Executive Summary of the Dagstuhl Seminar 08381 "Computational Complexity of Discrete Problems"

Peter Bro Miltersen, Aarhus, Denmark
Rüdiger Reischuk, Lübeck
Georg Schnitger, Frankfurt
Dieter van Melkebeek, Madison, Wisconsin, U.S.A.

Keywords:
computational complexity, discrete problems, Turing machines, circuits, proof complexity, pseudorandomness, derandomization, cryptography, computational learning, communication complexity, query complexity, hardness of approximation

1 Introduction and Goals

Estimating the computational complexity of discrete problems constitutes one of the central and classical topics in the theory of computation. Mathematicians and computer scientists have long tried to classify natural families of Boolean relations according to fundamental complexity measures like time and space, both in the uniform and in the nonuniform setting. Several models of computation have been developed in order to capture the various capabilities of digital computing devices, including parallelism, randomness, and quantum interference.

In addition, complexity theorists have studied other computational processes arising in diverse areas of computer science, each with their own relevant complexity measures. Several of those investigations have evolved into substantial research areas, including:

- proof complexity, interactive proofs, and probabilistically checkable proofs (motivated by verification),
- approximability (motivated by optimization),
- pseudorandomness and zero-knowledge (motivated by cryptography and security),
- computational learning theory (motivated by artificial intelligence),
- communication complexity (motivated by distributed computing), and
- query complexity (motivated by databases).

The analysis and relative power of the basic models of computation remains a major challenge, with intricate connections to all of the areas mentioned above. Several lower bound techniques for explicitly defined functions form the basis for results in propositional proof complexity. The structure that underlies the lower bounds has led to efficient sample-based algorithms for learning unknown functions from certain complexity classes. Close connections have been discovered between circuit lower bounds for certain uniform complexity classes and the possibility of efficient derandomization. The discovery of probabilistically checkable proofs for NP has revived the area of approximability and culminated in surprisingly tight hardness of approximation results for a plethora of NP-hard optimization problems.

Thus, new results that relate or separate different models of computation and new methods for obtaining lower and upper bounds on the complexity of explicit discrete problems are topics of general importance for the computer science community.

The seminar "Computational Complexity of Discrete Problems" has evolved out of the series of seminars entitled "Complexity of Boolean Functions," a topic that has been covered at Dagstuhl on a regular basis since the foundation of IBFI. Over the years, the focus on nonuniform models has broadened to include uniform ones as well. The change in title reflects and solidifies this trend.

A salient feature of the current research in computational complexity is the interpenetration of ideas from different subareas of computational complexity and from other fields in computer science and mathematics. By organizing a generic seminar on computational complexity we aimed to attract researchers from those various subareas and foster further fruitful interactions between them.

2 Organization of the Meeting

41 participants attended this Dagstuhl Seminar. 29 participants presented their results in plenary talks which ranged in length from 20 minutes to 60 minutes. Topics varied from specific problems and their solutions to surveys of larger subareas. In addition, there were many discussions in small groups in between talks as well as after lunch and dinner.

The plenary talks were structured into five morning and three afternoon sessions. Most sessions were focused on a special topic. In one additional evening session all participants were invited to present and discuss open problems informally.

3 Discussed Topics and Achievements

We list some of the major topics that have been discussed during the meeting. Further details as well as additional material can be found in the Abstracts Collection.

Two-Prover Games and Approximability

In a two-prover game, a referee chooses questions (x, y) according to a publicly known distribution, and sends x to the first player and y to the second player. The two players reply by sending $a = a(x)$ and $b = b(y)$ respectively. The players win if a publicly known predicate $V(x, y, a, b)$ holds. The value of the game is the maximally achievable probability of success.

In a parallel repetition the referee generates questions $x = (x_1, ..., x_n), y = (y_1, ..., y_n)$, where each pair (x_i, y_i) is chosen independently according to the original distribution. The players respond by sending $a = (a_1, ..., a_n)$ and $b = (b_1, ..., b_n)$ and win iff $V(x_i, y_i, a_i, b_i)$ holds for every i. The parallel repetition theorem states that for any two-prover game with value $1 - \epsilon$ the value of the game for n repetitions is at most $(1 - \epsilon^3)^{\Omega(n/s)}$, where s is the length of the answer in the original game. On the other hand, the strong parallel repetition conjecture has recently been refuted, in particular the value of the n times repeated game cannot in general be bounded by $(1 - \epsilon)^{\Omega(n/s)}$. The parallel repetition theorem has fundamental applications in communication complexity, quantum computing, algorithmic geometry and in particular in approximation complexity.

Thomas Holenstein reviewed his simplified proof of the parallel repetition theorem. He also discussed the role of the consistent sampling lemma, which is a central tool for the proof of the parallel repetition theorem as well as for the refutation of the strong parallel repetition conjecture.

Oded Regev considered unique games. (In a unique game, for any x, y and any a the predicate $V(x, y, a, b)$ holds for a unique value of b and conversely there is a unique value of a for any given b.) In particular he investigated a semidefinite relaxation of unique games and its behavior under parallel repetitions. Consequences, among others, are further counterexamples to the strong parallel repetition conjecture.

In the topological ordering problem for graphs, vertices of a given directed graph have to be ordered such that as few edges as possible start in a “late” vertex and end in an “early” vertex. By choosing a random ordering at most twice as many edges as required go “backwards”. Venkatesan Guruswami showed that this is essentially the best performance achievable by an efficient algorithm provided the Unique-Games conjecture holds, i.e., that it is NP-hard to distinguish between unique games with value at least $1 - \epsilon$ and unique games with value at most ϵ.

In a constraint satisfaction problem CSP (P) a collection of constraints is given, where each constraint is expressed by a predicate P (applied to a subset of Boolean variables). Johan Hastad considered the problem of approximating the maximal number of simultaneously satisfiable constraints. Call a predicate P approximation resistant if efficient algorithms cannot substantially outperform a random assignment. Assuming the Unique-Games Conjecture, it is shown for instance that a random predicate is approximation resistant with high probability.

Communication Complexity

Troy Lee discussed a norm based framework for showing lower bounds on communication complexity. In particular, the approximation rank of a matrix, one of the most powerful techniques to lower-bound quantum communication complexity, is shown to be approximable in polynomial time. Another consequence is that the logarithm of the approximation rank lower-bounds quantum communication complexity even with entanglement.

Alexander Sherstov described the pattern matrix approach to derive lower bounds on bounded-error and quantum communication complexity. In particular, for an arbitrary function $f : \{0,1\}^{n/4} \rightarrow \{0,1\}$ assign the input x to Alice and a subset $V \subseteq \{1,\dots,n\}$ of size $n/4$ to Bob. It is shown that $f(x|_V)$, i.e., f applied to the projection of x to the bits in V, has bounded-error communication complexity $\Omega(d)$, where d is the approximate degree of f. Among the applications is a new proof of Razborov's quantum lower bounds for all functions of the form $f(x,y) = D(|x \text{ AND } y|)$, where $D : \{0,1,...,n\} \rightarrow \{0,1\}$ is a given function, as well as a large new class of functions whose quantum communication complexity (regardless of prior entanglement) is only polynomially smaller than their classical complexity.

Communication lower bounds for the multiparty number-on-the-forehead model have so far been rather weak and even a separation of randomized and deterministic protocols was unknown. Philipp Woelfel described a first successful separation by a non-constructive argument: there exists a function f with linear deterministic communication complexity, but f has a one-sided error randomized protocol with at most logarithmic communication complexity.

Randomized Computations, Derandomization and Explicit Constructions

The "Walk on Spheres" (WoS) algorithm simulates Brownian Motion in a domain $\Omega \subseteq \mathbb{R}^d$. The WoS algorithm starts at a point $X_0 = x$ in a given bounded domain Ω until it gets ϵ-close to the boundary of Ω. At every step, the algorithm measures the distance d_k from its current position X_k to the boundary of Ω and jumps a distance of $d_k/2$ in a uniformly random direction from X_k to obtain X_{k+1}. The algorithm terminates when it reaches a point X_n that is ϵ-close to the boundary of Ω.

It is not hard to see that the algorithm requires at least $\Omega(\log 1/\epsilon)$ steps to converge. Mark Braverman analyzed the number of steps of the WoS algorithms and derived tight bounds for various domains Ω. For instance, whereas $O(\log 1/\epsilon)$ steps suffice for planar domains with connected exterior, $\Theta((1/\epsilon)^{2-4/(d-1)})$ steps may be required for domains $\Omega \subseteq \mathbb{R}^d$ with connected exterior for $d > 3$.

Eli Ben-Sasson described an explicit construction of a seedless disperser for affine subspaces of F_p^n of dimension greater than $2n/5 + 10$ as well as an improved "global" list-decoding algorithm for Reed-Muller codes. Both applications are obtained by observing that functions represented by low-degree multivariate polynomials in $F_p[x_1,\dots,x_n]$, when viewed as univariate polynomials over F_{p^n}, have a very special structure.

Ronen Shaltiel described unconditional weak derandomization results for communication games, constant depth circuits and streaming algorithms. In particular deterministic solutions are explicitly constructed which simulate a given randomized solution for most inputs using “roughly the same complexity”. For instance, in the context of constant depth circuits deterministic circuits of polynomial size are constructed improving on a classical result by Nisan and Wigderson. As in the conventional approach randomness is extracted from the input, but this time seedless randomness extractors are used rather than seeded extractors.

Amir Shpilka gave an explicit construction of a small ϵ-net S for the family of linear threshold functions. In particular, for any linear threshold function $f : \{-1,+1\}^n \rightarrow \{-1,+1\}$, if $f(x) = 1$ with probability at least ϵ, then there is a vector $x \in S$ with $f(x) = 1$. Up to a polynomial factor, the size of S is comparable to the size achieved by random sets. The construction uses tools such as k-wise independent distributions, random walks on expander graphs and families of perfect hash functions.

Amnon Ta-Shma proposed a generalization of the zig-zag graph product of Reingold, Vadhan and Wigderson: the generalized product combines a large graph and several small graphs instead of just a single small graph. The new product gives a fully-explicit combinatorial construction of regular graphs with an almost optimal spectral gap improving on the Reingold, Vadhan and Wigderson construction.

The Complexity of Non-Uniform Computations

Ordered binary decision diagrams (OBDDs) are one of the most common data structures for Boolean functions. Beate Bollig analyzed the required size of OBDDs which determine the most significant bit of the product of two n-bit integers. Lower and upper bounds are presented. In particular it is shown that exponential size in n is required, answering a question posed by Ingo Wegener.

Kristoffer Arnsfelt Hansen showed how to transform an AC^0 circuit, augmented by a bottom layer of modular counting gates, into an equivalent randomized depth-2 circuit of quasipolynomial size consisting of a modular counting gate or threshold-style gate at the top and a bottom layer of modular counting gates. Thus, in this particular situation constant depth can be reduced to two at the cost of a moderate increase in size.

Stasys Jukna considered unbounded fanin circuits of depth two with arbitrary Boolean functions as gates and showed that $\Omega(n^3)$ wires are necessary to multiply two $n \times n$ matrices. Previously only the lower bound $\Omega(n^2 \log n)$ was known.

Direct Product Theorems are formal statements of the intuition “if solving one instance of a problem is hard, then solving multiple instances is even harder”. Valentine Kabanets described uniform direct product theorems which can be applied to uniform models of computation (e.g., randomized algorithms), whereas most previous direct product theorems applied only to nonuniform models (e.g., circuits).

Michal Koucky observed that the Natural Proof framework of Razborov and Rudich does not apply to a separation of, say, TC^0 and NC^1 and hence there is no known barrier to achieve such a separation by existing “natural approaches”. The reason is that showing that certain self-reducible problems in NC^1 have no polynomial size TC^0-circuits is equivalent to showing that they have no TC^0-circuit of size $n^{1+\epsilon}$ for every $\epsilon > 0$.

Jakob Nordström investigated tradeoffs between space and length of resolution proofs. In particular a strongest possible separation is shown by exhibiting CNF formulae with proofs of linear length but space consumption $\Omega(n/\log n)$. Hence length and space are uncorrelated.

Computational Learning Theory

A stegosystem embeds secret messages into unsuspicious covertexts. In order to hide secret messages reliably, a stegosystem has to draw samples from a covertext source. Rüdiger Reischuk assumes a model where both stegoencoder and attacker have identical information on the covertext distribution; thus both have to learn the distribution by sampling. It is investigated how algorithmic learning techniques can be used to design secure, reliable and computationally efficient stegosystems. Positive results are obtained for covertext channels with simple descriptions and for pseudorandom channels.

Amir Shpilka investigated depth-3 arithmetic circuits, i.e., circuits computing sums of products of linear functions. Given oracle access to an unknown polynomial p of this type, a depth-3 circuit for p is to be found. It is shown that such a circuit can be constructed in time exp(polylog(s)), if p is computable by a depth-3 circuit of size s.

Stephan Waack described a generalization of the classification noise model for PAC learning to allow input-dependent noise. The model is applied for the problem of discriminating between protein-protein interfaces and random sets of pairs of protein surface residues.

Algebraic Computations

Markus Bläser constructed an identity test for polynomials in the black box model with an asymptotically optimal randomness consumption. In particular he obtained an efficient construction of a hitting set generator against the class of polynomials (of fixed degree in each of the variables), where the seed length of this generator is almost optimal.

Christopher Umans presented randomized algorithms for factoring degree n univariate polynomials over the p-element field F_p. His approach is asymptotically faster than the best previous algorithms, provided $\log p < n$ and matches the asymptotic running time of the best known algorithms for $\log p \geq n$. The improvements come from new algorithms for modular composition of degree n univariate polynomials.

Data Streams

Anna Gal investigated deterministic algorithms that make a constant number of passes over the input and give a constant factor approximation of the length of the longest increasing subsequence in a sequence of length n. It is shown that any such algorithm must use space $\Omega(\sqrt{n})$. The proof is based on deriving a direct sum property for a related communication problem.

Jaikumar Radhakrishnan considered the problem of finding a repeated element in a data stream of n elements, where the elements belong to a set of size m. A randomized algorithm is described which finds a repeated element in a single pass and uses space bounded by $O((\log m)^4)$. This is the first sub-linear space one-pass algorithm for this problem.

Further Topics

Martin Dietzfelbinger reduced the analysis of a “blind” search strategy for minimizing continuous functions over $[0,1]$ to the analysis of a one-dimensional random process: a token is initially placed in a random position p within the set $A = \{0, 1, \ldots, n\}$. In one step the token is moved to the new position $p = p - d$, where d is chosen according to a distribution μ. It is shown that the expected number of steps to reach position 0 is $\Theta((\log n)^2)$ for the best distribution μ.

Lance Fortnow talked about a topic in computational game theory. The notion of program equilibria by Tennenholtz is extended to allow for universal simulation; in particular, the players are allowed to be arbitrary Turing machines and payoffs based on running time are discounted. It is shown that several properties of the Tennenholtz model are preserved, for instance cooperation between players is still possible in the Prisoner’s Dilemma game.

A central open problem in descriptive complexity theory is the question whether there exists a logic capturing PTIME. Martin Grohe showed that fixed-point logic with counting captures polynomial time on all classes of graphs with excluded minors. As a consequence, the Weisfeiler-Lehman algorithm can be used as a polynomial time isomorphism test for graph classes with excluded minors.

Miroslaw Kutylowski investigated the degree of anonymity offered by communication protocols. He also described the topic of authentication mechanisms for low-end devices.

4 Conclusion

Understanding the complexity of discrete problems is one of the fundamental tasks in the theory of computation. There has been significant progress in understanding parallel repetitions of two-prover games and in utilizing the Unique-Games conjecture within approximation complexity. Methods are being constantly refined to determine the communication complexity of randomized or quantum protocols. Progress in explicit constructions of combinatorial objects and in derandomization is continuing. However there is still a long way to go until tight resource lower bounds for computations in non-trivial models will be obtained.