Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 39 (2013)

On the sum of $L1$ influences

Artūrs Bačkurs*   Mohammad Bavarian†

Abstract

For a real-valued function p over the Boolean hypercube the $L1$-influence of p is defined to be $\sum_{i=1}^n \mathrm{E}_{x\in\{-1,1\}^n}\left[\frac{|p(x)-p(x^i)|}{2}\right]$, where the string x^i is defined by flipping the i-th bit of x. For Boolean functions the notion of $L1$ influence will coincide with the usual notion of influences defined as $\sum_{i=1}^n \mathrm{E}_{x\in\{-1,1\}^n}\left[\frac{|p(x)-p(x^i)|^2}{4}\right]$. For general $[-1,1]$-valued functions, however, the $L1$-influence can be much larger than its $L2$ counterpart.

In this work, we show that the $L1$-influence of a bounded $[-1,1]$-valued function p can be controlled in terms of the degree of p's Fourier expansion, resolving affirmatively a question of Aaronson and Ambainis (Proc. Innovations in Comp. Sc., 2011). We give an application of this theorem to the maximal deviation of cut-value of graphs. We also discuss the relationship between the study of bounded functions over the hypercube and the quantum query complexity of partial functions which was the original context in which Aaronson and Ambainis encountered this question.

1 Introduction

The notion of the influence of a variable [13, 4] plays a fundamental role in the study of functions over product probability spaces. One canonical example of such space, which we shall mostly consider in this work, is the discrete cube $\{-1,1\}^n$ equipped with the uniform probability measure. Given a function $f:\{-1,1\}^n\to\{0,1\}$, the influence of the direction i is defined as

$$\mathrm{Inf}_i(f) := \Pr_{x\in\{-1,1\}^n}[f(x)\neq f(x^i)],$$

where x^i denotes the neighbor of x in the i-th direction, i.e., $x^i=(x_1,x_2,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)$. The notion of i-th influence of f has a clear geometric interpretation as the fraction of all the edges of the hypercube in the i-th direction that lie on the edge boundary of $\mathrm{supp}(f)$. Hence, the sum $\mathrm{Inf}(f)=\sum_{i=1}^n \mathrm{Inf}_i(f)$ gives us the "total influence" of a function, measuring the total edge boundary of the support.

Going beyond the Boolean valued functions, the notion of the influence of a variable can be generalized in several ways. The idea is to replace the term $\Pr[f(x)\neq f(x^i)]$ with an analytical expression such as $\mathrm{E}\left[\left(\frac{|f(x)-f(x^i)|}{2}\right)^\alpha\right]$ for some non-zero $\alpha\in\mathbb{R}$. Notice that since for Boolean functions the

*MIT. E-mail: backurs@mit.edu.
†MIT. E-mail: bavarian@mit.edu.

ISSN 1433-8092

term $\frac{|f(x)-f(x^i)|}{2}$ is either 0 or 1, all these different notions of influences coincide in this case. Although in this work we are concerned with the $L1$ influences which corresponds to the case of $\alpha = 1$, it is better to start off with the more familiar $L2$ influences and their properties.

The total $L2$ influence of a function $f : \{-1,1\}^n \to \mathbb{R}$ is defined as the sum over all ith $L2$ influences $\mathrm{Inf}_i^{\mathrm{sq}}(f) = \mathrm{E}[(\frac{f(x)-f(x^i)}{2})^2]$. One reason for working with $L2$ influence is that simplicity of $L2$ norm in Fourier analysis allows one to derive a nice characterization of $L2$ influence in terms of Fourier coefficients of the function. More precisely, we have

$$\mathrm{Inf}^{\mathrm{sq}}(f) = \sum_{S \subseteq [n]} |S|\, \hat{f}(S)^2.$$

This alternative view of $L2$ influence as the average weight of Fourier coefficients is crucial in proving the following simple but important fact.

Fact 1.1 *Let $f : \{-1,1\}^n \to \mathbb{R}$. Denoting by $\deg(f)$ the size of largest $|S|$ with $\hat{f}(S) \neq 0$, we have*

$$\mathrm{Inf}^{\mathrm{sq}}(f) \leq \deg(f) \|f\|_2^2 .$$

Aaronson and Ambainis [1] in their study of query complexity of partial functions raised the following question.

Question 1.2 *Does an analogue of fact 1.1 holds for bounded functions if one replaces $L2$ influences with $L1$ influences? More precisely, does it hold that*

$$\mathrm{Inf}(f) = \mathrm{E}_x\left[\sum_{i=1}^{n} \frac{|f(x)-f(x^i)|}{2}\right] = O(\deg(f)^{O(1)}),$$

for any function $f : \{-1,1\}^n \to [-1,1]$?

Certainly, the above inequality implies a related inequality for general functions as the restriction of f taking values in $[-1,1]$ can be dropped by normalizing the right hand side of the inequality by $\|f\|_\infty$ term. One reason to work with bounded functions is that in the applications to complexity theory and especially for randomized and quantum query complexity the functions that arise naturally corresponds to quantities related to acceptance probability profile of an algorithm. Since probabilities are bounded in $[0,1]$ these functions also have to be bounded in the corresponding range. This is especially important in the context of bounded-error query complexity of partial functions where usually not much information on other norms beside the L^∞ norm of the acceptance profile function is available.

There are two reasons to suspect that proving this variation of Fact 1.1 might be much harder. First of all, bounded functions can have much asymptotically larger total $L1$ influence than $L2$ influence. In other words, although proving a bound on $L1$ influences in terms of degree implies a corresponding upper bound for $L2$ influences of a bounded function, one cannot hope for a reverse implication. The second difficulty relies on the fact that $L1$ influences do not have an easy characterization in terms of Fourier coefficients. What allowed us to avoid this difficulty in the $L2$ case was the equivalent characterization of $\mathrm{Inf}^{\mathrm{sq}}(f)$ in terms of the Fourier coefficient, a luxury not available in any case beside the case of $L2$ norm. Hence, to prove such a result one needs to relate the $L1$ influences of a function, which is defined in terms of the values of the function, to its degree that is most easily understood in the value of its Fourier coefficients.

1.1 Results

Let $p : \{-1,1\}^n \to \mathbb{R}$. The $L1$ influence of the i-th variable of a function $p : \{-1,1\}^n \to \mathbb{R}$ is defined as

$$\mathrm{Inf}_i(p) := \mathrm{E}_x \left[\frac{|p(x) - p(x^i)|}{2} \right],$$

where $x^i = (x_1, x_2, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$. The total $L1$ influence is defined as the sum of all individual i-th influences.

$$\mathrm{Inf}(p) := \sum_{i=1}^{n} \mathrm{Inf}_i(p).$$

Every function $p : \{-1,1\}^n \to \mathbb{R}$ has a unique representation as a multilinear polynomial in terms of Walsh functions $\{\chi_S\}_{S \subseteq [n]}$ where $\chi_S(x) = \prod_{i \in S} x_i$. Let

$$p(x) = \sum_{S \subseteq [n]} \widehat{p}(S)\ \chi_S(x)$$

be such a representation of p. Then we define $\deg(p)$ to be the maximal $|S|$ with $\widehat{p}(S) \neq 0$.

In [1] Aaronson and Ambainis asked whether the total $L1$ influence of a $[-1,1]$-valued function p can be bounded by a polynomial in $\deg(p)$. In this work we resolve the question of Aaronson and Ambainis affirmatively.

Theorem 1.3 *Let $p : \{-1,1\}^n \to \mathbb{R}$ and $\deg(p) = d$. Then we have*

$$\mathrm{Inf}(p) = O(A_p d^3 \log d).$$

where we define $A_p := \max_x p(x) - \min_x p(x)$.

Notice that $A_p \leq 2\|p\|_\infty$. The advantage of the parameter A_p over $\|p\|_\infty$ is that A_p is invariant under any shift $p \mapsto p + \gamma$, which makes both sides of the inequality in Theorem 1.3 invariant under this operation.

In section 5, we show how we can use the inequality above to prove the following result of Erdös, Goldberg and Pach [10] on the cut-deviation of the graphs.

Theorem 1.4 *Given a graph $G = (V,E)$ with density $\rho_G = |E|/\binom{n}{2}$ there always exists a cut (S, S^c) such that*

$$|\mathrm{E}(S, S^c) - \rho_G |S||S^c|| = \Omega\left(\min(\rho_G, 1 - \rho_G)\, n^{\frac{3}{2}}\right).$$

We prove this result by applying our inequality to the following polynomial,

$$g_G(x) := \frac{|E|}{2} - \rho_G \frac{|V|(|V|-1)}{4} + \frac{\rho_G}{2} \sum_{i<j} x_i x_j - \left(\frac{1}{2} - \frac{\rho_G}{2}\right) \sum_{i \sim j} x_i x_j.$$

An interesting feature of the example above is that it exhibits a non-trivial $\Omega(\sqrt{n})$ separation between $L1$ and $L2$ influences. This, in turn, shows that the result of Theorem 1.4 cannot be proved using the simple $L2$ influence inequality 1.1. More precisely, applying the $L2$ inequality to polynomial g_G would only show the existence of cuts with deviation $\Omega(n)$ rather than $\Omega(n^{3/2})$. This confirms the intuition that in some settings Theorem 1.3 can be much stronger than its $L2$ counterpart.

Finally, in the last section we give some future directions and open problems related to the sum of $L1$ influences.

1.2 Related Work

The importance of concept of influences in the analysis of functions over product spaces was already recognized in the pioneering work of Kahn, Kalail and Linial [13] and Bourgain, Kalai, Katznelson and Linial [4]. Building upon these results, Friedgut [11] showed that a Boolean function with very low total influence is somewhat "simple" as it can be approximated with a function depending on few coordinates, showing that the total influence in some regime acts as a complexity measure of functions. Bourgain [3] further studied the interaction between the condition of Boolean-valuedness and influences proving very powerful results about the *spectrum* of such functions.[1] Later, Dinur, Friedgut, Kindler and O'Donnel [9] obtained a (exponentially weaker but optimal) generalization of Bourgain's result [3] for the $[-1, 1]$-valued functions.

Most of the results mentioned above, either implicitly or explicitly, investigated the effects of Boolean-valuedness on the spectrum of functions. As most computational and learning problems are specified by a truth table of the form $f : \{-1, 1\}^n \to \{0, 1\}$, one may assume understanding the spectral properties of Boolean functions should be sufficient for the applications to complexity theory. Indeed, for many applications such as the study of small-depth circuits, threshold circuits, decision trees and even, via an easy reduction, the bounded error query complexity of total functions, this is sufficient. The main point here is the distinction between *total functions* versus *partial functions*. A total function is a Boolean function $f : \{-1, 1\}^n \to \{0, 1\}$ defined on the whole hypercube whereas a partial function $f : A \to \{0, 1\}$ is only defined on a strict subset A of the hypercube. The distinction between partial functions and total functions is crucial in the applications to query and communication complexity. For example, although it has been known since the work of Simon [21] that for partial functions quantum algorithms can be exponentially more powerful than classical algorithms, for total functions quantum algorithms can only exhibit at most a polynomial speedup. (See [5, 7] for further discussion and [12, 14, 24] for similar issues in communication complexity.)

It turns out that the case of quantum and randomized query complexity of partial functions is way less-understood than the case of total functions. The work of Aaronson and Ambainis [1] is one of the first papers trying to investigate the relationship between the size and the structure of the domain A of a partial function f, and the quantum versus classical advantage achievable for computing f. The intuition is that unless the domain A is specially structured and rather small, quantum algorithms should not be able to outperform classical algorithms by too much. Unfortunately, our knowledge in this topic still remains in its infancy and the recent work [2] shows that even partial functions arising from the restriction of very simple functions, such as parity, can exhibit very interesting behavior.

[1] The spectrum of a function usually refers to the weight distribution of Fourier coefficients of a function. The spectral properties refer to, for example, the behavior of $S(m) = \sum_{|S|\geq m} |\hat{f}(S)|^2$ or $h(p) = \sum_{S\subseteq[n]} |S|^p|\hat{f}(S)|^2$ as a function of m and p .

One of the first complications that arise when trying to address the problems regarding the query complexity of partial functions is that, instead of Boolean functions, one has to deal with more general bounded functions. To see this, let us first recall how one usually associates a polynomial to any (say, quantum) algorithm solving a query problem.

Lemma 1.5 (See [5]) *Let Q be a quantum algorithm with a black box access to an input $X \in \{-1,1\}^n$ trying to solve a problem $f : \{-1,1\}^n \to \{0,1,*\}$.[2] If Q makes T queries to the black-box before accepting or rejecting the input, its acceptance probability of each $X \in \{-1,1\}^n$ can be seen as a real-valued multilinear polynomial $p(X)$ of degree at most $2T$.*

Hence, we see that if an algorithm manages to solve a query problem in few queries, this implies the existence of a polynomial $p(X)$ of low degree satisfying $|p(X) - f(X)| \leq 1/3$ for any X in the domain A of f. Hence, if the domain of f is a strict subset of Boolean hypercube one has no information on $p(X)$ for $X \in A^c$. Unlike the case of "essentially Boolean functions" with the range $[0,1/3] \cup [2/3,1]$, which can be studied by simple reductions to Boolean function, the spectral properties of bounded functions can be different to those of Boolean functions as demonstrated by the work of Dinur et al [9]. Thus it seems that one prerequisite for making progress on problems regarding the tradeoffs between the size and the structure of the domain of a partial function and the quantum and classical query complexity is to develop some analytical tools for studying the properties of bounded functions over the hypercube.

Returning back to the study of influence of a variable, we should mention a series of recent developments surrounding the alternative notions of influences in the Gaussian setting by Keller, Mossel and Sen [16, 17].(See also [15]) As pointed out by Cordero-Erausquin and Ledoux [8] the geometric influences of Keller, Mossel and Sen can be seen as the "$L1$ influences" in the Gaussian setting. It is interesting to further clarify the relation between the notions of influence in the Gaussian setting introduced recently in above works with the $L1$ influences in discrete cube as studied in this paper. Notice that it is well-known that the Gaussian space can be seen as a "special case" of the Boolean case since as the central limit theorem indicates, Gaussian random variables can be simulated with a number of independent Bernoulli random variables to very good precision.

Lastly, we should mention that [1] is not the first place where the notion of $L1$ influences has appeared. In a seminal paper, Talagrand [22] obtained a very powerful generalization of KKL theorem [13]. Talagrand's inequality explicitly uses the notion of $L1$ influences. However, to our knowledge, [1] is the first place where the question of the relationship between $L1$ influences and the traditional complexity measure of Boolean functions such as the Fourier degree is raised.

2 Preliminaries

In this work, we use concepts from the analysis over the hypercube $\{-1,1\}^n$. For more extensive introduction to the analysis of Boolean functions and its application to complexity theory we refer to the surveys of de Wolf and O'Donnel [18, 23]. We also refer to [7] for extensive introduction to the complexity measures of functions such as randomized, quantum and deterministic query complexity and their relation to more analytic concepts such as degree, approximate degree, etc.

[2] This is the alternative notation for partial query complexity problems with $A = \mathrm{dom}(f)$ consisting of points x where $f(x) \neq *$. We say that an algorithm accepts an input if it outputs 1 and it rejects an input if it outputs 0.

It is well-known that any function $f : \{-1,1\}^n \to \mathbb{R}$ can be represented as a polynomial with real coefficients over the monomials $\chi_S(x) = \prod_{i \in S} x_i$ which are called Fourier-Walsh characters. The notion of the influence of a variable is well-known in the context of the analysis of Boolean functions. For a Boolean valued function $g : \{-1,1\}^n \to \{-1,1\}$ the influence of i-th coordinate is defined to be $\inf_i(g) = \Pr_x[g(x) \neq g(x^i)]$, where $x^i \in \{-1,1\}^n$ is the point x with ith coordinate flipped.

For more more general non-Boolean functions $p : \{-1,1\}^n \to \mathbb{R}$ the notion of influence is usually extended in $L2$ form by following definition

$$\mathrm{Inf}_i^{\mathrm{sq}}(p) = \mathrm{E}_x \left[\left(\frac{p(x) - p(x^i)}{2} \right)^2 \right].$$

In this work, we work with a different generalization of notion of influences to non-Boolean functions,

$$\mathrm{Inf}_i(p) = \mathrm{E}_x \left[\frac{|p(x) - p(x^i)|}{2} \right].$$

Notice that our notation differs from many other works in computer science literature where the above typically represent the $L2$ influences which we denote by $\mathrm{Inf}_i^{\mathrm{sq}}(f)$ and $\mathrm{Inf}^{\,\mathrm{sq}}(f)$. But following Aaronson and Ambainis [1] and since $L1$ influences are central to this work we choose to work with above notation to simplify the writing.

Another central tool in our work is the noise operator which is defined as follows.

Definition 2.1 *Noise operator with rate $\rho \in \mathbb{R}$ applied to polynomial p is the following polynomial:*

$$T_\rho p(x) := \sum_{S \subseteq [n]} \widehat{p}(S)\, \rho^{|S|} \chi_S(x).$$

For $\rho \in [-1,1]$ there is an alternative characterization of $T_\rho p$ which will be useful later. Consider a bipartite distribution over $(x, y) \in \{-1,1\}^n \times \{-1,1\}^n$ defined as follows: we pick $x \in \{-1,1\}^n$ uniformly at random, and for each i independently we set $y_i = x_i$ with probability $(1+\rho)/2$ and $y_i = -x_i$ with the remaining probability. It is not too hard to see [18, 23] that the above distribution, denoted by $x \sim_\rho y$, is symmetric in x and y and that the operator T_ρ satisfies

$$T_\rho\, p(x) = \mathrm{E}_{y \sim_\rho x}[p(y)],$$

for $\rho \in [-1,1]$. This characterization has the a very useful consequence, for $\rho \in [-1,1]$ we have $\|T_\rho(p)\|_\infty \leq \|p\|_\infty$ and in fact $\|T_\rho(p)\|_q \leq \|p\|_q$ for all $q \geq 1$.

3 The case of homogeneous polynomials

Recall that $p(x) = \sum_{S \subseteq [n]} \widehat{p}(S) \chi_S(x)$ is a homogeneous polynomial if $\widehat{p}(S) = 0$ for all S such that $|S| \neq \deg(p)$. In this section we prove the following theorem.

Theorem 3.1 *Let $p : \{-1,1\}^n \to \mathbb{R}$ of $\deg(p) = d$ be a homogeneous polynomial. Then we have*

$$\mathrm{Inf}(p) = O(A_p d^2 \log d).$$

Let $p(x) = \sum_{R\subseteq[n]} \widehat{p}(R)\,\chi_R(x)$ be a homogeneous polynomial of degree d. Let $S \subseteq [n]$. Critical to our analysis is the following polynomial

$$q_S(x) = \sum_{R\subseteq[n]:|R\cap S|=1} \widehat{p}(R)\,\chi_R(x).$$

Lemma 3.2 *For all $S \subseteq [n]$,*

$$q_S(x) = O(A_p d \log d).$$

Proof

We define $v_\alpha \in \mathbb{R}^d$ such that for any $1 \le k \le d$: $(v_\alpha)_k = \Pr_P[|P \cap [k]| \equiv 1 (\bmod 2)]$, where we choose set P by putting each $i \in [n]$ in it independently with probability α.

$$(v_\alpha)_k = \sum_{i:i\equiv 1(\bmod 2)} \alpha^i (1-\alpha)^{k-i}\binom{k}{i}$$

$$= \frac{1}{2}\Big(((1-\alpha)+\alpha)^k - ((1-\alpha)-\alpha)^k)\Big) = (1-(1-2\alpha)^k)/2.$$

Let $S \subseteq [n]$ and $S' \subseteq S$ be chosen by including every $i \in S$ independently with probability α. Then

$$A_p \ge \mathrm{E}_{S'}[p(x) - p(x^{S'})] = 2 \sum_{R\subseteq[n]:|R\cap S|\ge 1} \widehat{p}(R)\;\cdot (v_\alpha)_{|R\cap S|}\chi_R(x).$$

We will choose $\alpha_1, \alpha_2, ..., \alpha_d$ and $x_1, x_2, ..., x_d$ such that $\sum_{i=1}^n v_{\alpha_i} x_i = \overrightarrow{e}$ and $\sum_{i=1}^d |x_i| = O(d \log d)$, where $\overrightarrow{e}$ is d-dimensional vector with the first entry 1 and with the remaining entries 0. This gives

$$q_S(x) = \sum_{R\subseteq[n]:|R\cap S|=1} \widehat{p}(R)\,\chi_R(x) = \sum_{i=1}^{d} x_i \sum_{R\subseteq[n]:|R\cap S|=1} \widehat{p}(R)\,\chi_R(x)(v_{\alpha_i})_{|R\cap S|}$$

$$\le \frac{A_p}{2}\sum_{i=1}^{d}|x_i| = O(Ad\log d).$$

For any $-1 \le \gamma \le 1$ we consider vectors v'_γ with $(v'_\gamma)_k = \gamma^k$ for $-1 \le \gamma \le 1$. Notice that since $v_{1/2}$ is the vector of all $1/2$ entries, $v'_\gamma = 2\left(v_{1/2} - v_{(1-\gamma)/2}\right)$. So instead of working with v_α's directly we instead choose to work with v'_γ.

If d is even we choose $\gamma_i = -1/2 + (i-1)/d$ for $i \le d/2$ and $\gamma_i = (i - d/2)/d$ for $i > d/2$. If d is odd we choose $\gamma_i = -1/2 + (i-1/2)/d$ for $i \le (d-1)/2$ and $\gamma_i = (i + 1/2 - d/2)/d$ for $i \ge (d+1)/2$. Consider matrix M with v'_γ as columns. We have to solve $Mx = \overrightarrow{e}$. Notice that M is similar to Vandermonde matrix. Using Cramer's rule we obtain

$$|x_k| = \left|\frac{\gamma_1 ... \gamma_{k-1}\gamma_{k+1} ... \gamma_d}{\gamma_k}\frac{1}{\prod_{j\ne k}(\gamma_j - \gamma_k)}\right|.$$

Now because of the choice of γs we get $|\gamma_1 ... \gamma_{k-1}\gamma_{k+1}...\gamma_d| \leq |\prod_{j\neq k}(\gamma_j - \gamma_k)|$. Thus $\sum_i |x_i| \leq \sum_i \frac{1}{|\gamma_i|} = O(d \log d)$. ∎

Now we will prove Theorem 3.1.

Proof of Theorem 3.1

Consider

$$B = \mathrm{E}_S \left[\mathrm{E}_{x^{S^c}} \left[\sum_{i \in S} \left| \sum_{R: R\cap S = \{i\}} \widehat{p}(R)\, \chi_{R\setminus\{i\}}(x) \right| \right] \right],$$

where in the first expectation we choose S by putting each $i \in [n]$ in it with probability $1/d$ and in the second expectation we choose values of variables in complement of S uniformly and independently at random. Now if for every $i \in S$ we choose

$$x_i = \mathrm{sgn}\left(\sum_{R: R\cap S=\{i\}} \widehat{p}(R)\, \chi_{R\setminus\{i\}}(x) \right),$$

we can use the previous upper bound and conclude that $B = O(A_p d \log d)$ as well.

Now we lower bound B:

$$\begin{aligned} B &= \frac{1}{d-1} \sum_{i=1}^{n} \mathrm{E}_{x,z} \left| \sum_{R: i\in R} \widehat{p}(R)\, \chi_R(x)\chi_R(z) \right| \\ &= \Omega\left(\frac{1}{d} \sum_{i=1}^{n} \mathrm{E}_x \left| \sum_{R: i \in R} \widehat{p}(R)\, \chi_R(x) \right| \right), \end{aligned}$$

where we choose each $z_i = 0$ with probability $1/d$ and $z_i = 1$ with the remaining probability. In the last equality we moved expectation over z inside the absolute value and then used $\mathrm{E}_z[\chi_R(z)] = (1 - 1/d)^d = \Omega(1)$. (We use the fact that there is no $\alpha_R \neq 0$ with $|R| < d$.)

Now it remains to notice that

$$\mathrm{Inf}(p) = \sum_{i=1}^{n} \mathrm{E}_x \left[|p(x) - p(x^i)|/2 \right] = \sum_{i=1}^{n} \mathrm{E}_x \left| \sum_{R: i\in R} \widehat{p}(R)\, \chi_R(x) \right|.$$

∎

4 The case of general polynomials

Now we will modify the proof of Theorem 3.1 to solve the case of non-homogeneous polynomial (Theorem 1.3).

To prove the theorem we also need another lemma:

Lemma 4.1 *Let q be a degree-d polynomial in ρ such that $|q(\rho)| \leq 1$ for $-1 \leq \rho \leq 1$. Then the following equality holds: $q\left(\frac{1}{1-\frac{1}{d^2}}\right) = O(1)$.*

Lemma 4.1 follows from properties of Chebyshev polynomials and lemma of Paturi [19].

Proposition 4.2 ([20]) *Define $\mathcal{P}_d$ as follows*

$$\mathcal{P}_d = \{p \in \mathbb{R}[x] \mid \deg(p) \leq d \,, \max_{x \in [-1,1]} |p(x)| \leq 1\}$$

Then we have

$$\forall p \in \mathcal{P}_d, x \notin [-1,1] \qquad |p(x)| \leq |T_d(x)|$$

Where T_d is the d-th Chebychev polynomial of the first kind.

$$T_d(\rho) = \frac{1}{2}\left(\left(\rho + \sqrt{\rho^2 - 1}\right)^d + \left(\rho - \sqrt{\rho^2 - 1}\right)^d\right).$$

Lemma 4.3 (Paturi) $T_d(1+\gamma) \leq e^{2d\sqrt{2\gamma+\gamma^2}}$ *for all* $\gamma \geq 0$.

Proof [6] $T_d(1+\gamma) \leq (1+\gamma+\sqrt{2\gamma+\gamma^2})^d \leq (1+2\sqrt{2\gamma+\gamma^2})^d \leq e^{2d\sqrt{2\gamma+\gamma^2}}$. ∎

Proof of Theorem 1.3

Now combining Lemma 3.2 and Lemma 4.1 with the fact that for all $-1 \leq \rho \leq 1$ we have $\max_x T_\rho p(x) \leq \max_x p(x)$ and $\min_x T_\rho p(x) \geq \min_x p(x)$ (Those inequalities follow from $T_\rho p(x) = \mathrm{E}_{y \sim_\rho x}[p(y)]$.) we get that

$$\sum_{R \subseteq [n]: |R \cap S| = 1} \widehat{p}(R)\ \rho'^{|R|}\, \chi_R(x) = O(A_p d \log d),$$

where $\rho' = 1/(1 - 1/d^2)$. (We fix x and consider $T_\rho q_S(x)$ as a polynomial in ρ and then apply Lemma 4.1.)

Consider

$$B = \mathrm{E}_S\left[\mathrm{E}_{x^{S^c}}\left[\sum_{i \in S}\left|\sum_{R: R \cap S = \{i\}} \widehat{p}(R)\, \rho'^{|R|} \chi_{R \setminus \{i\}}(x)\right|\right]\right],$$

where in the first expectation we choose S by putting each $i \in [n]$ in it with probability $1/d^2$ and in the second expectation we choose values of variables in complement of S uniformly and independently at random. Now if for every $i \in S$ we choose

$$x_i = \mathrm{sgn}\left(\sum_{R: R \cap S = \{i\}} \widehat{p}(R)\, \rho'^{|R|} \chi_{R \setminus \{i\}}(x)\right),$$

we can use the previous upper bound and conclude that $B = O(A_p d \log d)$ as well.

Now we lower bound B:

$$B = \frac{1}{d^2 - 1} \sum_{i=1}^{n} \mathrm{E}_{x,z}\left|\sum_{R: i \in R} \widehat{p}(R)\, \rho'^{|R|} \chi_R(x) \chi_R(z)\right|$$

$$= \Omega \left(\frac{1}{d^2} \sum_{i=1}^{n} \mathrm{E}_x \left| \sum_{R: i \in R} \widehat{p}(R)\, \chi_R(x) \right| \right),$$

where we choose each $z_i = 0$ with probability $1/d^2$ and 1 with the remaining probability. In the last equality we moved expectation over z inside the absolute value and then used $\mathrm{E}_z[\rho'^{|R|}\chi_R(z)] = \rho'^{|R|}(1-1/d^2)^{|R|} = 1$. ■

5 A Corollary on Maximal Deviation of Cut-value of Graphs

In this section we use a very special case of our original theorem to reprove a theorem of Erdös et al. [10] on the maximum discrepancy of cut-values in graphs. In a graph $G = (V, E)$, by S^c we denote $V \setminus S$ and we write $u \sim v$ if and only if $(u, v) \in E$.

Definition 5.1 *For any graph $G = (V, E)$ and $0 \le p \le 1$ the cut-deviation $D_p(G)$ is the maximum over all cuts $(S, V \backslash S)$ of the discrepancy between the cut-value $|E(S, S^c)|$ and the expected cut-value $p|S|(|V| - |S|)$ (where we choose each edge independently with probability p), i.e.,*

$$D_p(G) = \max_{S \subseteq V} \left| |E(S, S^c)| - p|S||S^c| \right|.$$

We are interested in lower bounding the quantity $D_p(G)$. Given a $G = (V, E)$, let $\rho_G := |E|/\binom{|V|}{2}$ be the edge density. Notice that for any $p \neq \rho_G$ a random cut will already give a deviation of $\Omega(n^2)$ for $D_p(G)$. So the interesting case is when $p = \rho_G$. For this choice of the parameter we prove the following Theorem,

Theorem 5.2 *For every graph $G = (V, E)$,*

$$D_{\rho_G}(G) = \Omega(\min(\rho_G, 1 - \rho_G) n^{\frac{3}{2}}).$$

We note that the above inequality is tight as it follows by applying standard tail inequalities to Erodös-Renyi graphs $G(n, p)$. Moreover, the one-sides variants of this inequality

$$\max_{S \subseteq V} E(S, S^c) - \rho_G|S||S^c| = \Omega(\min(\rho_G, 1 - \rho_G) n^{\frac{3}{2}})$$

which holds for random graphs, does not hold in general as can be seen from the example of the complement of complete bipartite graph $K_{n/2,n/2}$.

To prove this result we will use the following lemma.

Lemma 5.3 *Let $G = (V, E)$. For any $S \subseteq V$ let $x_S \in \{-1, 1\}^{|V|}$ be such that $(x_S)_i = 1$ iff $i \in S$ (assume that $V = [n]$). Then*

$$g_p(x) := \frac{|E|}{2} - p\frac{|V|(|V| - 1)}{4} + \frac{p}{2} \sum_{i<j} x_i x_j - (1/2 - p/2) \sum_{i \sim j} x_i x_j$$

satisfies $g_p(x_S) = E(S, S^c) - p|S||S^c|$.

Proof
We check that $|E(S,S^c)| = 1/2|E| - 1/2\sum_{i\sim j} x_i x_j$ and $|S||S^c| = \frac{|V|(|V|-1)}{4} - \frac{1}{2}\sum_{i<j} x_i x_j$. ∎

Proof of Theorem 5.2 Set $p := \rho_G$. First we notice that,

$$A_{g_p} = \max_x g_p(x) - \min_x g_p(x) \le 2\max_x |g_p(x)| = 2\max_{S\subseteq[n]} |E(S,S^c) - p|S||S^c|| = 2D_p,$$

where in the third equality we use the previous lemma.
Theorem 1.3 implies

$$\mathrm{Inf}(g_p) = \sum_{i=1}^n \mathrm{E}_x\left[\left|p/2\sum_{j\nsim i} x_j - (1-p)/2\sum_{j\sim i} x_j\right|\right] = O\Big(\max_{S\subseteq[n]} \big|E(S,S^c) - p|S||S^c|\big|\Big),$$

where we use the fact that $\deg(g_p) = 2$.
Now we just need to lower bound the left hand side. For a particular i we have,

$$\mathrm{E}_x\left[\left|p/2\sum_{j\nsim i} x_j - (1-p)/2\sum_{j\sim i} x_j\right|\right] = \Omega(\min(p,1-p)\sqrt{n}),$$

where the last equality follows when we consider random walk of length $n-1$ with steps of length $p/2$ or $(1-p)/2$. ∎

6 Conclusion and Open Problems

The main open problem is to improve the bound in Theorem 1.3. We believe that this bound is far from optimal. It is conceivable that the total $L1$ influence of a $[-1,1]$-valued function p is always bounded by a linear function of the degree of p. We expect that optimizing our techniques one could improve the bound in Theorem 1.3 by logarithmic factors. However, we suspect that improving the upper bound to $O(\deg(p))$ would require some new ideas.

As mentioned in the introduction, we hope that our results and techniques in this work would be useful in the study of quantum versus classical query complexity of partial functions. However, as demonstrated in Section 5, the applications of our inequality may not be limited to complexity theory. There, we gave a proof of a purely combinatorial result of Erdös et al. by applying Theorem 1.3 to an appropriately chosen polynomial.

Another possible future direction is to clarify the relationship between the notion of $L1$ influence in the discrete cube as studied in this work and the alternative notions of influence in the Gaussian setting as discussed by [8, 16, 17].

7 Acknowledgments

We are grateful to Scott Aaronson for introducing this problem to us and his encouragement. We would also like to thank Jelena Markovic and Madhu Sudan for helping us improve the exposition of the results. MB is grateful to National Sciense Foundation for financial support provided through STC-award 0939370.

References

[1] S. Aaronson and A. Ambainis. The need for structure in quantum speedups. *In Innovations in Computer Science*, pages 338-352. 2011.

[2] S. Aaronson, A. Ambainis and M. Bavarian. Quantum query complexity of restriction of parity and the sensitivity conjecture. In Preperation 2013.

[3] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. Israel Journal of Mathematics 131.1 (2002): 269-276.

[4] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson and N. Linial. The influence of variables in product spaces. Israel Journal of Mathematics 77.1-2 (1992): 55-64.

[5] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum Lower Bounds by Polynomials. In *39th IEEE Symposium on Foundations of Computer Science (FOCS 98)*, pp.352-361.

[6] H. Buhrman, R. Cleve, R. de Wolf, and Ch. Zalka. Bounds for Small-Error and Zero-Error Quantum Algorithms. In *40th IEEE Symposium on Foundations of Computer Science (FOCS 99)*, pp.358-368.

[7] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. Theoretical Computer Science 288.1 (2002): 21-43.

[8] D. Cordero-Erausquin and M. Ledoux. Hypercontractive measures, Talagrand's inequality, and influences. Geometric Aspects of Functional Analysis. Springer Berlin Heidelberg, 2012. 169-189.

[9] I. Dinur, E. Friedgut, G. Kindler and R. O'Donnell. On the fourier tails of bounded functions over the discrete cube. In Proceedings of the thirty-eighth annual ACM symposium on Theory of computing (STOC '06).

[10] P. Erdös. M. Goldberg, J. Pach and J. Spencer. Cutting a graph into two dissimilar halves. , J. Graph Theory 12, No.1, 121-131 (1988).

[11] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. Combinatorica 18.1 (1998): 27-35.

[12] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, pp. 516-525. ACM, 2007.

[13] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In Proceedings of the 29th Annual Symposium on Foundations of Computer Science (SFCS '88). IEEE Computer Society, Washington, DC, USA, 68-80.

[14] B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In Proceedings of the 43rd annual ACM symposium on Theory of computing, pp. 31-40. ACM, 2011.

[15] N. Keller. On the influences of variables on Boolean functions in product spaces. Combinatorics, Probability and Computing 20.01 (2011): 83-102.

[16] N. Keller, E. Mossel and A. Sen. Geometric influences. The Annals of Probability 40.3 (2012): 1135-1166.

[17] N. Keller, E. Mossel and A. Sen. Geometric Influences II: Correlation Inequalities and Noise Sensitivity. arXiv preprint arXiv:1206.1210 (2012).

[18] R. O'Donnell. 2008. Some topics in analysis of Boolean functions. In *Proceedings of the 40th annual ACM symposium on Theory of computing (STOC '08)*. ACM, New York, NY, USA, 569-578.

[19] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In *Proceedings of 24th STOC*, pages 468-474, 1992.

[20] T. J. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory.* Wiley-Interscience, second edition, 1990.

[21] D.R. Simon. On the power of quantum computation. SIAM Journal on Computing 26.5 (1997): 1474-1483.

[22] M. Talagrand. On Russo's approximate zero-one law. The Annals of Probability (1994): 1576-1587.

[23] R. de Wolf. A Brief Introduction to Fourier Analysis on the Boolean Cube. Theory of Computing (ToC Library, Graduate Surveys 1), 6, 2008.

[24] R. de Wolf. Quantum communication and complexity. Theoretical Computer Science 287.1 (2002): 337-353.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de