Research statement

Antonina Kolokolova

What does it mean for a problem or a concept to be "hard"? It can be a computational problem that we don't know how to solve without spending a lot of time and memory, or a concept hard to express in a (formal) language, or a theorem that requires a long proof. Complexity theory studies the notion of hardness in its different aspects and analyses the relationship among them. The unifying theme of my research is the study of concepts (patterns) and their complexity, for the patterns described in the language of logic.

The most common definition of complexity is in the framework of *computational complexity*, formulated in terms of limited computational resources such as time, space, circuit size or the number of processors. Traditionally, the research in this field is aimed at understanding the relationship between different amounts and types of resources, and classifying natural and interesting problems according to their complexity. This is an active field, which has seen several major results in the last several years, for example, a celebrated polynomial-time algorithm for testing if a number is prime. However, the main open questions, in particular the famous P vs. NP question, remain unresolved.

Complexity arises in many other contexts, seemingly unrelated to computation, such as mathematical logic. There are several notions of complexity in logic. The first is *descriptive complexity* that arises in finite model theory. Here the complexity of a set of formulae corresponds to the hardness of the properties they express. For example, first-order formulae cannot express graph reachability. This notion found its application in the field of databases, where it is used to estimate the expressive power of database languages. For example, in a first-order database language without a built-in reachability query, it is impossible to ask an airline flight database if there is a way to fly from city A to city B without limiting the number of transfers. Establishing the correspondence between logics and complexity classes allows one to use tools of one area to attack open questions in other area. Such was the case for the closure of non-deterministic logspace under complementation: Immerman's original insight and proof came from the descriptive complexity framework [Imm88].

The second notion is *proof complexity* that studies the power of reasoning. Given a weak system of arithmetic, how hard are the properties provable in this system? To relate this notion to complexity classes, consider the classes of functions that are provably total in such a system. For example, a system of arithmetic in which all quantifiers are bounded by polynomials cannot prove the Parity principle, the principle which states that there is no perfect matching on an odd number of vertices. This fact was proven by Ajtai [Ajt83] using the machinery of model theory; in the language of complexity theory, this states that the Parity function is not computable in AC_0, by polynomial-size circuits of constant depth.

Another aspect of logic in complexity theory is in analyzing the complexity of mathematical techniques needed to resolve complexity-theoretic questions. There were several meta-results in complexity theory stating that a certain class of techniques would not, by itself, be sufficient to resolve some open questions. The most well-known of them is Baker, Gill and Solovay [BGS75] result that diagonalization (or any technique insensitive to the presence of oracles) cannot resolve P vs. NP and some other important complexity questions. But what does it mean in the mathematical terms? Arora, Impagliazzo and Vazirani [AIV92] formalized this as independence of specific logical theory; in the recently submitted paper by Impagliazzo, Kabanets and myself [IKK] this framework is extended to the "algebrization barrier" by Aaronson and Wigderson [AW08].

Yet another side is analyzing the complexity of the objects used in proofs of complexity-theoretic results. For example, several major recent results make heavy use of the properties of expander graphs. In particular, Reingold used expanders in his proof that undirected graph reachability is in LogSpace [Rei05]. But how complex, in logical terms, is it to construct and reason about properties of such graphs?

Although there has been a lot of progress, the main questions about complexity theory remain unsolved, and the relations between many notions of complexity are not well understood. The goal of my research is to study the ways complexity occurs in various, primarily machine-independent settings such as mathematical logic characterizations, and the relations of these types of complexity to computational complexity and to each other. The hope is that combining different ways of viewing complexity can open new directions and help solve open questions in both complexity theory and logic.

1 Proving properties of "easy" patterns.

What can we prove if we are only allowed to use "small" objects, that is, objects that can be described using a weak logic? A natural framework to study such a question is proof complexity. One approach is to study complexity of reasoning in (predominantly propositional) proof systems. A uniform counterpart of proof systems is bounded arithmetic.

In my PhD thesis, I pinpoint the relationship between the expressive power of a logic and the strength of a system of bounded arithmetic that operates with objects described by that logic. The main result of my thesis states that if a logic is provably closed, then the corresponding system of arithmetic has the proving power of exactly the same complexity as the descriptive power of the logic; otherwise it is weaker. Here, "provably closed" means that not only is the logic closed under first-order operations, but also that there exist "simple" proofs of that closure (there are no formulas in the proof that are more complicated than the formulas from the logic). This is presented in [Kol05].

Based on the theory V_0 capturing complexity class AC^0 and thus the first-order logic, we developed systems of arithmetic for polynomial time [CK01, CK03] and non-deterministic log-space [CK04] complexity classes. The theory for polynomial time is based on second-order Horn logic, and for NL on second-order 2CNF, both due to Grädel [Grä92]. We show that it is "easy" to prove the basic properties of first-order logic augmented with transitive closure operator: transitivity, existence of a normal form, and the closure under complementation itself.

The situation is different for a seemingly similar complexity class SL of problems reducible to the reachability problem on undirected graphs. The only known proof of closure of SL, as well as the recent result that SL coincides with L (deterministic LogSpace) [Rei05], use expander graphs and algebraic reasoning, which is not known to be formalizable in such low complexity class as SL. We can still get a viable system of arithmetic, but its capture of SL is not as strong. One of my current projects is to estimate the complexity needed to reason about expander graphs. We make a step towards this in a paper with Koucky and Kabanets [KKK] (in preparation), where we try to analyze the complexity of a specific construction of expander graphs.

2 Model Expansion framework and problem solving

In bounded arithmetic theories, we can view the power of a system of arithmetic in terms of the complexity of functions that witness existential second-order quantifiers in fragments of second-order arithmetic. These ideas turned out to be very useful for a descriptive complexity-based solver project of Eugenia Ternovska, David Mitchell and their group. The key idea there was using existential second-order fragments of various logics by stating a search problem as finding an expansion of a given instance of a problem to a larger instance which contains a solution. For example, when solving 3-colourability, we encode it as a first-order formula with conditions on three free second-order variables representing colours, and the solution is an expansion of a model (in this case, a graph) by three relations corresponding to the colours. The commonly used approaches to problem solving either encode a problem as a SAT instance directly, or are based on stable-model semantic of Answer Set Programming.

A survey of results on complexity of this problem for various logics was presented at the Logic and Computational Complexity (LCC'2006). The full version appeared as an SFU technical report TR2007-29 [KLMT07]; a journal version is in preparation.

3 Impossibility as independence

The celebrated result by Baker, Gill and Solovay showed that diagonalization (or any technique that is "relativizing", i.e., insensitive to the presence of oracles) cannot resolve longstanding open questions such as P vs. NP, P vs. PSPACE and many other.

Since then, new techniques have been developed to overcome this relativization barrier. In particular, arithmetization was a non-relativizing technique which led to results unprovable with just relativizing tools, most notable IP=PSPACE result. However, there are recent papers arguing that arithmetization by itself is not sufficient to resolve P vs. NP and several other questions. In their STOC'08 paper [AW08], Aaronson and Wigderson present the concept of "algebrization barrier": the arithmetizing techniques cannot resolve the main open questions in complexity. Even before that, Fortnow [For94] showed that the results using arithmetization remain insensitive to the presence of a certain limited class of oracles, however he does not show that outcomes of P vs. NP differ for oracles from this class.

But what does it mean, in logic terms, that a technique is not applicable to resolve a certain open problem? Arora, Impagliazzo and Vazirani [AIV92] have developed a framework relating the notion of non-applicability of a technique with independence from a theory. They present a theory of arithmetic RCT in which all provable statements are relativizable. Thus, if model-theoretic tools show that there are models RCT with contradictory answers to a complexity-theoretic question then this question is not resolvable using relativizing technique. They also present an axiom, local checkability, which would strengthen the theory so much that a resolution of a question in this theory, although still insensitive to a very restricted class of oracles, would be essentially the same as in the non-oracle setting.

In the paper with Impagliazzo and Kabanets (submitted) [IKK] we extend the [AIV92] framework to capture the notion of an algebrizing technique. That is, the theory RCT is extended with an axiom limiting oracles to ones admitting arithmetization. In the resulting theory, ACT, provable statements are exactly such that are possible to prove using arithmetization, and independence of ACT implies impossibility of a proof based only on arithmetizations.

4 Future directions

My short-term research goals are to explore more complexity of formalizing expanders and the locality notions of [AIV92]. It would be very interesting to pinpoint exactly what is the complexity of reasoning using expander graphs (and thus establish a lower bound on proof complexity of existing proofs using expanders such as SL=L result or AKS sorting networks). There are variants of locality notions of [AIV92] which are used as bases for several theories of arithmetic capturing polynomial time (Cobham's framework, Horn satisfiability, alternating log-space). It would be interesting to study under which assumptions those theories are equivalent.

Another direction which I would like to explore in the near future is the connections with constraint satisfaction problems: in recent years there have been beautiful results relating subclasses of constraint satisfaction problems and expressibility in variants of Datalog.

The study of patterns and their complexity is an exciting field, presenting a new way to look at both theoretical and practical problems. I believe that the logic approach will continue being fruitful in such study. On the theoretical side, I am interested in the logic and complexity study of various algebraic and combinatorial objects (such as expander graphs), reasoning complexity of complexity-theoretic techniques, properties of logics (such as 0-1 laws) and other connections between logic and complexity theory. On the practical side, I am interested in finding new applications of the logical study of patterns and their complexity, and welcome collaboration from other areas.

References

[AIV92] S. Arora, R. Impagliazzo, and U. Vazirani. Relativizing versus nonrelativizing techniques: The role of local checkability. Manuscript, 1992.

[Ajt83] M. Ajtai. Σ_1^1-Formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

[AW08] S. Aaronson and A. Wigderson. Algebrization: A new barrier in complexity theory. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 731–740, 2008.

[BGS75] T. Baker, J. Gill, and R. Solovay. Relativizations of the P=?NP question. *SIAM Journal on Computing*, 4(4):431–442, 1975.

[CK01] S.A. Cook and A. Kolokolova. A second-order system for polynomial-time reasoning based on Grädel's theorem. In *Proceedings of the Sixteens annual IEEE symposium on Logic in Computer Science*, pages 177–186, 2001.

[CK03] S.A. Cook and A. Kolokolova. A second-order system for polytime reasoning based on Grädel's theorem. *Annals of Pure and Applied Logic*, 124:193–231, 2003.

[CK04] S.A. Cook and A. Kolokolova. Bounded arithmetic of NL. In *Proceedings of the Nineteens annual IEEE symposium on Logic in Computer Science*, pages 398–407, 2004.

[For94] L. Fortnow. The role of relativization in complexity theory. *Bulletin of the European Association for Theoretical Computer Science*, 52:229–244, February 1994. Columns: Structural Complexity.

[Grä92] E. Grädel. Capturing Complexity Classes by Fragments of Second Order Logic. *Theoretical Computer Science*, 101:35–57, 1992.

[IKK] R. Impagliazzo, V. Kabanets, and A. Kolokolova. An axiomatic approach to algebrization. submitted to STOC'09.

[Imm88] Immerman. Nondeterministic space is closed under complementation. In *SCT: Annual Conference on Structure in Complexity Theory*, 1988.

[KKK] M. Koucky, V. Kabanets, and A. Kolokolova. Expanders made elementary. in preparation.

[KLMT07] A. Kolokolova, Y. Liu, D. Mitchell, and E. Ternovska. Model expansion and the expressiveness of fo(id) and other logics. Technical Report TR2007-29, Simon Fraser University, 2007.

[Kol05] A. Kolokolova. Closure properties of weak systems of bounded arithmetic. In *Computer Science Logic: 19th International Workshop, CSL 2005, 14th Annual Conference of the EACSL, Oxford, UK, August 22-25, 2005. Proceedings*, volume 3634 of *LNCS*, pages 369–383, 2005.

[Rei05] Omer Reingold. Undirected st-connectivity in log-space. In *STOC*, pages 376–385, 2005.