Conspiracies between Learning Algorithms, Lower Bounds, and Pseudorandomness

Igor Carboni Oliveira
University of Oxford

Joint work with Rahul Santhanam (Oxford)
Minor algorithmic improvements imply lower bounds (Williams, 2010).

\textbf{NEXP} not contained in \textbf{ACC}^0 (Williams, 2011), and extensions.
This Work

Analogue of Williams’ celebrated lower bound program in Learning Theory.

Combining and extending existing connections.

Further applications of the “Pseudorandom Method”:

- Hardness of MCSP,
- Karp-Lipton Theorems for BPEXP.
- etc.
Lower bounds from learning
Learning Model  (Randomized, MQs, Uniform Dist.)

A Boolean circuit class $C$ is fixed.

$$f : \{0, 1\}^n \rightarrow \{0, 1\}$$  from $C[s(n)]$ is selected.

Learner must output w.h.p a hypothesis $h$ such that:

$$\Pr_{x \in \{0,1\}^n} [h(x) = f(x)] \geq 1 - 1/n.$$
Some learning algorithms

Combinatorial lower bounds

| DNF | $\not\subseteq$ | AC$^0$ | $\not\subseteq$ | AC$^0[p]$ | $\not\subseteq$ | ACC$^0$ | $\subseteq$ | TC$^0$ | $\subseteq$ | Formula[poly] | $\subseteq$ | Circuit[poly] |

Lower bounds are unknown, or obtained via diagonalization

[Jac97] DNFs can be learned in \textit{polynomial} time. Harmonic-Sieve/Boosting

[LMN93] AC$^0$ circuits learnable in \textit{quasi-polynomial} time. Fourier Concentration

[CIKK16] AC$^0[p]$ learnable in \textit{quasi-polynomial} time. Pseudorandomness/Natural Property
Can we learn $\text{AC}^0$ circuits with $\text{Mod } 6$ gates in sub-exponential time?

As far as I know, open even for:

$\text{AND } o \text{ OR } o \text{ MAJ}$ circuits, $\text{MOD}_2 o \text{ AND } o \text{ THR}$ circuits.

**Definition.** Non-trivial learning algorithm:

- Runs in randomized time $\leq \frac{2^n}{n^{\omega(1)}}$.
- For every function $f$ in $C$:

$$\Pr_{x \in \{0,1\}^n} [h(x) = f(x)] \geq \frac{1}{2} + \frac{1}{n}.$$
Non-trivial learning implies lower bounds

Let \( \text{BPE} = \text{BPTIME}[2^{O(n)}] \).

**Theorem.** Let \( C \) be any subclass of Boolean circuits closed under restrictions.

**Example:** \( C = \text{(depth-6)-ACC}^0, \ AND \ o \ OR \ o \ THR, \ etc. \)

If for each \( k>1 \), \( C[n^k] \) admits a non-trivial learning algorithm, then for each \( k > 1 \), \( \text{BPE} \) is not contained in \( C[n^k] \).
<table>
<thead>
<tr>
<th>Assumption</th>
<th>Non-trivial SAT/Proof System</th>
<th>Non-trivial Derandomization</th>
<th>Non-trivial Deterministic Exact Learning</th>
<th>Non-trivial Randomized Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proofs checked in deterministic time</strong></td>
<td>$2^n / n^{\omega(1)}$</td>
<td>$2^n / n^{\omega(1)}$</td>
<td>$&lt; 2^n$</td>
<td>Learner runs in randomized time $2^n / n^{\omega(1)}$</td>
</tr>
<tr>
<td><strong>Algorithm runs in deterministic time</strong></td>
<td>$2^n / n^{\omega(1)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learner runs in deterministic time</strong></td>
<td>$&lt; 2^n$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learner runs in randomized time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consequence</strong></td>
<td>LBs for NEXP</td>
<td>LBs for NEXP</td>
<td>LBs for EXP</td>
<td>LBs for BPEXP</td>
</tr>
<tr>
<td><strong>Reference</strong></td>
<td>[Wil10]</td>
<td>[Wil10], [SW13]</td>
<td>[KKO13]</td>
<td>[This Work]</td>
</tr>
</tbody>
</table>
Remarks on lower bounds from Learning

▶ Learning approach won’t directly work for classes containing PRFs.

▶ Conceivable that one can design non-trivial learning algorithms for a class $C$ under the assumption that $\text{BPEXP}$ is contained in $\text{P/poly}$.

▶ Learning connection applies to virtually any circuit class of interest, and there is **no depth blow-up**.

It can lead to new lower bounds for restricted classes such as $\text{THR} \circ \text{THR}$ and $\text{ACC}^0$. 
Previous work on learning vs. lower bounds

- Systematic investigation initiated about 10 years ago:
  
  [FK06] Lower bounds for $\text{BPEXP}$ from polynomial time learnability.
  
  [HH11] Lower bounds for $\text{EXP}$ from deterministic exact learning.
  
  [KKO13] Optimal lower bounds for $\text{EXP}$ from deterministic exact learning.
  
  [Vol14] Lower bounds for $\text{BPP}/1$ from polynomial time learnability.
  
A Challenge in Getting Lower Bounds from Randomized Learning

Williams’ lower bounds from non-trivial SAT algorithms: a non-trivial algorithm can be used to violate a tight hierarchy theorem for NTIME.

Challenge in Randomized Learning:
   lack of strong hierarchy theorems for BPTIME.

The approach has to be indirect, and we must do something different ...
Speedup Phenomenon in Learning Theory

**Speedup Lemma.** Let $\mathcal{C}$ be any class of Boolean circuits containing $\text{AC}^0[2]$.

Suppose that for each $k \geq 1$ the class $\mathcal{C}(n^k)$ admits a non-trivial learning algorithm.

Then for each $k \geq 1$ and $\varepsilon > 0$, the class $\mathcal{C}(n^k)$ is strongly learnable in time $O(2^{n^\varepsilon})$. 
SAT Algorithms vs. Learning Algorithms

Exponential Time Regime

“Nontrivial”
SETH is false
ETH is false
NP ⊆ SUBEXP
P = NP

(Running Time)

2^n/n^ω(1)
2^{cn}, c < 1
2^{o(n)}
2^{n^ε}
poly(n)

Complexity of CNF-SAT

Complexity of Learning C[poly]

Equivalent
[Speedup Phenomenon]
Main Techniques: “Speedup Lemma”

1. Given oracle access to $f : \{0, 1\}^n \rightarrow \{0, 1\}$ in $\mathbb{C}[\text{poly}]$, implicitly construct a “pseudorandom” ensemble of functions in $\mathbb{C}[\text{poly}]$ on $n^\delta$ bits.
   
   (using NW-generator + Hardness Amplification [CIKK16])

   **Intuition:** Non-trivial learner can distinguish this ensemble from random functions. This can be done in time $2^{O(n^\delta)}$.

2. This distinguisher (i.e. the non-trivial learner) and the reconstruction procedures of NW-generator and Hardness Amplification can be used to strongly learn $f$ in time $2^{O(n^\delta)}$. 
Main Techniques: “LBs from Learning”

1. Starting from non-trivial learner, apply the Speedup Lemma to obtain a sub-exponential time learner.

2. Adapting the techniques from [KKO13], randomized sub-exponential time learnability of $C[poly]$ implies BPE lower bounds against $C[n^k]$. 

3. Using an additional win-win argument, this holds under minimal assumptions on $C$, and with no blow-up in the reduction.
Combining and extending existing connections
Nontrivial SAT / $\text{ACC}^0$ lower bounds [Wil11]

P/poly lower bounds [BFT98]

Non-trivial learning [This work]

Well-known connection to PRGs/derandomization of BPP [IW97]

[OS17] Connections to pseudo-deterministic algorithms.

Further motivation for the following question:

Which algorithmic upper bounds imply lower bounds for ZPEXP and REXP, respectively?
One-sided error: Lower bounds for REXP

We combine the satisfiability and learning connections to lower bounds to show:

[ Informal ]
If a circuit class $C$ admits both non-trivial SAT and non-trivial Learning then REXP is not contained in $C$.

**Corollary. [ACC$^0$ lower bounds from non-trivial learning]**
If for every depth $d>1$ and modulo $m>1$ there is $\varepsilon > 0$ such that $\text{ACC}^0_{d,m}(2^{n^\varepsilon})$ has non-trivial learning algorithms, then $\text{REXP} \not\subseteq \text{ACC}^0(\text{poly}(n))$.

Indicates that combining the two frameworks might have further benefits.
Zero-error: Lower bounds for ZPEXP

[IKW02], [Wil13] Connections between natural properties without density condition, Satisfiability Algorithms, and NEXP lower bounds.

[CIKK16] Connections between BPP-natural properties and Learning Algorithms.

We give a new connection between P-natural properties and ZPEXP lower bounds.

Let $\mathcal{C}(\text{poly}) \subseteq \text{P/poly}$ be a circuit class closed under restrictions.

Theorem. [ZPEXP lower bounds from natural properties]
If for some $\delta > 0$ there are P-natural properties against $\mathcal{C}(2^{n^\delta})$ then ZPEXP $\not\subseteq \mathcal{C}(\text{poly}(n))$. 
Further Applications of our Techniques
A rich web of techniques and connections

Use of (conditional) **PRGs** and related tools, often in contexts where (**pseudo**)**randomness** is not intrinsic.
Karp-Lipton Collapses

Connection between uniform class and non-uniform circuit class:

[KL80] If $\text{NP} \subseteq \text{P/poly}$ then $\text{PH} = \Sigma_2^P \cap \Pi_2^P$.

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Consequence</th>
<th>Major Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP in P/poly</td>
<td>EXP = MA [BFT98]</td>
<td>$\text{MA}_{\text{EXP}}$ not in P/poly [BFT98]</td>
</tr>
<tr>
<td>NEXP in P/poly</td>
<td>NEXP = EXP [IKW02]</td>
<td>SAT / LB Connection [Wil10]</td>
</tr>
</tbody>
</table>

Randomized Exponential Classes such as $\text{BPEXP}$?
Karp-Lipton for randomized classes

**Theorem 1.** If $\text{BPE} \subseteq \text{i.o.SIZE}[n^k]$ then $\text{BPEXP} \subseteq \text{i.o.EXP}/O(\log n)$.

The advice is needed for technical reasons. But it can be eliminated in some cases:

**Theorem 2.** If $\text{BPE} \subseteq \text{i.o.SIZE}[n^k]$ then $\text{REXP} \subseteq \text{i.o.EXP}$.

- Check paper for Karp-Lipton collapses for $\text{ZPEXP}$, and related results.
Hardness of MCSP

Minimum Circuit Size Problem:

Given $1^s$ and a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ represented as an N-bit string,

Is it computed by a circuit of size at most $s$?

Recent work on MCSP and its variants: [KC00], [ABK+06], [AHM+08], [KS08], [AD14], [HP15], [AHK15], [MW15], [HP15], [AGM15], [HW16].

[ABK+06] MCSP is not in $\text{AC}^0$. Open. Prove that MCSP is not in $\text{AC}^0[2]$ !
We prove the first hardness result for \textbf{MCSP} for a standard complexity class beyond \textbf{AC}^0:

**Theorem.** If \textbf{MCSP} is in \textbf{TC}^0 then \textbf{NC}^1 collapses to \textbf{TC}^0.

The argument describes a non-uniform \textbf{TC}^0 reduction from \textbf{NC}^1 to \textbf{MCSP} via \textit{pseudorandomness}. 
Additional applications of our techniques

▶ Equivalences between truth-table compression [CKK+14] and randomized learning models in the sub-exponential time regime.

(For instance, equivalence queries can be eliminated in sub-exp time randomized learning of expressive concept classes.)

▶ A dichotomy between Learnability and Pseudorandomness in the non-uniform exponential-security setting:

“A circuit class is either learnable or contains pseudorandom functions, but not both.”

In other words, learnability is the only obstruction to pseudorandomness.

(Morally, ACC⁰ is either learnable in sub-exp time or contains exp-secure PRFs.)
Problems and Directions

Is there a speedup phenomenon for complex classes (say AC$^0[p]$ and above) for learning under the uniform distribution with random examples?

Can we establish new lower bounds for modest circuit classes by designing non-trivial learning algorithms?
Towards lower bounds against NC?

Non-trivial learning implies lower bounds:

First example of lower bound connection from non-trivial randomized algorithms.

**Problem.** Establish a connection between non-trivial randomized SAT algorithms and lower bounds.

(First step in a program to obtain unconditional lower bounds against NC.)
Thank you