THEORY OF COMPUTING, Volume 9 (29), 2013, pp. 889–896
www.theoryofcomputing.org

NOTE

SPECIAL ISSUE: ANALYSIS OF BOOLEAN FUNCTIONS

Hypercontractivity Via the Entropy Method

Eric Blais* Li-Yang Tan†

Received January 15, 2013; Revised June 30, 2013; Published December 7, 2013

Abstract: The Hypercontractive Inequality of Bonami (1968, 1970) and Gross (1975) is equivalent to the following statement: for every $q > 2$ and every function $f : \{-1,1\}^n \to \mathbb{R}$ of Fourier degree at most m,

$$\|f\|_q \leq (q-1)^{m/2}\|f\|_2\,.$$

The original proof of this inequality is analytical. Friedgut and Rödl (2001) gave an alternative proof of the slightly weaker Hypercontractive Inequality $\|f\|_4 \leq 28^{m/4}\|f\|_2$ by combining tools from information theory and combinatorics. Specifically, they recast the problem as a statement about multi-hypergraphs, generalized Shearer's lemma, and used probabilistic arguments to obtain the inequality.

We show that Shearer's Lemma and elementary arguments about the entropy of random variables are sufficient to recover the optimal Hypercontractive Inequality for all even integers q.

ACM Classification: G.2, G.3, F.1.3

AMS Classification: 68Q87

Key words and phrases: Boolean functions, harmonic analysis, hypercontractivity, information theory, entropy method

*Research supported by a Simons Postdoctoral Fellowship. Part of this research was completed while the author was a postdoctoral research fellow at Carnegie Mellon University.

†Part of this research was completed while the author was visiting CMU.

 DOI: 10.4086/toc.2013.v009a029

1 Introduction

The Hypercontractive Inequality plays a fundamental role in the analysis of Boolean functions. Discovered independently[1] by Aline Bonami [3, 4] and several years later by Leonard Gross [14], the inequality is concerned with the noise operator T_ρ that acts on functions $f : \{-1,1\}^n \to \mathbb{R}$ via $T_\rho f(x) = \mathbf{E}[f(y)]$, where y is a ρ-correlated copy of x (i. e., y is drawn from the product distribution where $\mathbf{E}[y_i x_i] = \rho$ for all $i \in [n]$). Intuitively, the noise operator smooths f by replacing each $f(x)$ with the average of f's values in a neighborhood around x; this effect is also evident when considering the Fourier expansion

$$T_\rho f = \sum_{S\subseteq[n]} \rho^{|S|}\widehat{f}(S)\chi_S\,, \quad \text{where} \quad \chi_S(x) = \prod_{i\in S} x_i\,.$$

Here we see that each Fourier coefficient $\widehat{f}(S)$ is attenuated by a factor $\rho^{|S|}$ that gets harsher the larger $|S|$ is. The Hypercontractive Inequality quantifies this smoothing effect by showing that applying T_ρ to f allows one to bound its 2-norm by a *smaller* norm:

Hypercontractive Inequality. *Let* $f : \{-1,1\}^n \to \mathbb{R}$ *and* $0 < \rho < 1$. *Then* $\|T_\rho f\|_2 \le \|f\|_{1+\rho^2}$.

The Hypercontractive Inequality is equivalent (via duality, see, e. g., [22]) to the following inequality.

Theorem 1.1. *For any* $f : \{-1,1\}^n \to \mathbb{R}$, *let* $f^{(=m)} = \sum_{|S|=m} \widehat{f}(S)\chi_S$ *denote the projection of* f *to its degree-*m *part. Then for all* $q > 2$,

$$\|f^{(=m)}\|_q \le (q-1)^{m/2} \cdot \|f^{(=m)}\|_2\,.$$

That is, the $2 \to q$ *norm of the projection operator* $\mathcal{P}^{=m}$ *that maps* $f : \{-1,1\}^n \to \mathbb{R}$ *to its degree-*m *part is at most* $(q-1)^{m/2}$. *We denote this as* $\|\mathcal{P}^{=m}\|_{2\to q} \le (q-1)^{m/2}$.

First introduced into theoretical computer science by the celebrated work of Kahn, Kalai, and Linial [15], the Hypercontractive Inequality has seen utility in a surprisingly wide variety of areas, spanning distributed computing, random graphs, k-SAT, social choice, inapproximability, learning theory, metric spaces, statistical physics, convex relaxation hierarchies, etc. [2, 6, 22, 8, 9, 10, 11, 5, 18, 17, 21, 13, 19, 1]. In almost every one of these results there are no known alternate proofs that do not require the use of hypercontractivity. (See de Wolf's survey [23, Sec. 4] and O'Donnell's monograph [20, Ch. 9] for more details on the Hypercontractive Inequality and its applications.)

The well-known analytic proof of the Hypercontractive Inequality proceeds by induction on n. The crux of the inductive step (sometimes referred to as the tensoring property of the Hypercontractive Inequality) is Minkowski's inequality, the triangle inequality for L^p spaces; the base case is a two-point inequality that reduces to standard analytic calculations but is nonetheless technically involved. Considering the ubiquity of the Hypercontractive Inequality in discrete settings, there has been significant interest in obtaining alternative proofs of the inequality.

Ehud Friedgut and Vojtech Rödl [12] obtained one such alternative proof by exploiting a novel connection between the Hypercontractive Inequality and Shannon entropy. Specifically, their main result is an information-theoretic/combinatorial proof of the inequality

$$\|\mathcal{P}^{=m}\|_{2\to 4} \le (\sqrt{28})^{m/2}\,.$$

[1] For a history of the discovery of the Hypercontractive Inequality we refer the reader to the Notes in [20, Ch. 9].

This is a slightly weaker version of the $q = 4$ special case of Theorem 1.1, but it is nonetheless qualitatively sufficient for many applications of the Hypercontractive Inequality, with a slight loss in the corresponding bounds. Friedgut and Rödl obtain this result by recasting the inequality in terms of multi-hypergraphs, invoking a generalization of Shearer's entropy lemma for such hypergraphs, and applying probabilistic arguments to complete the proof.

Friedgut and Rödl's result raises two fundamental questions: Is the combinatorial/information theoretic argument strong enough to recover the optimal value of $\|\mathcal{P}^{=m}\|_{2\to q}$? And, can this result be obtained directly by elementary information theoretic arguments? We give positive answers to both questions: we show that Friedgut and Rödl's argument can be simplified and sharpened by reasoning directly about the Fourier spectrum of f, without requiring the translation to hypergraphs or generalization of Shearer's lemma. With this direct approach, we obtain a simple proof of the optimal Hypercontractive Inequality

$$\|\mathcal{P}^{=m}\|_{2\to 2k} \le (2k-1)^{m/2}$$

for all $k \in \mathbb{N}$ (i. e., Theorem 1.1 for all even integers q) using only elementary information-theoretic facts.

Let us briefly mention two other alternative proofs for special cases of the Hypercontractive Inequality. First, there is a short and elegant inductive proof of the $q = 4$ special case of Theorem 1.1 that requires only the Cauchy-Schwarz inequality. This proof first appeared in the literature in [19], although Bonami's original paper [4] contains a proof along similar lines. This proof, however, does not appear to generalize beyond the $q = 4$ special case. Second, in independent recent work, Kauers *et al.* noted that the two-point inequality in the standard analytic proof of the Hypercontractive Inequality becomes particularly simple when q is an even integer [16]. However, this proof still uses the inductive step as a black-box (i. e., the fact that the Hypercontractive Inequality tensorizes, requiring Minkowski's inequality), and remains very much an analytic proof.

2 Basics of Shannon entropy

Let $\mathbf{X}$ be a (scalar or vector valued) random variable over the discrete sample space Ω and let $p : \Omega \to [0,1]$ be the probability mass function of $\mathbf{X}$. The *entropy* of $\mathbf{X}$ is

$$H(\mathbf{X}) = -\sum_{x\in\Omega} p(x)\log p(x)\,.$$

Here and throughout this note, logarithms are taken to base 2. The *conditional entropy* of $\mathbf{X}$ given $\mathbf{Y}$ is

$$H(\mathbf{X} \mid \mathbf{Y}) = \mathop{\mathbf{E}}_{y}\Big[-\sum_{x} p(x \mid y)\log p(x \mid y)\Big]\,.$$

We use the following basic properties of entropy in our analysis.

Lemma 2.1 (Universal upper bound)**.** *For any random variable* $\mathbf{X}$ *over the sample space* Ω*,*

$$H(\mathbf{X}) \le \log|\mathrm{supp}(\mathbf{X})| \le \log|\Omega|\,.$$

The equality $H(\mathbf{X}) = \log|\Omega|$ *holds if and only if* $\mathbf{X}$ *is uniformly distributed over* Ω*.*

Lemma 2.2 (Chain rule)**.** *The entropy of a sequence* $\mathbf{X}_1,\dots,\mathbf{X}_n$ *of random variables satisfies*

$$H(\mathbf{X}_1,\dots,\mathbf{X}_n)=\sum_{i=1}^{n} H(\mathbf{X}_i \mid \mathbf{X}_1,\dots,\mathbf{X}_{i-1}).$$

Finally we recall Shearer's lemma, a generalization of the subadditivity of entropy. (For a simple proof, see [7].) For a sequence of random variables $\mathbf{X}=(\mathbf{X}_1,\dots,\mathbf{X}_n)$ and a set $S\subseteq[n]$, we write $\mathbf{X}_S$ to denote the projection of $\mathbf{X}$ onto the coordinates in S, i. e., $\mathbf{X}_S=(\mathbf{X}_j)_{j\in S}$.

Lemma 2.3 (Shearer's Lemma)**.** *Let* $\mathbf{X}\in\Omega^n$ *be a vector of random variables and let* $S_1,\dots,S_m\subseteq[n]$ *be a collection of sets that cover each element in* $[n]$ *at least* t *times. Then*

$$H(\mathbf{X})\le \frac{1}{t}\sum_{j=1}^{m} H(\mathbf{X}_{S_j}).$$

3 Hypercontractivity via the entropy method

In this section we prove Theorem 1.1 for all even integers q using the entropy method.

Theorem 3.1 (Special case of Theorem 1.1)**.** *For any* $f:\{-1,1\}^n\to\mathbb{R}$ *and any even integer* $q>2$,

$$\|f^{(=m)}\|_q \le (q-1)^{m/2}\cdot\|f^{(=m)}\|_2\,.$$

Proof. By a limiting argument, it suffices to prove the inequality for $f:\{-1,1\}^n\to\mathbb{Q}$. By homogeneity, we may further assume that the Fourier coefficients of f are integral (i. e., $\widehat{f}(S)\in\mathbb{Z}$ for all $S\subseteq[n]$).

For each $S\in\binom{[n]}{m}$, let W_S denote a set of $|\widehat{f}(S)|$ elements which we call *witnesses* for S. The witness sets for any two distinct sets $S\neq T$ are disjoint, and we write $\mathcal{W}=\bigcup_S W_S$ to denote the disjoint union of all $\binom{n}{m}$ witness sets. We say that q witnesses in $\mathcal{W}$ are a *legal q-tuple* if their corresponding sets $S_1,\dots,S_q\in\binom{[n]}{m}$ satisfy $S_1\triangle\cdots\triangle S_q=\emptyset$.

Let $w=(w_1,\dots,w_q)\in\mathcal{W}^q$ be a random variable drawn uniformly at random from the collection of ordered legal q-tuples in $\mathcal{W}^q$ (where repetitions are allowed) and let $\mathbf{S}=(\mathbf{S}_1,\dots,\mathbf{S}_q)\in\binom{[n]}{m}^q$ be the sets corresponding to the q witnesses in w. By the chain rule, the joint entropy of w and $\mathbf{S}$ satisfies

$$\begin{aligned}
H(w,\mathbf{S})=H(w)+H(\mathbf{S}\mid w) &= \log\left(\sum_{\substack{S_1,\dots,S_q\in\binom{[n]}{m}\\ S_1\triangle\cdots\triangle S_q=\emptyset}} |\widehat{f}(S_1)\dots\widehat{f}(S_q)|\right)+0\\
&\ge \log\left|\sum_{\substack{S_1,\dots,S_q\in\binom{[n]}{m}\\ S_1\triangle\cdots\triangle S_q=\emptyset}} \widehat{f}(S_1)\dots\widehat{f}(S_q)\right| \;=\; \log(\|f^{(=m)}\|_q^q)\,, \qquad (1)
\end{aligned}$$

where the final equality uses the fact that q is even.

Let

$$\mathbf{T} = (\mathbf{T}_{\{1,2\}}, \mathbf{T}_{\{1,3\}}, \ldots, \mathbf{T}_{\{q-1,q\}}) \in \binom{[n]}{\leq m}^{\binom{q}{2}}$$

be a sequence of $\binom{q}{2}$ sets such that for every $i < j \in [q]$,

$$\mathbf{T}_{\{i,j\}} = \left\{x \in [n] \colon x \in (\mathbf{S}_i \cap \mathbf{S}_j) \setminus (\mathbf{S}_{i+1} \cup \cdots \cup \mathbf{S}_{j-1}) \text{ and } \#\{\ell \in [q] \colon \ell < i,\ x \in \mathbf{S}_\ell\} \text{ is even}\right\}.$$

Recall that the symmetric difference of the sets in $\mathbf{S}$ is empty, and so every $x \in [n]$ occurs in an even number of them. Therefore, if an element x satisfies $\{i \in [q] \colon x \in \mathbf{S}_i\} = \{i_1, \ldots, i_k\}$ where $i_1 \leq \ldots \leq i_k$, then x is added to the sets $\mathbf{T}_{\{i_1,i_2\}}, \mathbf{T}_{\{i_3,i_4\}}, \ldots, \mathbf{T}_{\{i_{k-1},i_k\}}$.

Let

$$\mathbf{T}_{\{i,*\}} := (\mathbf{T}_{\{i,j\}})_{j \neq i} = (\mathbf{T}_{\{1,i\}}, \mathbf{T}_{\{2,i\}}, \ldots, \mathbf{T}_{\{i-1,i\}}, \mathbf{T}_{\{i,i+1\}}, \ldots, \mathbf{T}_{\{i,q\}}).$$

This construction guarantees that for every $i \in [q]$, the sets in $\mathbf{T}_{\{i,*\}}$ partition $\mathbf{S}_i$ (since every $x \in \mathbf{S}_i$ is in exactly one set $\mathbf{T}_{\{i,j\}}$ for some $j \neq i$). In particular, $\mathbf{S}$ determines $\mathbf{T}$ and vice-versa so $H(\mathbf{S}) = H(\mathbf{T})$. With a different application of the chain rule to the joint entropy of w and $\mathbf{S}$, we obtain

$$H(w, \mathbf{S}) = H(\mathbf{S}) + H(w \mid \mathbf{S}) = H(\mathbf{T}) + H(w \mid \mathbf{S}). \tag{2}$$

The chain rule also yields $H(w \mid \mathbf{S}) = \sum_{i=1}^{n} H(w_i \mid w_1, \ldots, w_{i-1}, \mathbf{S}_1, \ldots, \mathbf{S}_n)$. Given $\mathbf{S}_i$, the random variable w_i is uniformly distributed among the witnesses for $\mathbf{S}_i$ and, in particular, is independent of w_j and $\mathbf{S}_j$ for each $j \neq i$. So the conditional entropy of w given $\mathbf{S}$ satisfies

$$H(w \mid \mathbf{S}) = \sum_{i=1}^{q} H(w_i \mid \mathbf{S}_i). \tag{3}$$

The random variables $\mathbf{T}_{\{1,*\}}, \ldots, \mathbf{T}_{\{q,*\}}$ cover each set in $\{\mathbf{T}_{\{i,j\}}\}_{i \neq j}$ twice, so by Shearer's Lemma

$$H(\mathbf{T}) \leq \frac{1}{2} \sum_{i=1}^{q} H(\mathbf{T}_{\{i,*\}}). \tag{4}$$

Since $\mathbf{T}_{\{i,*\}}$ is a partition of $\mathbf{S}_i$, $H(\mathbf{T}_{\{i,*\}}) = H(\mathbf{S}_i, \mathbf{T}_{\{i,*\}})$. Also, there are at most $(q-1)^m$ possible ordered partitions of the m elements of $\mathbf{S}_i$ into $q-1$ parts so the universal upper bound yields

$$H(\mathbf{T}_{\{i,*\}} \mid \mathbf{S}_i) \leq m \log(q-1).$$

These two observations imply that

$$H(\mathbf{T}_{\{i,*\}}) = H(\mathbf{S}_i, \mathbf{T}_{\{i,*\}}) = H(\mathbf{S}_i) + H(\mathbf{T}_{\{i,*\}} \mid \mathbf{S}_i) \leq H(\mathbf{S}_i) + m \log(q-1). \tag{5}$$

We can now combine (2)–(5) to obtain the upper bound

$$H(w, \mathbf{S}) \leq \frac{qm}{2} \log(q-1) + \frac{1}{2} \sum_{i=1}^{q} \Big(H(\mathbf{S}_i) + 2H(w_i \mid \mathbf{S}_i) \Big). \tag{6}$$

Finally, we observe that $H(\mathbf{S}_i)+2H(w_i \mid \mathbf{S}_i) = H(w_i, w_i', \mathbf{S}_i)$ where w_i, $\mathbf{S}_i$ are as above, and w_i' is another witness of $\mathbf{S}_i$ chosen uniformly at random from $W_{\mathbf{S}_i}$. By the universal upper bound,

$$H(\mathbf{S}_i)+2H(w_i \mid \mathbf{S}_i) = H(w_i, w_i', \mathbf{S}_i) \leq \log\left(\sum_{S\in\binom{[n]}{m}} |\widehat{f}(S)|^2\right) = \log\|f^{(=m)}\|_2^2. \tag{7}$$

Combining (1), (6), and (7) and rearranging completes the proof. □

Remark 3.2. For $f:\{-1,1\}^n \to \mathbb{R}$, define $f^{(\leq m)} = \sum_{|S|\leq m}\widehat{f}(S)\chi_S$. The argument above (with only minor changes in the choice of random variables) can also be used to show that

$$\|f^{(\leq m)}\|_q \leq (q-1)^{m/2}\cdot\|f^{(\leq m)}\|_2$$

for every even integer $q>2$, a slight generalization of Theorem 3.1.

Acknowledgements

We thank Ryan O'Donnell, Rocco Servedio, and Andrew Wan for helpful discussions. We also thank Elchanan Mossel and the anonymous referees whose comments helped improve the presentation.

References

[1] BOAZ BARAK, FERNANDO G. S. L. BRANDÃO, ARAM W. HARROW, JONATHAN A. KELNER, DAVID STEURER, AND YUAN ZHOU: Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. 44th STOC*, pp. 307–326. ACM Press, 2012. [doi:10.1145/2213977.2214006] 890

[2] MICHAEL BEN-OR AND NATHAN LINIAL: Collective coin flipping. In SILVIO MICALI AND FRANCO PREPARATA, editors, *Randomness and Computation*, volume 5 of *Advances in Computing Research: A research annual*, pp. 91–115. JAI Press, 1989. Preliminary version in FOCS'85. 890

[3] ALINE BONAMI: Ensembles $\Lambda(p)$ dans le dual de D^∞. *Annales de l'Institut Fourier*, 18(2):193–204, 1968. EuDML. 890

[4] ALINE BONAMI: Étude des coefficients Fourier des fonctions de $L^p(G)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970. EuDML. 890, 891

[5] JEAN BOURGAIN: On the distribution of the Fourier spectrum of Boolean functions. *Israel J. Math.*, 131(1):269–276, 2002. [doi:10.1007/BF02785861] 890

[6] JEAN BOURGAIN, JEFF KAHN, GIL KALAI, YITZHAK KATZNELSON, AND NATHAN LINIAL: The influence of variables in product spaces. *Israel J. Math.*, 77(1-2):55–64, 1992. [doi:10.1007/BF02808010] 890

[7] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Wiley, 1991. 892

[8] URIEL FEIGE AND JOE KILIAN: Zero knowledge and the chromatic number. *J. Comput. System Sci.*, 57(2):187–199, 1998. Preliminary version in CCC'96. [doi:10.1006/jcss.1998.1587] 890

[9] EHUD FRIEDGUT: Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998. [doi:10.1007/PL00009809] 890

[10] EHUD FRIEDGUT: Sharp thresholds of graph properties, and the k-SAT problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. AMS. 890

[11] EHUD FRIEDGUT, GIL KALAI, AND ASSAF NAOR: Boolean functions whose Fourier transform is concentrated on the first two levels. *Advances in Applied Mathematics*, 29(3):427–437, 2002. [doi:10.1016/S0196-8858(02)00024-6] 890

[12] EHUD FRIEDGUT AND VOJTECH RÖDL: Proof of a hypercontractive estimate via entropy. *Israel J. Math.*, 125(1):369–380, 2001. [doi:10.1007/BF02773387] 890

[13] CHRISTOPHE GARBAN, GÁBOR PETE, AND ODED SCHRAMM: The Fourier spectrum of critical percolation. *Acta Mathematica*, 205(1):19–104, 2010. [doi:10.1007/s11511-010-0051-x] 890

[14] LEONARD GROSS: Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. JSTOR. 890

[15] JEFF KAHN, GIL KALAI, AND NATHAN LINIAL: The influence of variables on boolean functions (extended abstract). In *Proc. 29th FOCS*, pp. 68–80. IEEE Comp. Soc. Press, 1988. [doi:10.1109/SFCS.1988.21923] 890

[16] MANUEL KAUERS, RYAN O'DONNELL, LI-YANG TAN, AND YUAN ZHOU: Hypercontractivity inequalities via SOS, and the Frankl-Rödl graph. Technical report, 2012. To appear in SODA'14. [arXiv:1212.5324] 891

[17] SUBHASH KHOT, GUY KINDLER, ELCHANAN MOSSEL, AND RYAN O'DONNELL: Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007. Preliminary version in FOCS'04. See also at ECCC. [doi:10.1137/S0097539705447372] 890

[18] SUBHASH KHOT AND NISHEETH K. VISHNOI: The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proc. 46th FOCS*, pp. 53–62. IEEE Comp. Soc. Press, 2005. [doi:10.1109/SFCS.2005.74] 890

[19] ELCHANAN MOSSEL, RYAN O'DONNELL, AND KRZYSZTOF OLESZKIEWICZ: Noise stability of functions with low influences: invariance and optimality. *Ann. of Math.*, 171(1):295–341, 2010. Preliminary version in FOCS'05. [doi:10.4007/annals.2010.171.295] 890, 891

[20] RYAN O'DONNELL: *The Analysis of Boolean Functions*. Cambridge Univ. Press, 2014, to appear. Preliminary version at analysisofbooleanfunctions.org. 890

[21] RYAN O'DONNELL AND ROCCO A. SERVEDIO: Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007. Preliminary version in CCC'06. [doi:10.1137/060669309] 890

[22] MICHEL TALAGRAND: On Russo's approximate zero-one law. *Ann. Probab.*, 22(3):1576–1587, 1994. JSTOR. 890

[23] RONALD DE WOLF: *A Brief Introduction to Fourier Analysis on the Boolean Cube*. Number 1 in Graduate Surveys. Theory of Computing Library, 2008. [doi:10.4086/toc.gs.2008.001] 890

AUTHORS

Eric Blais
Simons Postdoctoral Fellow
CSAIL, MIT. Cambridge, MA
eblais@csail.mit.edu
`http://people.csail.mit.edu/eblais`

Li-Yang Tan
Ph. D. Candidate
Department of Computer Science, Columbia University. New York, NY
liyang@cs.columbia.edu
`http://www.cs.columbia.edu/~liyang`

ABOUT THE AUTHORS

ERIC BLAIS is a Simons Postdoctoral Fellow in the Theory of Computing group at MIT. He graduated from Carnegie Mellon University in 2012; his advisor was Ryan O'Donnell. His research interests include the complexity of Boolean functions and sublinear algorithms.

LI-YANG TAN is a Ph. D. student at Columbia University, where he is fortunate to be advised by Rocco Servedio. His research interests lie in complexity theory, with a focus on discrete Fourier analysis, the complexity of Boolean functions, and computational learning theory.