Mathematisches Forschungsinstitut Oberwolfach

Report No. 54/2015

DOI: 10.4171/OWR/2015/54

Complexity Theory

Organised by
Peter Bürgisser, Berlin
Oded Goldreich, Rehovot
Madhu Sudan, Cambridge MA
Salil Vadhan, Cambridge MA

15 November – 21 November 2015

Abstract. Computational Complexity Theory is the mathematical study of the intrinsic power and limitations of computational resources like time, space, or randomness. The current workshop focused on recent developments in various sub-areas including arithmetic complexity, Boolean complexity, communication complexity, cryptography, probabilistic proof systems, pseudorandomness and randomness extraction. Many of the developments are related to diverse mathematical fields such as algebraic geometry, combinatorial number theory, probability theory, representation theory, and the theory of error-correcting codes.

Mathematics Subject Classification (2010): 68-06, 68Q01, 68Q10, 68Q15, 68Q17, 68Q25, 94B05, 94B35.

Introduction by the Organisers

The workshop *Complexity Theory* was organized by Peter Bürgisser (TU Berlin), Oded Goldreich (Weizmann Institute), Madhu Sudan (Harvard), and Salil Vadhan (Harvard). The workshop was held on November 15th–21st 2015, and attended by approximately 50 participants spanning a wide range of interests within the field of Computational Complexity. The plenary program, attended by all participants, featured fifteen long lectures and five short (8-minute) reports by students and postdocs. In addition, intensive interaction took place in smaller groups.

The Oberwolfach Meeting on Complexity Theory is marked by a long tradition and a continuous transformation. Originally starting with a focus on algebraic and Boolean complexity, the meeting has continuously evolved to cover a wide variety

of areas, most of which were not even in existence at the time of the first meeting (in 1972). While inviting many of the most prominent researchers in the field, the organizers try to identify and invite a fair number of promising young researchers.

Computational complexity (a.k.a. complexity theory) is a central field of computer science with a remarkable list of celebrated achievements as well as a vibrant research activity. The field is concerned with the study of the *intrinsic complexity* of computational tasks, and this study tends to *aim at generality*: it focuses on natural computational resources, and considers the effect of limiting these resources on the class of problems that can be solved. Computational complexity is related to and has substantial interaction with other areas of mathematics such as algebra, analysis, combinatorics, geometry, number theory, optimization, probability theory, and quantum computation.

The workshop focused on several sub-areas of complexity theory and its nature may be best illustrated by a brief survey of some of the meeting's highlights.

Randomness Extraction. The problem of extracting almost perfect randomness from sources of highly defected randomness is of great theoretical and practical importance, since perfect randomness is essential to cryptography and has numerous applications in algorithmic design, whereas the natural sources of randomness are quite defected. One important setting of the problem refers to the case in which one is given samples drawn from two independent sources of defected randomness, where the level of defect is captured by a lower bound on the probability that the outcome equals any specific value. The logarithm of the reciprocal of this probability, called *min-entropy*, is a main parameter in these studies.

While it is easy to prove the existence of two-source (randomness) extractors for sources of logarithmic min-entropy, the explicit construction of extractors that can handle min-entropy rate below half was open since 1985. In 2005, Jean Bourgain obtained an explicit construction for min-entropy rate slightly below half (i.e., 0.499), but no progress on this problem has been reported till July 2015, when Eshan Chattopadhyay and David Zuckerman announced a construction that can handle poly-logarithmic min-entropy.

The workshop's actual program started with a special session devoted to this breakthrough, with both authors present. David Zuckerman presented the history and wide context of the problem of constructing two-source extractors, and Gil Cohen presented an overview of the construction. One informal specialized session, which took place on a later day, featured more detailed descriptions of two components of the construction. Specifically, Eshan Chattopadhyay presented constructions of "non-malleable extractors" and Raghu Meka presented constructions of "resilient functions".

One interesting point regarding the construction of Chattopadhyay and Zuckerman is that its analysis makes explicit use of a celebrated result about the computational limitations of bounded-depth Boolean circuits (which was presented by Mark Braverman in the 2009 complexity meeting at Oberwolfach). This is remarkable because these two areas of complexity theory did not seem related

before and their history did not register any actual interaction so far. Another peculiar connection is the use of Uri Feige's leader election protocol for the construction of non-malleable extractors, whereas this protocol was discovered in the 1998 complexity meeting at Oberwolfach, following the presentation of a different protocol by David Zuckerman (which in turn drew on ideas from the area of pseudorandomness).

Boolean Circuit Lower Bounds. The project of establishing circuit lower bounds calls for presenting explicit functions that cannot be computed within limited computational resources. One direction of research is aimed at better understanding of very restricted computation devices such as (unbounded fan-in) bounded-depth circuits and formulae.

Ben Rossman outlined his proof that shows that the simple conversion of circuits of size s and depth d into formulae of size s^d and depth d is essentially the best possible. Specifically, he showed that the parity of n variables, which can be computed by a depth d circuit of size $\exp(n^{1/(d-1)})$, requires depth d formula of size $\exp(\Omega(d \cdot n^{1/d}))$.

Avishay Tal addressed the problem of presenting explicit functions that require depth-three circuits of size $\exp(\omega(\sqrt{n}))$. He presented a proof of such a result in a *restricted model* of depth three circuits, which arises from a natural model of multi-linear circuits for computing multi-linear functions, by studying the "rigidity" of random Boolean Toeplitz matrices. Specifically, he showed that such a random matrix disagrees with any rank r matrix on at least $\widetilde{\Omega}(n^3/r^2)$ entries, which improves over the previously known bound of $\Omega(n^2/r)$ when $r < n/\log^2 n$.

Fine-grained complexity. A relatively recent direction of research refers to the study of problems that are known to have polynomial-time algorithms, where the aim is to provide evidence that the known algorithms are actually the best possible. Ryan Williams surveyed research in this direction, known as fine-grained complexity, while highlighting the connection between it and the study of the exact complexity of problems that seem to require exponential-time such as 3SAT.

Doubly-efficient interactive proof systems. The invention of interactive proof systems and the exploration of their power are among the greatest success stories of computational complexity. While research in the 1980s referred to polynomial-time verification aided by a computationally unbounded prover, the term doubly-efficient refers to almost linear-time verification aided by a polynomial-time prover. Clearly, only polynomial-time solvable problems can have such a proof system, even if the soundness condition is relaxed to hold only with respect to polynomial-time cheating provers (who attempt to prove false claims).

This upper bound (on the complexity of problems having doubly-efficient interactive proof systems) is met by a result presented by Ron Rothblum, which uses only one round of communication and relies on standard intractability assumptions. A different system, presented by Rothblum in a specialized session, achieves information theoretic soundness (in a larger constant number of rounds)

for any problem that can be solved in polynomial time and space $n^{o(1)}$. (The space bound is the best possible, up to a constant power.)

Two-server PIR with improved communication complexity. While the computational assumption used by Rothblum refers to one-server computational *Private Information Retrieval* (PIR) schemes, two-server PIRs offer information theoretic security. Specifically, one can retrieve any desired bit in an n-bit long string, held by each of the two servers, by exchanging $O(n^{1/3})$ bits of communication with each server such that no single server gets information about the identity of the desired bit. The simple scheme, invented in 1995, stood unimproved for two decades. Zeev Dvir presented a vast improvement on this simple scheme, by building on results of Yekhanin and Efremenko, which were presented in past Oberwolfach meetings (in 2007 and 2009, resp). The new scheme uses $\exp(\widetilde{O}(\sqrt{\log n})) = n^{o(1)}$ bits of communication, and relies on a construction of "matching vectors" family over a finite ring.

High-rate locally-testable and locally-correctable codes. The aforementioned results of Yekhanin and Efremenko refer to the construction of codes that support the recovery of any bit in a corrupted codeword based on a constant number of random probes (i.e., it achieves constant "locality"). These known results refer to codes of sub-exponential length (i.e., the codeword has length that is sub-exponential in the length of the message), and it is also known that such level of locality cannot be supported by codes of almost linear length. In his presentation, Or Meir considered the opposite extreme of the length-vs-locality trade-off: The case in which one requires the code to have linear length (or even length that is optimal with respect to its distance), and tries to minimize the number of probes that suffices for recovering a single bit. The new result asserts $\exp(\widetilde{O}(\sqrt{\log n})) = n^{o(1)}$ proves suffice to the n-bit codeword, whereas the prior bound was $n^{1/O(1)}$.

With respect to local testability (i.e., testing whether a string is a valid codeword or far from it by making few queries), the results are better. In the constant-probe regime codes of almost-linear length are known, whereas the new work present linear-length codes that are testable by a quasi-poly-logarithmic number of probes (i.e., the number of probes is $(\log n)^{O(\log\log n)}$).

Computational assumptions in cryptography. Modern cryptography is based on computational assumptions, since its most basic primitive such as secure encryption and unforgeable signatures imply the existence of *one-way functions* (OWF), which in turn is a very strong version of the famous conjecture by which $\mathcal{P} \neq \mathcal{N}P$. In recent years, far stronger computational assumptions became popular in cryptographic research. One such assumption, known as the IO conjecture, postulates that it is feasible to obfuscate computer programs such that the obfuscations of functionally equivalent programs cannot be distinguished. Vinod Vaikuntanathan presented a unified framework in which a wide spectrum of cryptographic assumptions, ranging from the (very minimal) assumption by which OWF exist to the highly speculative IO conjecture. He also noted that the IO conjecture does not imply OWF (nor does it even imply $\mathcal{P} \neq \mathcal{N}P$).

Preventing false discovery in interactive data analysis. It may seem weird that such a title fits in a complexity theoretic workshop, but it turns out that a natural formulation of adaptive (or interactive) data analysis yields a natural computational problem. As explained by Jon Ullman, *interactive data analysis* refers to a setting in which first one obtains a sample of the data, and then one conducts a study of this sample by issuing queries and examining the answers (e.g., testing various hypotheses regarding the data). The point is that these queries are selected adaptively based on prior answers, and the problem is to avoid (false) discoveries that are tailored on the sample but do not reflect the original data. One key observation is that avoiding such a phenomenon is closely related to devising a "privacy preserving mechanism" for answering statistical queries to the data, whereas the design of such mechanisms is related to complexity theory. In particular, it was shown that if one-way functions exist, then false discoveries cannot be prevented when the researcher makes more than $\widetilde{O}(n^2)$ queries to a sample of size n.

Additional surveys of wide areas. In addition to the aforementioned survey on fine-grained complexity, the meeting featured a large number of surveys of wide areas. These included:

- *A survey on lower bounds for low-depth arithmetic circuits.* The survey, presented by Neeraj Kayal, visited some of the main themes and techniques in this area, starting from the observation that sufficiently strong lower bound on the size of depth four circuits would yield such lower bounds for general arithmetic circuits (of unbounded depth).
- *Two surveys of recent directions in communication complexity.* The first survey, given by Mark Braverman, focused on the gap between the total length of the messages exchanged between two parties and the information contents of their interaction, raising the question of the extent by which an interactive communication can be compressed to its information contents. It is known that, in general, the best compression is to an exponential amount, but a quadratic amount is possible when the distribution of each input is independent of the distribution of the other input.

  The relation between multi-party communication complexity and distributed computing was the focus of Rotem Oshman's presentation, which highlighted the difference between the "local" model (where messages of unbounded length are allowed in each round) and the "congest" model (in which only short messages are allowed in each round). In both models, in each round, each party can only communicate with its neighbors in the fixed communication network.
- *Machine learning and complexity theory.* Rocco Servedio surveyed some of the known algorithms and lower bounds on the complexity of machine learning. He concluded his presentation suggesting to lower the expectations; that is, aim at better-than-obvious algorithms rather at algorithms that meet or approach the information-theoretic bound.

- *Random CSP instances and complexity theory.* Ryan O'Donnell surveyed the state-of-art regarding the complexity of solving random CSP instances, focusing on the use of the conjecture that it is hard to solve random instances of density that is close to the satisfiability threshold.

Informal specialized sessions. In addition to the formal plenary program, intense interaction between the participants took place in smaller groups. Part of these took place in the form of specialized sessions, some of which were already mentioned above. Other specialized sessions featured the following presentations.

- Amir Shpilka provided an inspiring exposition of a very recent construction of a (deterministic) quasi-NC algorithm for the bipartite matching problem. The said result by Fenner, Gurjar, and Thierauf was posted on ECCC a few days before the meeting (see TR15-177).
- Peter Bürgisser organized a specialized session on geometric complexity theory. This started by an outline of the geometric complexity theory program by him and then was followed by a report of Christian Ikenmeyer on recent advances in our understanding of the complexity of Kronecker coefficients. Klim Efremenko sketched the main ideas of his result on the limits of the method of shifted partial derivatives, which lead to an intense discussion with Neeraj Kayal and Pascal Koiran.
- In a specialized session on Coding theory, Venkatesan Guruswami reported recent advances on recovery of Reed-Solomon codes, and Amir Shpilka showed that Reed-Muller codes achieve the capacity of certain channels and gave a decoding algorithm from random errors in these codes.
- Or Meir organized a special session on open problems in Boolean circuit complexity. He presented a open question, which asks whether solving a computational problem on one of several distinct instances is easier than solving a single instance (this question is a close variant of a question posed by Beimel, Ben-Daniel, Kushilevitz, and Weinreb). Pascal Koiran presented an open problem in arithmetic circuit complexity which concerns finding an explicit polynomial that is hard to compute by polynomials of very restricted form. Oded Goldreich presented a line of research that concerns derandomization of randomized algorithms with very small error, and in particular, with respect to constant-depth circuits. Prasad Raghavendra presented an observation regarding a connection between the circuit complexity of a function and the properties of related polytopes.
- Avi Wigderson described (including extensive historical comments) his recent deterministic polynomial time algorithm for noncommutative rational identity testing (with Garg, Gurvits, and Oliveira). He highlighted the fact that questions and methods from very different origins (including invariant and representation theory, quantum information theorem, and optimization) interconnect and naturally combine for the solution of this problem.
- Boaz Barak talked about a Sum-of-Squares lower bound for the planted clique problem. He outlined the ideas behind a work in progress which

has still not been fully verified (with Sam Hopkins, Jon Kelner, Pravesh Kothari, Ankur Moitra and Aaron Potechin) showing that for every constant degree d and $\epsilon > 0$, the degree d Sum-of-Squares algorithm cannot certify that a random Erős-Rényi graph on n vertices does not contain a clique of size $n^{1/2-\epsilon}$.

- Separations in query and communication complexity: The aim of this session was to showcase some striking recent results (both for their strength and simplicity) giving separations, often tight, between various notions of query complexity for decision trees, and the surprising lifting of these bounds to similar separations between various models of communication complexity. These constitute progress on 30 year old questions in complexity theory. The first talk (given by Venkat Guruswami) discussed results for query complexity and the second talk (given by Raghu Meka) discussed the lifting approach for rectangle based measures of communication complexity.
  - The first talk was titled "Pointer Functions and Query complexity," given by Venkat Guruswami. It discussed a clever Boolean function construction of Goos-Pitassi-Watson which gives an optimal separation between nondeterministic and unambiguous decision tree complexities. It then discussed subsequent work by other authors showing that this function was also very useful in giving optimal quadratic separations between randomized and deterministic decision tree complexities and refuting an old 1986 conjecture by Saks and Wigderson on the largest possible gap between these models. The new developments lead to many more separations, such as between quantum query complexity and classical models, but this wasn't discussed in the talk.
  - Raghu Meka described the general method to transform query lower bounds into communication lower bounds for "composed functions". This is based on his recent works (with Mika Goos, Shachar Lovett, Thomas Watson, and David Zuckerman, and with Pravesh Kothari and Prasad Raghavendra). He presented ideas of the main structure theorem, which states that each rectangle in the communication matrix of the composed function can be simulated by a nonnegative combination of juntas. Consequently, this allows a characterization of the complexity of the composed functions in most known one-sided zero-communication models (capturing NP, co-NP, lower-bound measures such as corruption, smooth-rectangle bound, relaxed partition bound, *etc*) by a corresponding query complexity measure.
- Li-Yang Tan talked about his joint work with Ben Rossman and Rocco Servedio, where they proved an average-case depth hierarchy theorem for Boolean circuits over the standard basis of AND, OR, and NOT gates. The hierarchy theorem says that for every $d \geq 2$, there is an explicit n-variable Boolean function f, computed by a linear-size depth-d formula, which is

such that any depth-$(d-1)$ circuit that agrees with f on $(1/2 + o_n(1))$ fraction of all inputs must have size $\exp(n^{\Omega(1/d)})$. This answers an open question posed by Håstad in his Ph.D. thesis.

- Gil Cohen presented his recent work on improved explicit constructions of Ramsey graphs. Erdős, in 1947 proved the existence of $2\log n$-Ramsey graphs on n vertices, and matching this result with a constructive proof is considered a central problem in combinatorics. The new result achieves an exponential improvement over previous results, and provides explicit $\exp((\log\log n)^c)$-Ramsey graphs.
- Or Meir presented a new proof for a special case of the Karchmer, Raz, and Wigderson conjecture. If this conjecture is proved in full generality, it will imply super-polynomial formula lower bounds which is one of major challenges of the research in circuit complexity. While this case was already proved implicitly in Håstad's work on random restrictions, the new proof uses an entirely different approach based on communication complexity, and seems more likely to be generalizable to other cases of the conjecture.
- Alexander Razborov described recent work on Continuous Combinatorics as well as its context. He noted that Combinatorics was conceived, and then developed over centuries as a discipline about finite structures. However, currently, its applications increasingly pertain to structures that, although finite, are extremely large (e.g., the Internet network, social networks, statistical physics, to name just a few). Moreover, the numerical characteristics that researchers are normally interested in are "continuous" in the sense that small perturbations in the structure do not change the output very much. This makes it very natural to try to think of the "limit theory" of such objects by pretending that "very large" actually means "infinite". It turns out that this mathematical abstraction is very useful and instructive and leads to unexpected connections with many other things, both in mathematics and computer science. Two complementing approaches to constructing such a theory and applying it elsewhere are known as *graph limits* and *flag algebras*, and some of this theory was reviewed.
- Prasad Raghavendra presented exciting new results on lower bounds for linear programs and semidefinite programs based on his work (with Lee and Steurer).
- Zvika Brakerski described progress in the study of Fully Homomorphic Encryption (FHE) in the past couple of years. FHE is an encryption scheme that allows to compute arbitrary function "underneath" the encryption; that is, to go from $\mathrm{Enc}(x)$ to $\mathrm{Enc}(f(x))$ for all f, without any knowledge of the key. This allows to "outsource" computation to a third party without foregoing privacy. In particular, he focused on the "approximate eigenvector approach" based on work by Gentry, Sahai and Waters, and optimizing its performance via "sequentalization" based on joint his work with Vaikuntanathan.

- Raghu Meka described progress on constructing pseudorandom generators for small-space computation. His talk was based on his work with Gopalan and Daniel Kane.
- Zeev Dvir reviewed the Brascamp-Lieb Inequality and described its proof given by Franck Brathe. The proof introduces a normalization technique which allows one to apply a change of basis that puts a set of directions in radial isotropic positions. This technique has found several applications including Foster's sign-rank lower bound and recent work on Sylvester-Gallai type theorems and Locally-Correctable Codes.
- Salil Vadhan, following up on Ryan O'Donnell's plenary survey talk, described the proof of Daniely, Linial, and Shalev-Shwartz that polynomial-time PAC learning of DNF formulas (a long-standing open problem) is impossible if random k-SAT formulas on $n^{f(k)}$ clauses are hard to "refute" for some $f(k) \to \infty$. This led to an intensive small-group discussion on directions for obtaining stronger hardness results, and better understanding the relationship between learning and refutation.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, "US Junior Oberwolfach Fellows". Moreover, the MFO and the workshop organizers would like to thank the Simons Foundation for supporting Luca Trevisan in the "Simons Visiting Professors" program at the MFO.

Workshop: Complexity Theory

Table of Contents