The influence lower bound via query elimination[*]

Rahul Jain[†]   Shengyu Zhang[‡]

Abstract

We give a simpler proof, via query elimination, of a result due to O'Donnell, Saks, Schramm and Servedio, which shows a lower bound on the zero-error randomized 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 distributional query complexity of f, and it allows an immediate extension which can be used to prove stronger lower bounds for some functions.

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 value. A k-query ϵ-error randomized query algorithm is one that, on all inputs, has at most ϵ error probability and makes at most k queries over all random coins. The ϵ-error randomized query complexity of f, denoted $\mathsf{R}_\varepsilon(f)$, is the minimum number k such that there exists a k-query ϵ-error randomized query algorithm. The influence of a variable is another important quantity which measures the importance of the variable to the function value (on average over other variables). More precisely, for a function $f : \mathcal{X}^n \to \mathcal{Z}$ and a distribution μ on $\mathcal{X}$, the influence of the i-th variable is defined as $\inf_i(f, \mu) = \Pr[f(X) \neq f(X^i)]$, where $X = X_1 \ldots X_n$ is drawn from $\mu^{\otimes n}$ and X^i is obtained from X by re-randomizing X_i; namely $X^i = X_1 \ldots X_{i-1} Y_i X_{i+1} \ldots X_n$, where Y_i is drawn from μ and Y_i is independent of X. Both query complexity and influence are well-studied subjects; see [BdW02] for a survey of the former (with many other complexity measures) and [O'D08] for a survey of the latter (and Fourier analysis on Boolean functions).

Randomized query complexity can be lower bounded in terms of influence. In [OSSS05], O'Donnell, Saks, Schramm and Servedio proved that for all Boolean functions $f : \{-1, +1\}^n \to \{-1, +1\}$,

$$\mathsf{R}_0(f) = \Omega \left(\frac{\mathrm{Var}[f]}{\max_i \inf_i(f, \mu_p)} \right) \ . \tag{1}$$

Above μ_p is the distribution on $\{-1, +1\}$ with -1 picked with probability p; $\mathrm{Var}[f]$ is the variance of $f(X)$ with X drawn from $\mu_p^{\otimes n}$ and $\mathsf{R}_0(f)$ represents the zero-error randomized query complexity of f; namely the minimum over all randomized query algorithms with no error on each input, of

the maximum expected (over the random coins) number of queries made by the query algorithm on any input. Recently Lee [Lee10] gave another proof of this fact. Together with another bound $\mathsf{R}_0(f) \geq (\sum_i \inf_i(f, \mu_p))^2/(4p(1-p))$ for monotone functions [OS07], it gives a lower bound of $\Omega(n^{2/3})$ for all monotone functions invariant to a transitive group of permutations (on variables). This in particular reproduces the $\Omega(n^{4/3})$ lower bound for all monotone graph properties in [Haj91], which is $O(\log^{1/3}(n))$ shy of record [CK01].

In this paper we give a new proof of Eq. (1), arguably shorter and simpler than both previous ones [OSSS05, Lee10]. In fact we prove a stronger statement that applies to the two-sided error case. The basic idea is 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 lower bounds 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.

Since we lower bound distributional query complexity (defined in the next section), we get a smoothed version of the influence bound as an immediate consequence. As in the cases with the rectangle bound and the discrepancy bound in communication complexity and query complexity, where the smoothed versions can prove strong lower bounds [Kla07, She08, SZ09, LZ10, Kla10, JK10, CR11], this smoothed influence lower bound also gives stronger bounds for some functions than Eq. (1).

2   Main result and proof

Definition 1 (Influence) *Let $f : \mathcal{X}^n \to \mathcal{Z}$ be a function, and X_i's and Y_i's (for $i = 1, ..., n$) be random variables i.i.d. distributed according to μ on $\mathcal{X}$. 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 X_i on f 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 $\varepsilon > 0$, a deterministic k-query algorithm has the λ-distributional error ϵ if it makes at most k queries over all possible inputs, and for a random input drawn from λ, the average error probability is ϵ. The ϵ-error λ-distributional query complexity of f, denoted $\mathsf{D}_\varepsilon^\lambda(f)$, is the minimum number k such that there exists a k-query algorithm which has the μ-distributional error ϵ. We show the following.

Theorem 1 *Let $f : \mathcal{X}^n \to \mathcal{Z}$ be a function, μ be a distribution on $\mathcal{X}$ and $\varepsilon > 0$. Let X be drawn from $\mu^{\otimes n}$. Then,*
$$\mathsf{D}_\varepsilon^{\mu^{\otimes n}}(f) \geq \frac{1 - \max_{z \in \mathcal{Z}} \Pr[f(X) = z] - \varepsilon}{\inf_{\max}(f, \mu)}.$$

Proof: Let $\mathcal{P}_k$ be a deterministic k-query algorithm for f with $\mu^{\otimes n}$-distributional error at most δ. We present a deterministic $(k-1)$-query algorithm $\mathcal{P}_{k-1}$ for f with $\mu^{\otimes n}$-distributional error at most $\delta + \inf_{\max}(f, \mu)$. This way, starting from an algorithm which makes $\mathsf{D}_\varepsilon^{\mu^{\otimes n}}(f)$ queries and has average error at most ε, repeating the above procedure gives another algorithm $\mathcal{P}_0$ which makes

no queries and has average error at most $\varepsilon + \mathsf{D}_\varepsilon^{\mu^{\otimes n}}(f) \cdot \inf_{\max}(f, \mu)$. It is easily seen that $\mathcal{P}_0$ must have error at least $1 - \max_{z \in \mathcal{Z}} \Pr[f(X) = z]$ and hence we get the desired result.

Now we show how to obtain $\mathcal{P}_{k-1}$ from $\mathcal{P}_k$. We will show a randomized algorithm $\mathcal{P}'_{k-1}$ with at most $k-1$ queries on any input and any random coins and average error under $\mu^{\otimes n}$ at most $\delta + \inf_{\max}(f, \mu)$. From $\mathcal{P}'_{k-1}$, using an easy averaging argument (and fixing coins of $\mathcal{P}'_{k-1}$ appropriately), we can get a deterministic algorithm $\mathcal{P}_{k-1}$ with at most $k-1$ queries on any input and the same average error bound as in $\mathcal{P}'_{k-1}$.

Let X_i be the first query of $\mathcal{P}_k$ and without loss of generality we can assume that $\mathcal{P}_k$ does not query X_i any more afterward. In $\mathcal{P}'_{k-1}$ 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 $\mathcal{P}'_{k-1}$ proceeds identically to $\mathcal{P}_k$. By construction the maximum number of queries made by $\mathcal{P}'_{k-1}$ is at most $k-1$. Let $\mathrm{ans}(\mathcal{P}, X)$ represent the answer of algorithm $\mathcal{P}$ on input X. Since $\mathrm{ans}(\mathcal{P}_k, X^i) \neq f(X)$ implies either $\mathrm{ans}(\mathcal{P}_k, X^i) \neq f(X^i)$ or $f(X^i) \neq f(X)$, we have

$$
\begin{aligned}
\Pr[\mathcal{P}'_{k-1} \text{ makes error on input } X] =\ & \Pr[\mathrm{ans}(\mathcal{P}_k, X^i) \neq f(X)] \\
\leq\ & \Pr[\mathrm{ans}(\mathcal{P}_k, X^i) \neq f(X^i)] + \Pr[f(X^i) \neq f(X)] \quad \text{(from union bound)} \\
=\ & \Pr[\mathrm{ans}(\mathcal{P}_k, X) \neq f(X)] + \Pr[f(X^i) \neq f(X)] \quad \text{(since } X \text{ is identically distributed to } X^i) \\
\leq\ & \delta + \inf{}_{\max}(f, \mu)\ .
\end{aligned}
$$

∎

It is easily argued that $\mathsf{R}_0(f) = \Omega(\mathsf{R}_\varepsilon(f)) = \Omega(\mathsf{D}_\varepsilon^{\mu^{\otimes n}}(f))$ for ε, μ as above. Also $1 - \max_z \Pr[f(X) = z] = \Omega(\mathrm{Var}[f])$ for Boolean functions f, therefore the above theorem implies Eq. (1).

Next we improve the lower bound by going to a function g, which is close to f but could potentially have smaller $\inf_{\max}$. 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 $\mu^{\otimes n}$ being at most ε also works as an algorithm for g with average error under $\mu^{\otimes n}$ being at most $\varepsilon + \delta$. Therefore $\mathsf{D}_\varepsilon^{\mu^{\otimes n}}(f) \geq \mathsf{D}_{\varepsilon+\delta}^{\mu^{\otimes n}}(g)$. Hence as a corollary of Theorem 1 we get that a smoothed version of the influence bound also applies as a lower bound on the distributional query complexity of f.

Corollary 2 *Let* $f : \mathcal{X}^n \to \mathcal{Z}$ *be a function,* μ *be a distribution on* $\mathcal{X}$ *and* $\varepsilon > 0, \delta \geq 0$. *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 $\inf_{\max}$ but close to some other function g with small $\inf_{\max}$. For example, Tribes is OR of $s \approx n/\log_2 n$ AND gates, each of degree $t \approx \log_2 n - \log_2 \log_2 n$. The parameters s, t are so set to make exactly half the inputs being 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 δ-fraction of inputs x and changing their function values to $f(x) = x_1$ (x_1 is the first bit of x) and for the rest $f(x) = g(x)$. Then the first variable has influence at least $\Omega(\delta)$, so applying the old bound only gives a constant lower bound. But g is δ-close to f with $\inf_{\max}(g) = \Theta(\log n/n)$. So the above corollary gives a much better lower bound of $\Theta(n/\log n)$ for the distributional query complexity of f, which the original bound Eq. (1) only gives a constant.

A final comment is that our proof does not 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.

Acknowledgment: We thank Ronald de Wolf for detailed and helpful comments on an earlier draft of the paper.

References

[BdW02]   Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[CK01]   Amit Chakrabarti and Subhash Khot. Improved lower bounds on the randomized complexity of graph properties. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming*, pages 285–296, 2001.

[CR11]   Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of Gap-Hamming-Distance. In *Proceedings of the 43th Annual ACM Symposium on Theory of Computing*, 2011. To appear.

[Haj91]   Péter Hajnal. An $\Omega(n^{4/3})$ lower bound on the randomized complexity of graph properties. *Combinatorica*, 11(2):131–143, 1991.

[JK10]   Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 25th IEEE Conference on Computational Complexity*, pages 247–258, 2010.

[Kla07]   Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM Journal of Computing*, 1:20–46, 2007.

[Kla10]   Hartmut Klauck. A strong direct product theorem for Disjointness. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 77–86, 2010.

[Lee10]   Homin Lee. Decision trees and influence: an inductive proof of the OSSS inequality. *Theory of Computing*, 6(1):81–84, 2010.

[LZ10]   Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 475–489, 2010.

[O'D08]   Ryan O'Donnell. Some topics in analysis of Boolean functions. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 569–578, 2008.

[OS07]   Ryan O'Donnell and Rocco Servedio. Learning monotone decision trees in polynomial time. *SIAM Journal on Computing*, 37(3):827–844, 2007.

[OSSS05] Ryan O'Donnell, Michael E. Saks, Oded Schramm, and Rocco A. Servedio. Every decision tree has an influential variable. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 31–39, 2005.

[She08]   Alexander Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 85–94, 2008.

[SZ09]   Yaoyun Shi and Yufan Zhu. The quantum communication complexity of block-composed functions. *Quantum Information and Computation*, 9(5&6):444–460, 2009.