2018-2019 Annual Progress Report for Simons Investigator Award

Russell Impagliazzo
CSE Department
University of California, San Diego

July 31, 2020

1 Main Research Accomplishments:

Smooth boosting, hardcore distributions, and dense model theorems
Boosting is an important technique in machine learning, where a weak learner that gets a small advantage in prediction is amplified to a strong learner whose predictions are highly accurate. Klivans and Servidio observed that the Hardcore Distribution Lemmas of Impagliazzo and Holenstenin are essentially proven using boosting algorithms with a smoothness property, and that these boosting algorithms have other applications in learning theory. In recent work, I and my students have made simplifications in the design and analysis of smooth boosting algorithms. These simplifications allow a deeper understanding of their sample complexities. Mark Bun and my Ph.D. students Marco Carmosino and Jess Sorrell have used this new understanding to design learning algorithms that both achieve differential privacy and use privacy techniques to reduce their sample complexity.

Rex Liu and I are working on matching lower bounds on sample complexity for smooth boosting. With Rex Liu, Jess Sorrell and other researchers, we are using these new analysis techniques to design noise-tolerant boosting algorithms. Sam McGuire and I showed that smooth boosting requires the ability to compute majority, and that the hardcore distribution lemma and dense model theorems are actually false for a class AC_0 that cannot compute majority.

Algebraic proof systems Constant depth circuits with prime modular counting gates are the largest class for which constructive lower bounds within EXP are known. A major problem for three decades is to give a lower bound for the analogous proof system. Although they have found many

applications since then, the algebraic proof systems such as Nullstellensatz and polynomial calculus were origionally motivated as stepping stones towards such a lower bound. However, these systems seem just short of being able to simulate Frege proofs with arbitrary constant depth circuits. There is a reaonably straight-forward way to strengthen such systems to allow deeper circuits to be expressed. However, here, we show that these strengthenings go well beyond constant depth Frege proofs. Even in constant prime characteristic fields, depth seven algebraic proofs can simulate any constant depth of mod p Frege (even for p a different characteristic), cutting planes and other TC_0-Frege systems, and sum of squares proofs. Thus, lower bounds for such systems will be very difficult. This is joint work with my student Sasank Mouli and Toniann Pitassi, started at the Simons Institute program on lower bounds. See: Russell Impagliazzo, Sasank Mouli, Toniann Pitassi: The Surprising Power of Constant Depth Algebraic Proofs. LICS, 2020.

In work in preparation, the same authors have a lower bound on such proofs when a moderate (sub-quardratic) number of extension variables can be introduced. We also are working on "SETH" style lower bounds for polynomial-calculus, pushing degree lower bounds closer to n, and size lower bounds closer to the maximal 2^n.

Fine-grained complexity of extensions of first-order logic In previous work with my student Jiawei Gao and Antoninia Kolokolova and Ryan Williams, we showed that the well studied orthogonal vectors problem was complete under fine-grained reductitons for the class of all first-order expressible properties. This gave the first fine-grained completeness result for a well-studied class, and at the same time, moderately improved algorithms for all first-order properties. In: Jiawei Gao, Russell Impagliazzo: The Fine-Grained Complexity of Strengthenings of First-Order Logic. Electronic Colloquium on Computational Complexity (ECCC) 26: 9 , we extend this characterization to several stronger forms of logic. Surprisingly, the fine-grained complexity of logics seems uncorrelated to their expressive power. Extensions such as adding an ordering that add power to the logic do not change the fine-grained complexity of the class (with orthogonal vectors still being a complete problem, resulting in similarly improved algorithms.) On the other hand, adding function symbols does not change expressive power, but increases the fine-grained complexity by almost a linear factor. Related sole-authored work by Jiawei Gao, to appear in IPEC 2019, captures the fine-grained complexity of problems involving reachability in multi-trees.

While we resolved the fine-grained complexity of allowing a single ordering in addition to first order properties, we left open the question of multiple orderings. I have been working with a team of undergraduates just starting their junior years (Haosze An, Harrison Ku, Mohit Gurumukhani,

Michael Jaber, and Maria Parga-nina) on this question, and, with Marvin Kunneman, we recently submitted the resulting paper to SODA 2021. advanced undergraduates such as Cody Murray before, this is my first experience with such early interest by a team that work so well together.

2 People supported

This year, the Simons Investigator Award was my chief source of funds for supporting my graduate students. Next year, I also will have NSF funding.

- Marco Carmosino studies connections between lower bounds and algorithms, and in particular, between lower bounds and learning. He graduated last summer, but he was supported for about a month during the current reporting period. He now is a post-doctoral fellow at Simons Fraser University, and will soon be a post-doctoral fellow at Boston University working with Mark Bun. Bun, Carmosino, and UCSD grad student Jess Sorell have a project on connections between privacy and generalizations in learning, that potentially explains how and why learning algorithms can achieve good generalization when the amount of data is much less than the number of parameters. A paper on this subject, partially funded by this award, appeared in COLT 2020.

- Jiawei Gao also graduated last summer, and is now at Google. Her work looked at the fine-grained complexity of logically defined complexity classes. Her recent work was described above. She was not directly supported by the grant this period, but expenses related to presenting her work were covered.

- Anant Dhayal (year 5) has a joint paper in submission on extensions of the Easy Witness method to other complexity classes and its consequences for connections between algorithms and lower bounds. He will defend in the Fall quarter.

- Venkata Sai Sasank Mouli Gali (year 5) is working on algebraic circuit lower bounds and algebraic propositional proof systems. His recent work was mentioned above.

- Ivan Mikhailin (year 5) works on both circuit lower bounds and the fine grained complexity of NP-complete and parameterized problems. He plans to defend this Summer.

- Jess Sorrell (year 4) is working on connections between privacy and generalization in learning, and on lattice algorithms and applications to cryptography. She is co-advised by Daniele Micciancio. In addition to lattice cryptography and the work mentioned above with Mark Bun and Marco

Carmosino, she is working with Rex Liu, myself, and others, on using techniques from hardcore distribution lemmas to design boosting-based learning algorithms in the noisy and agnostic settings. She was not directly supported by the grant this period, but expenses related to presenting her work were covered.

- Kenneth Hoover (year 3) is working on reductions between variants of MCSP.

- Sam McGuire, second year, co-advised with Shachar Lovett is working on connections between pseudo-randomness and learning. We have a joint paper in submission on the necessity of majority for boosting algorithms, hardcore distribution lemmas, and dense model theorems.

- Rex Liu, second year, is working on sample complexity of learning algorithms and the project mentioned above on learning in highly noisy and agnostic settings.

The Simons Investigator Award has both paid the tuition and stipends for these students, and given them support to travel and participate in programs such as the Simons Institute Pseudo-randomness and Lower Bounds programs.

3 Talks

I have not given any grant related talks this period, but the students above have presented their work at the conferences listed and workshops such as the Banff workshop on proof complexity.