Sensitivity, Query Complexity, Communication Complexity
and
Fourier Analysis of Boolean Function

February 19-21, 2020

Indian Statistical Institute
203, B.T. Road, Kolkata - 700108

Organized by

Advanced Computing & Microelectronics Unit
Indian Statistical Institute, Kolkata

Foreword

Advanced Computing & Microelectronics Unit (ACMU), Indian Statistical Institute, Kolkata is organizing a workshop on "Sensitivity, Query Complexity, Communication Complexity and Fourier Analysis of Boolean Function" during 19-21 February 2020 at the Indian Statistical Institute, Kolkata Campus. This area has seen a lot of interesting advances in recent times. In fact in 2018 a major conjecture in this area was solved. The idea of the workshop is to bring together all the researchers working/interested in this area of research and discuss the latest results in this field.

Contents

List of Speakers

Yuval Filmus	Technion, Haifa, Israel
Nitin Saurabh	Technion, Haifa, Israel
Swagato Sanyal	IIT Kharagpur, India
Anupa Sunny	IRIF, Paris, France
Rajat Mittal	IIT Kanpur, India
Arkadev Chattopadhyay	TIFR, Mumbai, India
Rahul Jain	NUS/CQT, Singapore
Suhail Sherif	TIFR, Mumbai, India
Manaswi Paraashar	ISI Kolkata, India
Amey Bhangale	Simons Institute, Berkeley, USA
Diptarka Chakraborty	NUS, Singapore
Arnab Bhattacharya	NUS, Singapore
Prahladh Harsha	TIFR, Mumbai, India

Program

Day 1: Wednesday 19th February 2020

- 10:00-10:15 Introductory talk
- 10:15-11:15 Talk by Yuval Filmus : AND testing
- 11:45-12:45 Talk by Nitin Saurabh : Improved upper bounds on Fourier entropy
- 12:45-14:30 Lunch
- 14:30-15:30 Talk by Prahladh Harsha : High Dimensional Expanders: Introduction and some recent applications
- 16:00-17:00 Talk by Anupa Sunny : Sensitivity lower bounds from linear dependencies

Day 2: Thursday 20th February 2020

- 10:00-11:00 Talk by Rajat Mittal : Exact degree of a symmetric Boolean function
- 11:30-12:30 Talk by Arkadev Chattopadyay : Equality Alone Does Not Simulate Randomness
- 12:30-13:00 Photo and trip to geology museum
- 13:00-14:30 Lunch
- 14:30-15:30 Talk by Rahul Jain : Partition Bound Is Quadratically Tight for Product Distributions
- 16:00-17:00 Talk by Suhail Sherif : Refuting the Log-Approximate-Rank Conjecture
- 17:00-18:00 Talk by Manaswi Paraashar : Quantum Query-to-Communication Simulation Needs a Logarithmic Overhead

Day 3: Friday 21st February 2020

- 10:00-11:00 Talk by Amey Bhangale : Hardness of approximating hypergraph coloring
- 11:30-12:30 Talk by Diptarka Chakraborty : Hardness of Approximation of (Multi-) LCS over Small Alphabet
- 12:30-13:30 Talk by Arnab Bhattacharyya : New approaches for proving lower bounds on the length of locally decodable codes
- 13:30-15:30 Lunch
- 15:30-16:30 Talk by Swagato Sanyal : On decision tree complexity of Boolean functions
- 16:30-16:45 Concluding Remarks

Abstracts

Day 1: 19th February, 2020

10:15 a.m. - 11:15 a.m.

- **Title :** AND testing
- **Speaker :** Yuval Filmus
- **Affiliation :** Technion, Haifa, Israel

Abstract : A classical result in Boolean function analysis states that a Boolean function satisfying $f(x+y) = f(x) + f(y)$ mod 2 for most inputs is close to an XOR. We show that a function satisfying $f(xy) = f(x)f(y)$ for most inputs is close to an AND.

Joint work with Dor Minzer (IAS) and Elchanan Mossel (MIT).

11:45 a.m. - 12:45 p.m.

- **Title :** Improved upper bounds on Fourier entropy
- **Speaker :** Nitin Saurabh
- **Affiliation :** Technion, Haifa, Israel

Abstract : Given a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, define the Fourier distribution to be the distribution on subsets of $[n]$, where each subset S of $[n]$ is sampled with probability $f^(S)^2$. The Fourier Entropy-Influence (FEI) conjecture of Friedgut and Kalai (1996) seeks to relate two fundamental measures associated with the Fourier distribution: there exist a universal constant $C > 0$ such that Shannon entropy of the Fourier distribution of f is at most C times the total influence of f.

In this talk I will present an upper bound of average unambiguous (parity)-certificate complexity on the Fourier entropy of f. This result can be seen as a relaxation of the FEI conjecture where we upper bound Fourier entropy by combinatorial quantities larger than Influence.

Joint work with S. Arunachalam, S. Chakraborty, M. Koucky, and R. de Wolf.

14:30 p.m. - 15:30 p.m.

- **Title :** High Dimensional Expanders: Introduction and some recent applications
– **Speaker :** Prahladh Harsha
– **Affiliation :** TIFR, Mumbai, India

Abstract : Expander graphs, over the last few decades, have played a pervasive role in almost all areas of theoretical computer science. Loosely, speaking an expander graph is an extremely well-connected graph despite being sparse. Recently, various high-dimensional analogues of these objects have been studied in mathematics and even more recently, there have been some surprising applications in computer science, especially in the area of property testing, coding theory and approximate counting. In this talk, I'll give a high-level introduction to these high-dimensional expanders (HDXs). In particular, we will view them through the perspective of random walks on graphs. Time permitting, we will then some see applications of these HDXs towards property testing, coding theory and matroid counting.

16:00 p.m. - 17:00 p.m.

- **Title :** Sensitivity lower bounds from linear dependencies
- **Speaker :** Anupa Sunny
- **Affiliation :** IRIF, Paris, France

Abstract : Recently, using spectral techniques, H. Huang proved that every subgraph of the hypercube of dimension n induced on more than half the vertices has maximum degree at least $\sqrt{n}$. Combined with some earlier work, this completed a proof of the sensitivity conjecture. In this talk, a proof of Huang's result will be shown using only linear dependency and independence of vectors associated with the vertices of the hypercube. This approach leads to several improvements of the result. In particular, we prove that that in any induced subgraph of H_n with more than half the number of vertices, there are two vertices, one of odd parity and the other of even parity, each with at least n vertices at distance at most 2. As an application we show that for any Boolean function f, the polynomial degree of f is bounded above by $s_0(f)s_1(f)$, a strictly stronger statement which implies the sensitivity conjecture.

Based on joint work with Sophie Laplante and Reza Naserasr.

Day 2: 20th February, 2020

10:00 a.m. - 11:00 a.m.

- **Title :** Exact degree of a symmetric Boolean function
- **Speaker :** Rajat Mittal
- **Affiliation :** IIT Kanpur, India

Abstract : Symmetric functions are one of the simplest subclass of Boolean functions. A natural question is, what is the minimum possible degree of a non-constant symmetric Boolean function? Unfortunately, we don't know the exact answer for this natural question. This talk will cover some of the earlier results and different formulations to attack this problem.

We will look at the result of Gathen and Roche (97), showing that the degree is at least $n - O(n^{0.525})$ and almost every such function has degree n. Then, we will see different approaches to this problem using the method of dual witness and discrete derivatives.

11:30 a.m. - 12:30 p.m.

- **Title :** Equality Alone Does Not Simulate Randomness
- **Speaker :** Arkadev Chattopadhyay
- **Affiliation :** TIFR, Mumbai, India

Abstract : What precisely is the power of randomness? This is a central question in the theory of computing. We instantiate this question in Yao's classical 2-party model of communication. It is folklore that the simple function Equality does not have efficient deterministic protocols but has very cheap randomised ones. Several other functions like Greater-Than which are deterministically hard, have efficient randomised solutions exploiting the protocol for Equality. Do deterministic protocols equal in power to randomised ones, once they are augmented with the

ability to solve Equality? We answer this question completely and raise several questions which one needs to answer to understand randomness.

Joint work with Shachar Lovett and Marc Vinyals.

14:30 p.m. - 15:30 p.m.

- **Title :** Partition Bound Is Quadratically Tight for Product Distributions

– **Speaker :** Rahul Jain

– **Affiliation :** NUS/CQT, Singapore

Abstract : Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a 2-party function. For every product distribution μ on $\{0,1\}^n \times \{0,1\}^n$, we show that $CC^{\mu}_{0.49}(f) = O((log\, prt_{\frac{1}{8}}(f) \cdot \, loglog\, prt_{\frac{1}{8}}(f))^2)$, where $CC^{\mu}_{\varepsilon}(f)$ is the distributional communication complexity of f with error at most ε under the distribution μ and $prt_{\frac{1}{8}}(f)$ is the partition bound of f, as defined by Jain and Klauck [Proc.25th CCC, 2010]. We also prove a similar bound in terms of $IC_{\frac{1}{8}}(f)$, the information complexity of f, namely,

$$CC^{\mu}_{0.49}(f) = O((IC_{\frac{1}{8}}(f) \cdot log\, IC_{\frac{1}{8}}(f))^2)$$

The latter bound was recently and independently established by Kol [Proc. 48th STOC, 2016] using a different technique. We show a similar result for query complexity under product distributions. Let $g : \{0,1\}^n \to \{0,1\}$ be a function. For every bit-wise product distribution μ on $\{0,1\}^n$, we show that $QC^{\mu}_{0.49}(g) = O((log\, qprt_{\frac{1}{8}}(g) \cdot \, loglog\, qprt_{\frac{1}{8}}(g))^2)$, where $QC^{\mu}_{\varepsilon}(g)$ is the distributional query complexity of f with error at most ε under the distribution μ and $qprt_{\frac{1}{8}}(g))$ is the query partition bound of the function g. Partition bounds were introduced (in both communication complexity and query complexity models) to provide LP-based lower bounds for randomized communication complexity and randomized query complexity. Our results demonstrate that these lower bounds are polynomially tight for product distributions.

Joint work with Prahladh Harsha (TIFR) and Jaikumar Radhakrishnan (TIFR)

16:00 p.m. - 17:00 p.m.

- **Title :** Refuting the Log-Approximate-Rank Conjecture
– **Speaker :** Suhail Sherif
– **Affiliation :** TIFR, Mumbai, India

Abstract : Communication complexity is an interesting computational model. It is rich enough to allow interesting protocols, yet it is a model for which we can prove lower bounds. This makes it a good testing ground to analyze, for instance, the power of randomness. It also has a very wide range of applications in other areas of theoretical computer science.

One of the most important open problems in the area is the Log-Rank Conjecture, which posits that there are protocols that bring the deterministic communication complexity of a function all the way down to polynomial in the logarithm of the rank of the function matrix. The Log-Approximate-Rank Conjecture is the analogous conjecture for randomized communication protocols and it happens to imply the Log-Rank Conjecture.

In this talk, we will take a closer look at these conjectures. We will then show a simple function that refutes the Log-Approximate-Rank Conjecture along with a slew of other conjectures. The rough reason behind these refutations is that rank is subadditive, and we show that randomized communication complexity can not emulate this behaviour.

Joint work with Arkadev Chattopadhyay and Nikhil Mande.

17:00 p.m. - 18:00 p.m.

- **Title :** Quantum Query-to-Communication Simulation Needs a Logarithmic Overhead

– **Speaker :** Manaswi Paraashar

– **Affiliation :** ISI Kolkata, India

Abstract : Buhrman, Cleve and Wigderson (STOC'98) observed that every Boolean function $F = (f \circ g)$, where $f : \{-1,1\}^n \to \{-1,1\}$ and $g : \{-1,1\}^2 \to \{-1,1\}$, has two-party bounded-error quantum communication complexity $O(Q(f) logn)$, where $Q(f)$ is the bounded-error quantum query complexity of f. Note that the bounded-error randomized communication complexity of $(f \circ g)$ is bounded by $O(R(f))$, where $R(f)$ denotes the bounded-error randomized query complexity of f. Thus, the BCW simulation has an extra $O(logn)$ factor appearing that is absent in classical simulation. A natural question is if this factor can be avoided. Hoyer and de Wolf (STACS'02) showed that for the Set-Disjointness function, this can be reduced to c^{log*n} for some constant c, and subsequently Aaronson and Ambainis (FOCS'03) showed that this factor can be made a constant. That is, the quantum communication complexity of the Set-Disjointness function (which is $(NOR_n \circ AND_2)$) is $O(Q(NOR_n))$.

Perhaps somewhat surprisingly, we show that when $g = XOR_2$, then the extra $(logn)$ factor in the BCW simulation is unavoidable. In other words, we exhibit a total function $f : \{-1,1\}^n \to \{-1,1\}$ such that quantum communication complexity of $(f \circ XOR_2)$ is $\Theta(Q(f) \log n)$.

To the best of our knowledge, it was not even known prior to this work whether there existed total functions f, g, where g is a 2-input function, such that quantum communication complexity of $(f \circ g)$ is $\omega(Q(f))$.

Joint work with Sourav Chakraborty, Arkadev Chattopadhyay and Nikhil Mande.

Day 3: 21th February, 2020

10:00 a.m. - 11:00 a.m.

- **Title :** Hardness of approximating hypergraph coloring

– **Speaker :** Amey Bhangale

– **Affiliation :** Simons Institute, Berkeley, USA

Abstract : A c-coloring of a k-uniform hypergraph is a coloring of the vertices with c colors such that no hyperedge is monochromatic. Deciding if a given hypergraph is c-colorable or not is NP-hard, even for c = 2 (unlike the graph case). In this talk, I will talk about various hardness of approximation results on hypergraph coloring. The analysis of such reductions relies heavily on the Fourier analysis of Boolean functions. I will also discuss our recent result which gives a simple alternate proofs of the previously know NP-hardness results, using the upper bounds on the maximum size of intersecting family.

Joint work with Per Austrin and Aditya Potukuchi.

11:30 a.m. - 12:30 p.m.

- **Title :** Hardness of Approximation of (Multi-)LCS over Small Alphabet

– **Speaker :** Diptarka Chakraborty

– **Affiliation :** NUS, Singapore

Abstract : The problem of finding longest common subsequence (LCS) is one of the fundamental problems in computer science, which finds application in fields such as computational biology, text processing, information retrieval, data compression etc. It is well known that the problem of finding LCS of an arbitrary number of input sequences (which we refer as Multi-LCS problem) is NP-complete. Jiang and Li showed that under the assumption of $P \neq NP$, there exists no constant $\delta > 0$ such that the length of

LCS of arbitrary number of input sequences of length n can be approximated within $n^{-\delta}$-factor in polynomial time [Jiang and Li, SICOMP'95]. However their reduction assumes the alphabet size to be at least $\Omega(n)$. So far no hardness result is known for the problem of approximating (Multi-)LCS over sub-linear sized alphabet. On the other hand, it is easy to get $1/|\Sigma|$-factor approximation for strings of alphabet Σ.

In this work we make significant progress towards proving hardness of approximation over small alphabet by showing a polynomial-time reduction from the well-studied densest k-subgraph problem to approximating Multi-LCS over alphabet of size $poly(n/k)$. As a consequence, from the known hardness result of densest k-subgraph problem (e.g. [Manurangsi, STOC'17]) we get that no polynomial-time algorithm can give any constant factor approximation of Multi-LCS over alphabet of size $n^{o(1)}$ unless the Exponential Time Hypothesis (ETH) is false.

Joint work with Amey Bhangale and Rajendra Kumar.

12:30 p.m - 13:30 p.m.

- **Title :** New approaches for proving lower bounds on the length of locally decodable codes
- **Speaker :** Arnab Bhattacharya
- **Affiliation :** NUS, Singapore

Abstract : A code is called a q-query locally decodable code (LDC) if there is a randomized decoding algorithm that, given an index i and a received word w close to an encoding of a message x, outputs x_i by querying only at most q coordinates of w. Understanding the tradeoffs between the dimension, length and query complexity of LDCs is a fascinating and unresolved research challenge.

I will discuss two new techniques for analyzing the length of LDCs:

(1) For 2-query LDCs, we show that $n > exp(k)$ where n is the length of the LDC and k is its dimension. This was already proven by Kerenedis and de Wolf in 2002. However, unlike their argument, ours is purely elementary and quite short. We rely on a matrix concentration result that can be proven combinatorially.

(2) For "strong" 3-query LDCs, we show that $n > \Omega(k^2/log(k))$ where n is the length of the LDC and k is its dimension. This was already proven by Woodruff in 2007. However, unlike his argument, we use graph-theoretic techniques and do not rely on a direct reduction to 2-query LDCs.

For both of these results, we speculate about how the techniques may generalize.

Joint work with Sunil Chandran, Suprovat Ghoshal, and Sivakanth Gopi.

15:30 p.m. - 16:30 p.m.

- **Title :** On decision tree complexity of Boolean functions
- – **Speaker :** Swagato Sanyal
- – **Affiliation :** IIT Kharagpur, India

Abstract : The decision tree model, also known as the query model, is a simple model of computation. Besides having an intuitive appeal, and offering a natural paradigm for algorithm design, it is often possible to theoretically nail down the exact complexity of many interesting algorithmic tasks using currently available methods, completely and unconditionally; for more complex models of computation such an aspiration is currently a distant dream. In this talk we will first introduce and motivate query model, and then present a recent work of ours on the query complexity of an important subclass of problems, called composed functions.

Joint work with Dmitry Gavinsky, Troy Lee and Miklos Santha.