x})$  and  $\vec{y}$ , where  $\vec{x}$  and  $\vec{y}$  are truly random.

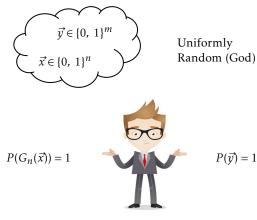
#### Definition (Hardness)

Let *P* be a proof system. Let  $b \in \{0,1\}^m$  be arbitrary. A generator is *hard* for *P* if *P* cannot prove in polynomial size that  $b \notin \text{Im}(G_n)$ .

# Pseudorandom $G_n$

Pseudorandom Generators in Proof Complexity

> N. Hayek & B. Kyle



(I'm a polynomial algorithm)

# Hardness of $G_n$

Pseudorandom Generators in Proof Complexity

> N. Hayek & B. Kyle



$$\exists \ \overrightarrow{x} : \ P(G_n(\overrightarrow{x})) = b$$

(I'm a polynomial P - proof)

## Motivation 1

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle ▶ If no proof system can deduce the **most basic property** of  $G_n$  efficiently (notably, what is in its image), then it certainly can't **distinguish** between the image of  $G_n$  and a proper random distribution efficiently.

Hard in Especially Strong System  $P \stackrel{?}{\Longrightarrow}$  Pseudorandom

# Motivation 1

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle ▶ If no proof system can deduce the **most basic property** of  $G_n$  efficiently (notably, what is in its image), then it certainly can't **distinguish** between the image of  $G_n$  and a proper random distribution efficiently.

Hard for Especially Strong System  $P \stackrel{?}{\Longrightarrow}$  Pseudorandom

▶ Conversely, if  $G_n$  is pseudorandom, it is hard for most proof systems (if there existed an algorithm that could efficiently prove  $b \notin \text{Im}(G_n)$ , we could use this to distinguish  $G_n$  from random, and break the generator).

Pseudorandom  $\stackrel{?}{\Longrightarrow}$  Hard for all P

# Motivation 2

Pseudorandom Generators in Proof Complexity

N. Hayek B. Kyle

▶ The existence of hard generators for P (in particular, when m > n) provides lower bounds for a class of tautologies in P.

## **Tseitin Generators**

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Example

Tseitin tautologies provide a good context for constructing hard generators. Let G be a connected graph, with |E|=n and |V|=m. Let  $\vec{x}$  be a vector of variables on E. Enumerate V arbitrarily  $v_1,...,v_m$ . Then we have the generator

$$T_G: \{0,1\}^n \to \{0,1\}^m \quad \vec{x} \mapsto \begin{bmatrix} \bigoplus_{e \ni v_1} x_e \\ \vdots \\ \bigoplus_{e \ni v_m} x_e \end{bmatrix}$$

where  $\oplus$  is the typical XOR (i.e.  $\equiv_2$  1). One can show that  $\vec{\sigma} \in \{0,1\}^m$  is *not* in the image of  $T_G \iff \bigoplus_{i=1}^m \sigma_i = 1$  (i.e. "odd"). When is it hard, then, for a proof system to show  $\vec{\sigma} \notin \text{Im}(T_G)$ ?

# $T_G$ hard $\equiv$ Tseitin hard to refute

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Example (cont.)

Fix  $\vec{\sigma}$  which is *odd*. Then

$$T_G(\vec{x}) = \vec{\sigma}$$

*cannot* happen, since  $\vec{\sigma} \notin \text{Im}(T_G)$ . In other words, there is no satisfying assignment to  $\vec{x}$ :  $T_G(\vec{x}) = \vec{\sigma}$ , and so the tautologies

$$\bigoplus_{e\ni v_1} x_e = \sigma_1$$

$$\vdots$$

$$\bigoplus_{e\ni v_m} x_e = \sigma_m$$

are unsatisfiable.

We know of some good (exponential) lower bounds on refuting the Tseitin tautologies, e.g. in resolution. We conclude that, in these systems,  $T_G$  is a hard generator, since the choice of mapping  $\vec{\sigma}$  is arbitrary when proving such lower bounds.

# Main Example: Nisan-Wigderson Generators

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Let  $g_i : \vec{x} \to \{0,1\}$  be a function on *n*-dimensional vector of variables  $\langle x_1,...,x_n \rangle$ . We call each  $g_i$  a **base function**.

# Main Example: Nisan-Wigderson Generators

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Let  $g_i : \vec{x} \to \{0,1\}$  be a function on *n*-dimensional vector of variables  $\langle x_1,...,x_n \rangle$ . We call each  $g_i$  a **base function**.

(Caveat: Fix a binary matrix A of dimensions  $m \times n$ . We impose that  $g_i$  depend only on variables  $x_i$  for which the j-th entry in the i-th row is 1.)

# Main Example: Nisan-Wigderson Generators

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Let  $g_i : \vec{x} \to \{0,1\}$  be a function on *n*-dimensional vector of variables  $\langle x_1,...,x_n \rangle$ . We call each  $g_i$  a **base function**.

(Caveat: Fix a binary matrix A of dimensions  $m \times n$ . We impose that  $g_i$  depend only on variables  $x_j$  for which the j-th entry in the i-th row is 1.)

Then define

$$G_n(\vec{x}) = \left\langle g_1(\vec{x}), ..., g_m(\vec{x}) \right\rangle$$

These are believed to be pseudorandom in certain contexts.

Subject to conditions on A and  $g_i$ , we will show that these are hard for some standard proof systems.

# Propositionalizing NW Generators

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Definition (Matrix Restriction)

$$J_i(A) = \{j \in [n] : a_{ij} = 1\} \qquad X_i(A) = \{x_j : j \in J_i(A)\}$$

Above, we related the hardness of  $T_n$  to the Tseitin tautologies. We are interested now in the hardness of NW generators, i.e. refuting the tautologies

$$\begin{cases} g_1(\vec{x}) = 1 \\ \vdots & \text{Vars}(g_i) \subseteq X_i(A) \\ g_m(\vec{x}) = 1 \end{cases}$$
 (1)

in some common proof systems.

# Propositionalizing NW Generators

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Definition (Matrix Restriction)

$$J_i(A) = \{j \in [n]: a_{ij} = 1\} \qquad X_i(A) = \{x_j: j \in J_i(A)\}$$

Above, we related the hardness of  $T_n$  to the Tseitin tautologies. We are interested now in the hardness of NW generators, i.e. refuting the tautologies

$$\begin{cases} g_1(\vec{x}) = 1 \\ \vdots \\ g_m(\vec{x}) = 1 \end{cases}$$
 Vars $(g_i) \subseteq X_i(A)$  (1)

in some common proof systems.

(Caveat: we later impose hardness conditions on  $g_i$ . By allowing these conditions to be satisfied by  $g_i \iff$  they are satisfied by  $\overline{g_i}$ , it is sufficient to consider (1), i.e.  $b = \langle 1, ..., 1 \rangle$ , for the sake of refuting  $\vec{b} \notin \text{Im}(G_n)$  for any  $\vec{b}$ .)

# **Background Definitions**

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle We're interested in propositionalizing (1). In our paper, circuit-based, linear, and functional encodings are provided. We will focus on the latter.

#### Definition (Extension Variable)

Fix  $i \in [m]$ . Let f be a boolean function for which  $Vars(f) \subseteq X_i(A)$ . Then  $y_f$  is an *extension variable* for f.

Denote by  $Vars(A) = \{y_f : \exists i \in [m] : Vars(f) \subseteq X_i(A)\}$  all possible extension variables (and hence functions) on the variables  $X_i(A) : i \in [m]$ .

## Example

Let 
$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$
. Then Vars( $A$ ) is in correspondence with all boolean functions on a subset of  $\{x_1, x_3\}$  or  $\{x_2\}$ .

# More definitions...

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Definition (Vars(A) $\rightarrow \vec{x}$ )

Let  $C = y_{f_1}^{\varepsilon_1} \lor \cdots \lor y_{f_k}^{\varepsilon_k}$  be a clause on Vars(A). Then

$$\|C\|:=f_1^{\varepsilon_1}\vee\cdots\vee f_k^{\varepsilon_k}$$

is a boolean function on the variables  $\vec{x}$ .

## Example

With *A* as above, let  $f_1 = x_1 \wedge \overline{x_3}$  and  $f_2 = x_2$ . Let  $C = \overline{y_{f_1}} \vee y_{f_2}$ . Then

$$||C|| = \overline{x_1 \wedge \overline{x_3}} \vee x_2 = \overline{x_1} \vee x_3 \vee x_2$$

# The Functional Encoding

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Let  $g_1,...,g_m$  be functions on  $\vec{x}$  which constitute a generator  $G_n$ . Note that  $Vars(g_i) \subseteq X_i(A)$ . We encode (1) as follows:

## Definition (Functional Encoding of NW Generator Hardness)

Fix *A*. Let  $\tau(A, G_n)$  denote the collection of clauses of the form  $C = y_{f_1}^{\varepsilon_1} \vee \cdots \vee y_{f_k}^{\varepsilon_k}$  for which

$$Vars(f_i) \subseteq X_i(A)$$
  $i = 1,...,k$  and  $g_i \models ||C||$ 

 $\tau(A, G_n)$  is the functional encoding of (1).

# The Functional Encoding is Correct

Pseudorandom Generators in Proof Complexity

N. Hayek B. Kyle

## Theorem ( $\tau(A, G_n)$ Corresponds with (1))

 $\tau(A, G_n)$  is satisfiable  $\iff$  (1) has a mapping that satisfies it.

#### Proof.

 $(\Leftarrow)$  Let  $\vec{x_0}$  be a solution to the system

$$\begin{cases} g_1(\vec{x_0}) = 1 \\ \vdots \\ g_m(\vec{x_0}) = 1 \end{cases}$$

Consider  $f_y \in \text{Vars}(A)$ . Let  $\rho : \text{Vars}(A) \to \{0,1\}$  be the truth assignment  $y_f \mapsto f(\vec{x_0})$ . Let  $C \in \tau(A, G_n)$ , i.e.

$$C = y_{f_1}^{\varepsilon_1} \lor \cdots \lor y_{f_k}^{\varepsilon_k}$$
 with  $Vars(f_j) \subseteq X_i(A) \forall j$ , some  $i$ 

Since  $g_i \models ||C||$ , and  $g_i(\vec{x_0}) = 1$ , we have  $||C|| = f_1^{\varepsilon_1} \lor \cdots \lor f_k^{\varepsilon_k} = 1$ , so  $\exists i : f_i^{\varepsilon_i}(\vec{x_0}) = 1$ . Therefore,  $\rho$  will satisfy  $y_{f_i}^{\varepsilon_i}$ , and hence C.

# The Functional Encoding is Correct

Pseudorandom Generators in Proof Complexity

> N. Hayek & B. Kyle

#### Proof.

 $(\Longrightarrow)$  Let  $\rho$  be an assignment on Vars(A) satisfying  $\tau(A, G_n)$ . Define

$$\vec{x_0} = \begin{bmatrix} y_{x_1} \\ \vdots \\ y_{x_n} \end{bmatrix}$$

Note that the formula  $x_i$  belong to Vars(A) so long as we have no zero columns in A (we will impose this later). One can show by induction that  $\rho(y_f) = f(\vec{x_0})$  as above. Since  $Vars(g_i) \subseteq X_i(A)$  and clearly  $g_i \models g_i$ , we have  $g_i \in \tau(A, G_n)$  as a clause.  $\rho$  is satisfying for  $\tau(A, G_n)$ , so  $\rho(g_i) = 1$  (as a bit assignment). But then  $\rho(g_i) = g_i(\vec{x_0}) = 1$ , as desired.

# Main Result: $\tau(A, G_n)$ Width Bounds in Resolution

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Without yet defining  $r,s,c \in \mathbb{R}$ , (r,s,c)-expanders, or  $\ell$ -robustness, we state the following theorem:

## Theorem (Width of $\tau(A, G_n)$ in Resolution)

Let  $A \in M_{m \times n}(\{0,1\})$  be an (r,s,c)-expander, and let  $g_i$  be  $\ell$ -robust for i=1,...,m. Let  $c+\ell \geq s+1$ . Then

$$w_{Res}(\tau(A,G_n)) > \frac{r(c+\ell-s)}{2\ell} = \Omega(r)$$

The foreign terms constitute the "hardness conditions" on A and  $g_i$ . We will define the following

- 1 (r,s,c)-expanders: these are sparse matrices which generalize well-connectedness for graphs. In such a way, tight groupings of variables between base functions are discouraged, preventing localized contradictions.
- **2**  $\ell$ -robust functions  $g_i$  resist partial assignments.

# Hardness of $A \leftrightarrow (r, s, c)$ -expanders

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Definition ((r, s, c)-expanders)

Let  $A \in M_{m \times n}(\{0,1\})$ . For a set of rows  $I \subseteq [m]$ , let  $\partial_A(I)$  (called the *boundary* of I) denote all columns which, when restricted to I, contain one "1."

Then, *A* is called an (r,s,c)-expander if  $|J_i(A)| \le s$  and, for all choices *I* as above,  $|I| \le r \implies |\partial_A(I)| \ge c|I|$ .

What does this say: first, the number of 1s in any given row is bounded (by s). This allows sparseness.

# Hardness of $A \leftrightarrow (r, s, c)$ -expanders

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Definition ((r, s, c)-expanders)

Let  $A \in M_{m \times n}(\{0,1\})$ . For a set of rows  $I \subseteq [m]$ , let  $\partial_A(I)$  (called the *boundary* of I) denote all columns which, when restricted to I, contain one "1."

Then, *A* is called an (r,s,c)-expander if  $|J_i(A)| \le s$  and, for all choices *I* as above,  $|I| \le r \implies |\partial_A(I)| \ge c|I|$ .

What does this say: first, the number of 1s in any given row is bounded (by s). This allows sparseness. Secondly, up to a selectivity threshold (r), we may lower bound the density (c) of boundary columns in I by a linear factor.

# Hardness of $A \leftrightarrow (r, s, c)$ -expanders

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Definition ((r, s, c)-expanders)

Let  $A \in M_{m \times n}(\{0,1\})$ . For a set of rows  $I \subseteq [m]$ , let  $\partial_A(I)$  (called the *boundary* of I) denote all columns which, when restricted to I, contain one "1."

Then, *A* is called an (r,s,c)-expander if  $|J_i(A)| \le s$  and, for all choices *I* as above,  $|I| \le r \implies |\partial_A(I)| \ge c|I|$ .

What does this say: first, the number of 1s in any given row is bounded (by s). This allows sparseness. Secondly, up to a selectivity threshold (r), we may lower bound the density (c) of boundary columns in I by a linear factor. For instance, a (1,2,.5)-expander could look like:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

# Hardness of $g_i \leftrightarrow \ell$ -robustness

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Our hardness condition on  $g_i$  (and its motivation) is more straight-forward:

## Definition (*ℓ*-robustness)

A function  $g_i$  is called  $\ell$ -robust if every assignment  $\rho$  such that  $g_i(\rho) \in \{0,1\}$  (i.e. not  $\star$ ) satisfies  $|\rho| \ge \ell$ .

In other words, no short assignments satisfy or  $\emph{dis}$ satisfy  $\ell$ . For instance...

# Hardness of $g_i \leftrightarrow \ell$ -robustness

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Our hardness condition on  $g_i$  (and its motivation) is more straight-forward:

#### Definition (*ℓ*-robustness)

A function  $g_i$  is called  $\ell$ -robust if every assignment  $\rho$  such that  $g_i(\rho) \in \{0,1\}$  (i.e. not  $\star$ ) satisfies  $|\rho| \ge \ell$ .

In other words, no short assignments satisfy or *dissatisfy*  $\ell$ . For instance...

## Example

 $x_1 \oplus x_2 \oplus \cdots \oplus x_n$  is *n*-robust, since we need to map *all* variables to determine if the sum is  $\equiv_2 1$ .

Conversely,  $\ell_1 \vee \cdots \vee \ell_n$  is only 1-robust (take the assignment  $\ell_1 = 1$ ).

By selecting robust (i.e. large enough  $\ell$ ) functions, we increase the number of variable assignments a prover would need to check implicitly (no shortcuts).

# Proof of Theorem: Measure

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Theorem (Width of $\tau(A, G_n)$ in Resolution)

Let  $A \in M_{m \times n}(\{0,1\})$  be an (r,s,c)-expander, and let  $g_i$  be  $\ell$ -robust for i=1,...,m. Let  $c+\ell \geq s+1$ . Then

$$w_{Res}(\tau(A,G_n)) > \frac{r(c+\ell-s)}{2\ell}$$

*Proof.* We will first define a measure  $\mu$  on clauses.

## Definition $(\mu)$

For a clause *C* in Vars(*A*),  $\mu$ (*C*) is the size of a minimal  $I \subseteq [m]$  such that:

- (a)  $\forall y_f^{\varepsilon} \in C \ \exists i \in I : Vars(f) \subseteq X_i(A)$
- (b)  $\{g_i \mid i \in I\} \models ||C||$

*Remark.*  $\mu$  is sub-additive: if  $C_0$  and  $C_1$  resolve to C then  $\mu(C) \le \mu(C_0) + \mu(C_1)$ . Furthermore,  $\mu(C) = 1$  for  $C \in \tau(A, G_n)$ .

# Proof of Theorem: Roadmap

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Theorem (Width of $\tau(A, G_n)$ in Resolution)

Let  $A \in M_{m \times n}(\{0,1\})$  be an (r,s,c)-expander, and let  $g_i$  be  $\ell$ -robust for i=1,...,m. Let  $c+\ell \geq s+1$ . Then

$$w_{Res}(\tau(A,G_n)) > \frac{r(c+\ell-s)}{2\ell}$$

- **1** We'll first establish a connection between  $\mu(C)$  and w(C): medium- $\mu$  clauses are wide.
- 2 We'll then show that  $\mu(\perp)$  is large, and therefore, by our remark, that any resolution refutation of  $\tau$  must contain a medium- $\mu$  clause which is wide.

Pseudorandom Generators in Proof Complexity

For a clause *C* in Vars(*A*),  $\mu$ (*C*) is the size of a minimal  $I \subseteq [m]$  such that:

(a) 
$$\forall y_f^{\varepsilon} \in C \ \exists i \in I : Vars(f) \subseteq X_i(A)$$
 (b)  $\{g_i \mid i \in I\} \models ||C||$ 

(b) 
$$\{g_i \mid i \in I\} \models ||C||$$

## Claim (#1)

For a clause C with 
$$\frac{r}{2} < \mu(C) \le r$$
,  $w(C) > \frac{r(c+\ell-s)}{2\ell}$ .

#### Proof.

Let I satisfy  $\mu(C)$ . Let  $I_0 \subseteq I$  be minimal such that (a) still holds. Then, for,  $I_1 := I \setminus I_0$ ,  $\{g_i : i \in I \setminus k\} \not\models ||C||$  for any  $k \in I_1$ .

Fix  $k \in I_1$ . We make the sub-claim that  $|I_k(A) \cap \partial_A(I)| \le s - \ell$ . Let  $\alpha$  be an assignment such that  $g_i(\alpha) = 1 \ \forall i \in I \setminus k$ , but  $||C||(\alpha) = 0$ . (This exists by the above). Then

$$\rho(x_i) := \begin{cases} \alpha(x_i) & i \notin \partial_A(I) \cap J_k(A) \\ \star & \text{otherwise} \end{cases}$$

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle cont.

$$\rho(x_i) := \begin{cases} \alpha(x_i) & i \notin \partial_A(I) \cap J_k(A) \\ \star & \text{otherwise} \end{cases}$$

Let  $i \neq k$  be arbitrary. We claim that, if  $x_s \in \operatorname{Vars}(g_i)$ , then  $s \notin \partial_A(I) \cap J_k(A)$ , and hence  $\rho$  is defined totally on each  $g_i$ . Let  $x_s \in \operatorname{Vars}(g_i)$ . Suppose  $s \in \partial_A(I)$ . Then s is a column in which only one "1" exists. But  $s \in J_i(A)$ , so  $s \notin J_k(A)$  for any  $k \neq i$  (since this would constitute a second "1"). A similar argument shows that variables  $x_s \in ||C||$  are such that  $s \notin \partial_A(I) \cap J_k(A)$ , with the additional rationale that  $k \notin I_0$ .

 $\Longrightarrow s - |\partial_A(I) \cap I_k(A)| \ge \ell \implies |\partial_A(I) \cap I_k(A)| \le s - \ell$ 

⇒ 
$$g_i|\rho = 1$$
 and  $C|\rho = 0$ . By (b),  $g_k|\rho = 0$ . But  $g_k$  is  $\ell$ -robust:  
⇒  $\#J_k(A) \setminus [\partial_A(I) \cap J_k(A)] \ge \ell$ 

Pseudorandom Generators in Proof Complexity

> N. Hayek & B. Kyle

#### cont.

To restate: so far, we have, for any  $k \in I_1$ , the inequality

$$|\partial_A(I) \cap J_k(A)| \le s - \ell$$

Hence, we sum up

 $<(s-\ell)|I|+\ell\cdot w(C)$ 

$$\begin{split} c|I| &\leq |\partial_A(I)| & \text{by } (r,s,c)\text{-properties of } A \\ &\leq \sum_{i \in I_0} |J_i(A) \cap \partial_A(I)| + \sum_{i \in I_1} |J_i(A) \cap \partial_A(I)| \\ &\leq \sum_{i \in I_0} |J_i(A)| + (s-\ell)|I_1| & \text{by sub-claim above} \\ &\leq s|I_0| + (s-\ell)|I_1| & \text{by prop of } A \text{ (sparseness)} \\ &= (s-\ell)|I| + \ell|I_0| & \end{split}$$

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle cont.

In this last step, we use  $|I_0| \le w(C)$ . Recall that  $I_0 \subseteq I$  is minimal such that

(a) 
$$\forall y_f^{\varepsilon} \in C \ \exists i \in I : Vars(f) \subseteq X_i(A)$$

But then  $I_0 \subseteq \{i : \operatorname{Vars}(f) \subseteq X_i \text{ for some } f \leftrightarrow y_f^{\varepsilon} \in C\}$ , and the magnitude of this set is bounded by  $\{f \leftrightarrow y_f^{\varepsilon}\}$ , which is bounded by w(C).

To restate: 
$$c|I| \le (s-\ell)|I| + \ell \cdot w(C)$$
. But  $|I| = \mu(C) > \frac{r}{2}$  by assumption, so  $w(C) \ge \frac{(c+\ell-s)|I|}{\ell} > \frac{r(c+\ell-s)}{2\ell}$ .

This is the bulk of our theorem! Now that we have shown that medium- $\mu$  clauses attain our width bound, we just need to show they exist in a resolution proof.

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

#### Claim (#2)

Any resolution refutation  $\Pi$  of  $\tau$  contains a clause C with  $\frac{r}{2} < \mu(C) \le r$ .

#### Proof.

We first show  $\mu(\bot) > r$ . Suppose not:  $\mu(\bot) \le r$ .

Then we arrive at the same inequalities, i.e.  $c \cdot |I| \le (s - \ell)|I| + \ell |I_0|$ 

This time,  $I_0$  is empty, and we get  $c \le s - l$ . But the expansion property was that  $c \ge s - l + 1$  \*\*

Now, since  $\Pi$  derives  $\bot$  from clauses  $C_i$  in  $\tau$  with  $\mu(C_i) = 1$ , and  $\mu$  is sub-additive, we are done.

# Putting It All Together

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle Adding Claim 1 with Claim 2 completes our lower bound on resolution width:

$$w_{\mathrm{Res}}(\tau(A,G_n)) > \frac{r(c+\ell-s)}{2\ell}$$

Later in the paper, it is shown that nearly all matrices satisfy c > 0.9s, and that most functions satisfy  $\ell > 0.9s$ . Observe that when this is true, our bound is linear in r.

$$w_{\text{Res}}(\tau(A,G_n)) = \Omega(r)$$

Note that r is bounded above my m, but can be taken to be roughly  $\frac{n}{s}$ . This width lower-bound to a strong size-lower bound by the known relation from class. In Section 4 of the paper, the method of random restriction in Polynomial Calculus is used to extend a width lower bound to a size one. The results are *stronger* bounds on the *weaker* linear encoding.

# Corollary: Size Lower Bound

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

## Corollary

Let  $\varepsilon > 0$  be an arbitrary fixed constant, A be an  $(r,s,\varepsilon s)$ -expander of size  $(m \times n)$ , and  $g_1,...,g_m$  be  $(1-\varepsilon/2)s$ -robust functions. Then every resolution refutation of  $\tau(A,\overline{g})$  must have size  $\exp(\Omega(\frac{r^2}{m\cdot 2^{2^s}}))/2^s$ 

# **Concluding Remarks**

Pseudorandom Generators in Proof Complexity

N. Hayek & B. Kyle

- We have strong lower bounds for both resolution and algebraic systems in refuting (1), and the authors believe the same for stronger systems.
- These strong lower bounds on proving these tautologies imply even stronger lower bounds on breaking the generator itself, further affirming the strength of Nisan's and Wigderson's construction.
- For proof systems P with the Efficient Interpolation Property, there is an easy way of converting any computationally secure generator to another generator which is hard for P. But in general no such method exists.