THEORY OF COMPUTING, Volume 7 (2011), pp. 147–153
www.theoryofcomputing.org

NOTE

The Influence Lower Bound Via Query Elimination

Rahul Jain* Shengyu Zhang†

Received: March 4, 2011; published: July 19, 2011.

Abstract: We give a simple proof, via query elimination, of a result due to O'Donnell, Saks, Schramm, and Servedio, which shows a lower bound on the zero-error expected query complexity of a function f in terms of the maximum influence of any variable of f. Our lower bound also applies to the two-sided error expected query complexity of f, and it allows an immediate extension which can be used to prove stronger lower bounds for some functions.

ACM Classification: F.1.3

AMS Classification: 68Q17

Key words and phrases: randomized query complexity, influence of variables

1 Introduction

Query complexity measures the hardness of computing a function f by the minimum number of input variables one needs to read before knowing the function's value. For $\varepsilon \geq 0$ and a distribution λ on the inputs, a k-expected query randomized algorithm has λ-distributional error ε, if its expected queries (over the inputs drawn from λ and its random coins) is at most k, and its expected error (over the inputs drawn from λ and its random coins) is at most ε. The ε-error λ-distributional expected query complexity of f, denoted $\mathsf{D}_{\varepsilon}^{\lambda}(f)$, is the minimum number k such that there exists a k-expected query randomized algorithm which has λ-distributional error ε.

*Supported by the internal grants of The Centre of Quantum Technologies.

†Partially supported by Research Grants Council of Hong Kong S.A.R. (Project numbers: CUHK419309 and CUHK418710.)

DOI: 10.4086/toc.2011.v007a010

The influence of a variable is another important quantity which measures the importance of the variable to the function's value (on average over other variables). Both query complexity and influence are well-studied subjects; see [1] for a survey of the former (with many other complexity measures) and [11, 4] for surveys of the latter (and Fourier analysis on Boolean functions).

Definition 1.1 (Influence)**.** Let $f : \mathcal{X}^n \to \mathcal{Z}$ be a function, and X_i's and Y_i's (for $i = 1, \ldots, n$) be random variables independently and identically distributed according to μ on $\mathcal{X}$. Let $X = X_1 \cdots X_n$, and for each $i \in [n]$, let X^i represent the random variable $X_1 \ldots X_{i-1} Y_i X_{i+1} \ldots X_n$. The *influence* of variable i on f with respect to μ is defined as

$$\inf_i(f, \mu) = \Pr[f(X) \neq f(X^i)] .$$

The *maximum influence* of f with respect to μ is defined as

$$\inf_{\max}(f, \mu) = \max_i \inf_i(f, \mu) .$$

For Boolean functions, the influence as defined above is half of that in [12].

For any randomized algorithm $\mathcal{P}$ and any fixed input distribution (which is clear from the context), let $\varepsilon(\mathcal{P})$ denote the error probability (over random coins and inputs) of $\mathcal{P}$, and $\delta_i(\mathcal{P})$ denote the probability (over random coins and inputs) that $\mathcal{P}$ queries input variable i. In [12], O'Donnell, Saks, Schramm and Servedio proved the following:

Theorem 1.2 ([12])**.** *Let $f : \{-1, +1\}^n \to \{-1, +1\}$. Let μ be a distribution on $\{-1, +1\}$, and X be drawn from $\mu^{\otimes n}$. Then for any zero-error randomized query algorithm $\mathcal{P}$ (querying X),*

$$\sum_{i=1}^{n} \delta_i(\mathcal{P}) \inf_i(f, \mu) \geq \frac{\mathsf{Var}[f]}{2} ,$$

where $\mathsf{Var}[f] = \mathsf{E}[f^2] - \mathsf{E}[f]^2$ is the variance of $f(X)$. In particular, let $\mathcal{P}$ be a randomized algorithm with no error and $\mathsf{D}_0^{\mu^{\otimes n}}(f)$ expected queries. Then,

$$\mathsf{D}_0^{\mu^{\otimes n}}(f) = \sum_{i=1}^{n} \delta_i(\mathcal{P}) \geq \frac{\mathsf{Var}[f]}{2 \cdot \inf_{\max}(f, \mu)} .$$

Recently Lee [9] gave another proof of this fact.

Together with another bound for monotone functions [13],

$$\mathsf{D}_0^{\mu_p^{\otimes n}}(f) \geq \Big(\sum_i \inf_i(f, \mu_p) \Big)^2 \log_2 \frac{1}{p(1-p)} ,$$

where μ_p is the distribution on $\{-1, +1\}$ with -1 picked with probability p, Theorem 1.2 gives a lower bound of $\Omega(n^{2/3})$ for the randomized query complexity of all nontrivial monotone Boolean functions invariant under a transitive group of permutations (on the variables). For such functions we let p be the "critical threshold," namely the probability such that the function takes value 1 with probability exactly half. Observe that all the variables have the same influence; thus, depending on whether the influence is large or small, one of the two bounds can be applied to yield the $\Omega(n^{2/3})$ lower bound. This in particular reproduces the $\Omega(|V|^{4/3})$ lower bound for all monotone graph properties in [5], which is $O(\log^{1/3}(|V|))$ shy of the current record [2].

In this paper we give a new proof of Theorem 1.2, arguably shorter and simpler than both previous ones [12, 9]. In fact, we prove a stronger statement that applies to the two-sided error case.

Theorem 1.3. *Let $f : \mathcal{X}^n \to \mathcal{Z}$ be a function, μ be a distribution on $\mathcal{X}$, $\varepsilon \geq 0$, and X be drawn from $\mu^{\otimes n}$. Then for any randomized query algorithm $\mathcal{P}$ (querying X),*

$$\sum_{i=1}^{n} \delta_i(\mathcal{P})\inf_i(f,\mu) \geq 1 - f_{\max} - \varepsilon(\mathcal{P}), \tag{1}$$

where $f_{\max} = \max_{z\in\mathcal{Z}} \Pr[f(X) = z]$. In particular, let $\mathcal{P}$ be an algorithm with expected error ε and expected queries $\mathsf{D}_{\varepsilon}^{\mu^{\otimes n}}(f)$. Then,

$$\mathsf{D}_{\varepsilon}^{\mu^{\otimes n}}(f) = \sum_{i=1}^{n} \delta_i(\mathcal{P}) \geq \frac{1 - f_{\max} - \varepsilon}{\inf_{\max}(f,\mu)}. \tag{2}$$

It is easily seen that $1 - f_{\max} \geq \mathsf{Var}[f]/4$ for Boolean functions f, therefore the above theorem implies Theorem 1.2 (up to a factor of 2) as a special case.

We present the proof of Theorem 1.3 in the next section. The original proof in [12] used a hybrid argument on decision trees, and Lee in [9] gave a martingale-based proof. Our approach proceeds by query elimination: we can save one query without increasing the error by more than $\max_i \inf_i(f,p)$ and eventually eliminate all queries to obtain a zero-query algorithm, which must have a large error probability on a hard distribution. This bounds from below the number of queries of the original algorithm. The analysis for the increase in error due to eliminating one query is quite simple and follows from the union bound (applied just once) and the observation that X^i is identically distributed to X.

We can improve the lower bound by considering a function g, which is close to f but could potentially have smaller $\inf_{\max}$, a method often referred to as *smoothing*. As is the case with the rectangle bound and the discrepancy bound in communication complexity and query complexity, where the smoothed versions can prove strong lower bounds [7, 14, 15, 10, 8, 6, 3], this smoothed influence lower bound also gives stronger bounds than Equation (2) for some functions.

Let $g : \mathcal{X}^n \to \mathcal{Z}$ be a function such that $\Pr[f(X) \neq g(X)] \leq \delta$, where X is drawn from $\mu^{\otimes n}$ as above and $\delta \geq 0$. It is easily noted that an algorithm for f with average error under $\mu^{\otimes n}$ at most ε also works as an algorithm for g with average error under $\mu^{\otimes n}$ at most $\varepsilon + \delta$. Therefore $\mathsf{D}_{\varepsilon}^{\mu^{\otimes n}}(f) \geq \mathsf{D}_{\varepsilon+\delta}^{\mu^{\otimes n}}(g)$ and we have the following corollary.

Corollary 1.4. *Let $f : \mathcal{X}^n \to \mathcal{Z}$ be a function, μ be a distribution on $\mathcal{X}$ and $\varepsilon, \delta \geq 0$. Let X be drawn from $\mu^{\otimes n}$. Let $g : \mathcal{X}^n \to \mathcal{Z}$ be a function such that $\Pr[f(X) \neq g(X)] \leq \delta$. Then*

$$\mathsf{D}_{\varepsilon}^{\mu^{\otimes n}}(f) \geq \mathsf{D}_{\varepsilon+\delta}^{\mu^{\otimes n}}(g) \geq \frac{1 - \max_{z\in\mathcal{Z}} \Pr[g(X) = z] - \varepsilon - \delta}{\inf_{\max}(g,\mu)}.$$

Note that there are functions f with large $\inf_{\max}$ which are close to some other function g with small $\inf_{\max}$. For example, Tribes is the OR of $s \approx n/\log_2 n$ AND gates, each of $t \approx \log_2 n - \log_2 \log_2 n$ variables. The parameters s,t are set to make approximately half the inputs take the value 1. It is well known that for this function, all influences $\inf_i = \Theta(\log n/n)$, where the distribution is uniform on all inputs. Let g be Tribes, and obtain f from g by picking a subset $S = \{0,1\} \times S'$ of inputs, where $S' \subseteq \{0,1\}^{n-1}$ has density δ, and defining

$$f(x) = \begin{cases} x_1 & \text{if } x \in S, \\ g(x) & \text{otherwise,} \end{cases}$$

where x_1 denotes the first bit of x. Then the first variable has influence at least $\Omega(\delta)$, so applying the bound of Equation (2) yields only a constant lower bound for the distributional query complexity of f. As g is δ-close to f and $\inf_{\max}(g) = \Theta(\log n/n)$, the above corollary yields the stronger lower bound of $\Theta(n/\log n)$.

Remark 1.5.

1. Our proof does not need to assume that the distributions of the different variables are the same. The proof goes through and the bound applies analogously as long as these distributions are independent.

2. The paper [12] also extends Theorem 1.2 to the general metric case where the influence is defined as

$$\inf_i^d(f,\mu) = \mathsf{E}[d(f(X), f(X^i))] .$$

   Our proof also extends to this general case. To see this, it suffices to note that all we use is the triangle inequality, which holds for any metric space.

2 Proof of main result

Proof of Theorem 1.3. Equation (2) is an easy corollary of Equation (1). For Equation (1), suppose that $\mathcal{P}$ makes at most k queries for any input and any coins. We will prove the statement by induction on k. The base case of $k = 0$ is trivial since an algorithm that does not make any query succeeds with probability at most $f_{\max}$. For the general $k > 0$, we will first show the statement when $\mathcal{P}$ is deterministic (assuming the induction hypothesis on randomized query algorithms). Suppose X_i is the first query of $\mathcal{P}$; without loss of generality we can assume that $\mathcal{P}$ does not query X_i afterwards. We will show a randomized algorithm $\mathcal{P}'$ making at most $k-1$ queries with $\varepsilon(\mathcal{P}') \leq \varepsilon(\mathcal{P}) + \inf_i(f,\mu)$. In $\mathcal{P}'$ we do not make this query, but assume the answer to this query to be Y_i, where Y_i is distributed according to μ and is independent of X. From here on $\mathcal{P}'$ proceeds identically to $\mathcal{P}$.

By construction the maximum number of queries made by $\mathcal{P}'$ (over coins and inputs) is at most $k-1$. Let $\mathrm{ans}(\mathcal{P}, X)$ represent the answer of algorithm $\mathcal{P}$ on input X. Recall the definition of X^i from

Definition 1.1. Since $\mathrm{ans}(\mathcal{P},X^i) \neq f(X)$ implies either $\mathrm{ans}(\mathcal{P},X^i) \neq f(X^i)$ or $f(X^i) \neq f(X)$, we have,

$$\begin{aligned}
\varepsilon(\mathcal{P}') &= \Pr[\mathrm{ans}(\mathcal{P},X^i) \neq f(X)] \\
&\leq \Pr[\mathrm{ans}(\mathcal{P},X^i) \neq f(X^i)] + \Pr[f(X^i) \neq f(X)] && \text{(from union bound)} \\
&= \Pr[\mathrm{ans}(\mathcal{P},X) \neq f(X)] + \Pr[f(X^i) \neq f(X)] && \text{(since } X \text{ is identically distributed to } X^i\text{)} \\
&= \varepsilon(\mathcal{P}) + \mathrm{inf}_i(f,\mu)\,. && (3)
\end{aligned}$$

Therefore (below we abbreviate $\mathrm{inf}_i(f,\mu) = \mathrm{inf}_i$),

$$\begin{aligned}
\sum_{i'} \delta_{i'}(\mathcal{P})\mathrm{inf}_{i'} &= 1 \cdot \mathrm{inf}_i + \sum_{i' \neq i} \delta_{i'}(\mathcal{P}) \\
&= \mathrm{inf}_i + \sum_{i' \neq i} \delta_{i'}(\mathcal{P}') && \text{(since } X \text{ is identically distributed to } X^i\text{)} \\
&= \mathrm{inf}_i + \sum_{i'=1}^{n} \delta_{i'}(\mathcal{P}') && \text{(since } \mathcal{P}' \text{ does not query } X_i\text{)} \\
&\geq \mathrm{inf}_i + 1 - f_{\max} - \varepsilon(\mathcal{P}') && \text{(by the induction hypothesis)} \\
&\geq \mathrm{inf}_i + 1 - f_{\max} - \varepsilon(\mathcal{P}) - \mathrm{inf}_i && \text{(from Equation (3))} \\
&= 1 - f_{\max} - \varepsilon(\mathcal{P})\,.
\end{aligned}$$

Now let us show Equation (1) when $\mathcal{P}$ is a randomized algorithm. We can think of $\mathcal{P}$ as invoking deterministic algorithm $\mathcal{P}^j$ with probability p_j, where each $\mathcal{P}^j$ makes at most k queries on any input. Then (assuming Equation (1) for all deterministic algorithms making at most k queries)

$$\begin{aligned}
\sum_i \delta_i(\mathcal{P})\mathrm{inf}_i = \sum_i \sum_j p_j \delta_i(\mathcal{P}^j)\mathrm{inf}_i &= \sum_j p_j \sum_i \delta_i(\mathcal{P}^j)\mathrm{inf}_i \\
&\geq 1 - f_{\max} - \sum_j p_j \cdot \varepsilon(\mathcal{P}^j) = 1 - f_{\max} - \varepsilon(\mathcal{P})\,.
\end{aligned}$$

□

3 Acknowledgment

We thank Ronald de Wolf for detailed and helpful comments on an earlier draft of the paper. The original paper only proves Equation (2). Ryan O'Donnell and an anonymous referee pointed out that the same argument can be easily extended to show Equation (1), yielding the current presentation.

References

[1] HARRY BUHRMAN AND RONALD DE WOLF: Complexity measures and decision tree complexity: A survey. *Theoret. Comput. Sci.*, 288(1):21–43, 2002. [doi:10.1016/S0304-3975(01)00144-X] 148

[2] AMIT CHAKRABARTI AND SUBHASH KHOT: Improved lower bounds on the randomized complexity of graph properties. In *Proc. 28th Intern. Colloq. Autom. Lang. Program. (ICALP)*, number 2076 in Lecture Notes in Comput. Sci., pp. 285–296. Springer, 2001. [doi:10.1007/3-540-48224-5_24] 149

[3] AMIT CHAKRABARTI AND ODED REGEV: An optimal lower bound on the communication complexity of Gap-Hamming-Distance. In *Proc. 43th STOC*, pp. 51–60. ACM Press, 2011. [doi:10.1145/1993636.1993644] 149

[4] 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] 148

[5] PÉTER HAJNAL: An $\Omega(n^{4/3})$ lower bound on the randomized complexity of graph properties. *Combinatorica*, 11(2):131–143, 1991. [doi:10.1007/BF01206357] 149

[6] RAHUL JAIN AND HARTMUT KLAUCK: The partition bound for classical communication complexity and query complexity. In *Proc. 25th IEEE Conf. Comput. Complexity (CCC)*, pp. 247–258. IEEE Comp. Soc. Press, 2010. [doi:10.1109/CCC.2010.31] 149

[7] HARTMUT KLAUCK: Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 1:20–46, 2007. [doi:10.1137/S0097539702405620] 149

[8] HARTMUT KLAUCK: A strong direct product theorem for Disjointness. In *Proc. 42nd STOC*, pp. 77–86. ACM Press, 2010. [doi:10.1145/1806689.1806702] 149

[9] HOMIN LEE: Decision trees and influence: An inductive proof of the OSSS inequality. *Theory of Computing*, 6(1):81–84, 2010. [doi:10.4086/toc.2010.v006a004] 148, 149

[10] TROY LEE AND SHENGYU ZHANG: Composition theorems in communication complexity. In *Proc. 37th Intern. Colloq. Autom. Lang. Program. (ICALP)*, number 6198 in Lecture Notes in Comput. Sci., pp. 475–489. Springer, 2010. [doi:10.1007/978-3-642-14165-2_41] 149

[11] RYAN O'DONNELL: Some topics in analysis of Boolean functions. In *Proc. 40th STOC*, pp. 569–578. ACM Press, 2008. [doi:10.1145/1374376.1374458] 148

[12] RYAN O'DONNELL, MICHAEL E. SAKS, ODED SCHRAMM, AND ROCCO A. SERVEDIO: Every decision tree has an influential variable. In *Proc. 46th FOCS*, pp. 31–39. IEEE Comp. Soc. Press, 2005. [doi:10.1109/SFCS.2005.34] 148, 149, 150

[13] RYAN O'DONNELL AND ROCCO SERVEDIO: Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007. [doi:10.1137/060669309] 148

[14] ALEXANDER SHERSTOV: The pattern matrix method for lower bounds on quantum communication. In *Proc. 40th STOC*, pp. 85–94. ACM Press, 2008. [doi:10.1145/1374376.1374392] 149

[15] YAOYUN SHI AND YUFAN ZHU: The quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5&6):444–460, 2009. `http://www.rintonpress.com/journals/qic/`, `http://arxiv.org/pdf/0710.0095v4/`. 149

AUTHORS

Rahul Jain
Assistant professor
National University of Singapore
rahul@comp.nus.edu.sg
http://comp.nus.edu.sg/~rahul

Shengyu Zhang
Assistant professor
The Chinese University of Hong Kong
syzhang@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/~syzhang

ABOUT THE AUTHORS

RAHUL JAIN obtained his Ph. D. in computer science from the Tata Institute of Fundamental Research, Mumbai, India in 2003. He was a postdoctoral fellow for two years at the University of California, Berkeley (2004-2006) and for two years at the Institute for Quantum Computing (IQC), University of Waterloo, Canada (2006-2008). In 2008, he joined NUS as an Assistant Professor in the Computer Science Department with cross appointment with CQT. His research interests are in the areas of information theory, quantum computation, cryptography, communication complexity, and computational complexity theory.

SHENGYU ZHANG received his B. S. in Mathematics at Fudan University in 1999, his M. S. in Computer Science at Tsinghua University in 2002, and his Ph. D. in Computer Science at Princeton University in 2006. After working in NEC Laboratories America for a summer, and in California Institute of Technology for two years as a postdoctoral researcher, he joined The Chinese University of Hong Kong as an assistant professor. His main research interests are complexity theories in various randomized and quantum models.