Electronic Colloquium on Computational Complexity, Report No. 14 (2018)

A Composition Theorem via Conflict Complexity

Swagato Sanyal *

January 10, 2018

Abstract

Let $\mathsf{R}(\cdot)$ stand for the bounded-error randomized query complexity. We show that for any relation $f \subseteq \{0,1\}^n \times \mathcal{S}$ and partial Boolean function $g \subseteq \{0,1\}^n \times \{0,1\}$, $\mathsf{R}_{1/3}(f \circ g^n) = \Omega(\mathsf{R}_{4/9}(f) \cdot \sqrt{\mathsf{R}_{1/3}(g)})$. Independently of us, Gavinsky, Lee and Santha [3] proved this result. By an example demonstrated in their work, this bound is optimal. We prove our result by introducing a novel complexity measure called the *conflict complexity* of a partial Boolean function g, denoted by $\chi(g)$, which may be of independent interest. We show that $\chi(g) = \Omega(\sqrt{\mathsf{R}(g)})$ and $\mathsf{R}(f \circ g^n) = \Omega(\mathsf{R}(f) \cdot \chi(g))$.

1 Introduction

Let $f \subseteq \{0,1\}^n \times \mathcal{S}$ be a relation and $g \subseteq \{0,1\}^m \times \{0,1\}$ be a partial Boolean function. In this work, we bound the bounded-error randomized query complexity of the composed relation $f \circ g^n$ from below in terms of the bounded-error query complexitites of f and g. Our main theorem is as follows.

Theorem 1 (Main Theorem)**.** *For any relation $f \subseteq \{0,1\}^n \times \mathcal{S}$ and partial Boolean function $g \subseteq \{0,1\}^n \times \{0,1\}$,*

$$\mathsf{R}_{1/3}(f \circ g^n) = \Omega\left(\mathsf{R}_{4/9}(f) \cdot \sqrt{\mathsf{R}_{1/3}(g)}\right).$$

Prior to this work, Anshu et. al. [1] proved that $\mathsf{R}_{1/3}(f \circ g^n) = \Omega(\mathsf{R}_{4/9}(f) \cdot \mathsf{R}_{1/2-1/n^4}(g))$. Although in the statement of their result g is stated to be a Boolean function, their result holds even when g is a partial Boolean function.

In the special case of g being a total Boolean function, ben-David and Kothari [2] showed that $\mathsf{R}(f \circ g^n) = \Omega\left(\mathsf{R}(f) \cdot \sqrt{\frac{\mathsf{R}(g)}{\log \mathsf{R}(g)}}\right)$.

Gavinsky, Lee and Santha [3] independently proved Theorem 1 (possibly with different values for the error parameters). They also prove this bound to be tight by exhibiting an example that matches this bound. We believe that our proof is sufficiently different and significantly shorter and simpler than theirs. We draw on and refine the ideas developed in the works of Anshu et. al. and ben-David and Kothari to prove our result.

We define a novel measure of complexity of a partial Boolean function g that we refer to as the *conflict complexity* of g, denoted by $\chi(g)$ (see Section 3 for a definition). This quantity is inspired by the *Sabotage complexity* introduced by ben-David and Kothari. However, the two measures also have important differences. For example, we could show that for any partial function g, $\chi(g)$ and $\mathsf{R}(g)$ are related as follows.

*Division of Mathematical Sciences, Nanyang Technological University, Singapore and Centre for Quantum Technologies, National University of Singapore, Singapore. ssanyal@ntu.edu.sg

ISSN 1433-8092

Theorem 2. *For any partial Boolean function* $g \subseteq \{0,1\}^n \times \{0,1\}$,

$$\chi(g) = \Omega\left(\sqrt{\mathsf{R}_{1/3}(g)}\right).$$

See Section 3 for a proof of Throrem 2. Sabotage complexity is known to be similarly related to the bounded-error randomized query complexity (up to a logarithmic factor) when g is a total Boolean function. For partial Boolean functions, unbounded separation is possible between sabotage complexity and $\mathsf{R}(\cdot)$.

We next prove the following composition theorem.

Theorem 3. *Let* $\mathcal{S}$ *be an arbitrary set,* $f \subseteq \{0,1\}^n \times \mathcal{S}$ *be a relation and* $g \subseteq \{0,1\}^m \times \{0,1\}$ *be a partial Boolean function. Then,*

$$\mathsf{R}_{1/3}(f \circ g^n) = \Omega(\mathsf{R}_{4/9}(f) \cdot \chi(g)).$$

To prove Theorem 3 we draw on the techniques developed by Anshu et. al. and ben-David and Kothari. See Section 5 for a proof of Theorem 3. Theorem 1 follows from Theorems 2 and 3.

2 Preliminaries

A partial Boolean function g is a relation in $\{0,1\}^m \times \{0,1\}$. For $b \in \{0,1\}$, $g^{-1}(b)$ is defined to tbe the set of strings x in $\{0,1\}^n$ for which $(x,b) \in g$ and $(x,\bar{b}) \notin g$. $g^{-1}(0) \cup g^{-1}(0)$ is referred to as the set of valid inputs to g. We assume that for all strings $y \notin g^{-1}(0) \cup g^{-1}(1)$, both $(y,0)$ and $(y,1)$ are in g. For a string $x \in g^{-1}(0) \cup g^{-1}(1)$, $g(x)$ refers to the unique bit b such that $(x,b) \in g$. All the probability distributions μ over the domain of a partial Boolean function g in this paper are assumed to be supported entirely on $g^{-1}(0) \cup g^{-1}(1)$. Thus $g(x)$ is well-defined for any x in the support of μ.

Definition 1 (Bounded-error Randomized Query Complexity)**.** Let $\mathcal{S}$ be any set. Let $h \subseteq \{0,1\}^k \times \mathcal{S}$ be any relation and $\epsilon \in [0,1/2)$. The 2-sided error randomized query complexity $\mathsf{R}_\epsilon(h)$ is the minimum number of queries made in the worst case by a randomized query algorithm $\mathcal{A}$ (the worst case is over inputs and the internal randomness of $\mathcal{A}$) that on each input $x \in \{0,1\}^k$ satisfies $\Pr[(x, \mathcal{A}(x)) \in h] \geq 1 - \epsilon$ (where the probability is over the internal randomness of $\mathcal{A}$).

Definition 2 (Distributional Query Complexity)**.** Let $h \subseteq \{0,1\}^k \times \mathcal{S}$ be any relation, μ a distribution on the input space $\{0,1\}^k$ of h, and $\epsilon \in [0,1/2)$. The distributional query complexity $\mathsf{D}^\mu_\epsilon(h)$ is the minimum number of queries made in the worst case (over inputs) by a deterministic query algorithm $\mathcal{A}$ for which $\Pr_{x\sim\mu}[(x, \mathcal{A}(x)) \in h] \geq 1 - \epsilon$.

In particular, if h is a function and $\mathcal{A}$ is a randomized or distributional query algorithm computing h with error ϵ, then $\Pr[h(x) = \mathcal{A}(x)] \geq 1 - \epsilon$, where the probability is over the respective sources of randomness.

The following theorem is von Neumann's minimax principle stated for decision trees.

Fact 1 (minimax principle)**.** *For any integer* k*, set* $\mathcal{S}$*, and relation* $h \subseteq \{0,1\}^k \times \mathcal{S}$,

$$\mathsf{R}_\epsilon(h) = \max_\mu \mathsf{D}^\mu_\epsilon(h).$$

Let μ be a probabilty distribution over $\{0,1\}^k$. $x \sim \mu$ implies that x is a random string drawn from μ. Let $C \subseteq \{0,1\}^k$ be arbitrary. Then $\mu \mid C$ is defined tobe the probability distribution obtained by conditioning μ on the event that the sampled string belongs to C, i.e.,

$$\mu \mid C(x) = \begin{cases} 0 & \text{if } x \notin C \\ \frac{\mu(x)}{\sum_{y\in C}\mu(y)} & \text{if } x \in C \end{cases}$$

For a partial Boolean function $g : \{0,1\}^m \to \{0,1\}$, probability distribution μ and bit b,

$$\mu_b := \mu \mid g^{-1}(b).$$

Notice that μ_0 and μ_1 are defined with respect to some Boolean function g, which will always be clear from the context.

Definition 3 (Subcube, Co-dimension)**.** A subset $\mathcal{C}$ of $\{0,1\}^m$ is called a subcube if there exists a set $S \subseteq \{1,\ldots,m\}$ of indices and an *assignment function* $A : S \to \{0,1\}$ such that $\mathcal{C} = \{x \in \{0,1\}^m : \forall i \in S, x_i = A(i)\}$. The co-dimension $\mathsf{codim}(\mathcal{C})$ of $\mathcal{C}$ is defined to be $|S|$.

Now we define composition of two relations.

Definition 4 (Composition of relations)**.** We now reproduce from the Section 1 the definition of composed relations. Let $f \subseteq \{0,1\}^n \times \mathcal{S}$ and $g \subseteq \{0,1\}^m \times \{0,1\}$ be two relations. The composed relation $f \circ g^n \subseteq (\{0,1\}^m)^n \times \mathcal{S}$ is defined as follows: For $x = (x^{(1)},\ldots,x^{(n)}) \in (\{0,1\}^m)^n$ and $s \in \mathcal{S}$, $(x,s) \in f \circ g^n$ if and only if there exists $b = (b^{(1)},\ldots,b^{(n)}) \in \{0,1\}^n$ such that for each $i = 1,\ldots,n$, $(x^{(i)}, b^{(i)}) \in g$ and $(b,s) \in f$.

We will often view a deterministic query algorithm as a binary decision tree. In each vertex v of the tree, an input variable is queried. Depending on the outcome of the query, the computation goes to a child of v. The child of v corresponding to outcome b to the query made is denoted by v_b.

It is well known that the set of inputs that lead the computation of a decision tree to a certain vertex forms a subcube. We will denote use the same symbol (e.g. v) to refer to a vertex as well as the subcube associated with it.

The depth of a vertex v in a tree is the number of vertices on the unique path from the root of the tree to v in the tree. Thus, the depth of the root is 1.

Definition 5. Let $\mathcal{A}$ be a decision tree on m bits. Let η_0 and η_1 be two probability distributions with disjoint supports. Let v be a vertex in $\mathcal{A}$. Let variable x_i be queried at v. Then,

$$\Delta^{(v)} := \begin{cases} |\Pr_{x\sim\eta_0}[x_i = 0] - \Pr_{x\sim\eta_1}[x_i = 0]| & \text{if } v \neq \perp. \\ 1 & \text{if } v = \perp. \end{cases}$$

Note that $\Delta^{(v)}$ is defined with respect to distributions η_0 and η_1. In our application, we will often consider a decision tree $\mathcal{A}$, a partial Boolean function g and a probability distributions μ over the inputs. $\Delta^{(v)}$, for a vertex v of $\mathcal{A}$, will then be assumed to be with respect to the distributions $(\mu_b \mid v)_{b\in\{0,1\}}$.

Claim 2. Let $\mathcal{A}$ be a decision tree on m bits. Let g be a partial Boolean function. Let $x \sim \{0,1\}^n$ be sampled from a distribution μ. Let v be a vertex in $\mathcal{A}$. Let variable x_i be queried at v. Then,

$$\mathsf{I}_\mu(g(x) : x_i \mid x \in v) = \mathsf{I}_{\mu|v}(g(x) : x_i) \geq 32\left(\Pr_{x\sim\mu|v}[g(x) = 0] \cdot \Pr_{x\sim\mu|v}[g(x) = 1] \cdot \Delta^{(v)}\right)^2,$$

where $\Delta^{(v)}$ is with respect to the distributions $(\mu_b \mid v)_{b\in\{0,1\}}$.

Proof of Claim 2. Define $b := g(x)$. Condition on the event $x \in v$. Let $(b \otimes x_i)$ be the distribution over pairs of bits, where the bits are distributed independently according to the distributions of b and x_i respectively. We use the equivalence: $\mathsf{I}(b : x_i) = \mathsf{D}((b,x_i)||(b \otimes x_i))$. Now, an application of *Pinsker's inequality* implies that

$$\mathsf{D}((b,x_i)||(b \otimes x_i)) \geq 2||(b,x_i) - (b \otimes x_i)||_1^2. \tag{1}$$

Next, we bound $|(b, x_i) - (b \otimes x_i)||_1$. To this end, we fix bits $z_1, z_2 \in \{0,1\}$, and bound $|\Pr[(b, x_i) = (z_1, z_2)] - \Pr[(b \otimes x_i) = (z_1, z_2)]|$. We have that,

$$\Pr[(b, x_i) = (z_1, z_2)] = \Pr[b = z_1] \Pr[x_i = z_2 \mid b = z_1]. \tag{2}$$

Now,

$$\begin{aligned}\Pr[(b \otimes x_i) = (z_1, z_2)] &= \Pr[b = z_1] \Pr[x_i = z_2] \\ &= \Pr[b = z_1](\Pr[b = z_1] \Pr[x_i = z_2 \mid b = z_1] + \\ &\qquad \Pr[b = \overline{z_1}] \Pr[x_i = z_2 \mid b = \overline{z_1}]).\end{aligned} \tag{3}$$

Taking the absolute difference of (3) and (2) we have that,

$$\begin{aligned}&|\Pr[(b, x_i) = (z_1, z_2)] - \Pr[(b \otimes x_i) = (z_1, z_2)]| \\ &= \Pr[b = z_1] \cdot \Pr[b = \overline{z_1}] \cdot \Delta^{(v)} = \Pr[b = 0] \cdot \Pr[b = 1] \cdot \Delta^{(v)}\end{aligned} \tag{4}$$

The Claim follows by adding (4) over z_1, z_2 and using (1). □

3 Conflict Complexity

In this section, we introduce a randomized process $\mathcal{P}$ (formally given in Algorithm 1). This process is going to play a central role in the proof of our composition theorem (Theorem 3). Later in the section, we use $\mathcal{P}$ to define the *conflict complexity* of a Boolean function g.

Let $n > 0$ be any integer and $\mathcal{B}$ be any deterministic query algorithm that runs on inputs in $(\{0,1\}^m)^n$. $\mathcal{B}$ can be though of as just a query procedure that queries various input variables, and then terminates without producing any output. Let $x = (x_i^{(j)})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ be a generic input to $\mathcal{B}$, and x_i stand for $(x_i^{(j)})_{j=1,\ldots,m}$. For a vertex v of $\mathcal{B}$, $v^{(i)}$ denotes the subcube in v corresponding to x_i, i.e., $v = \times_{i=1}^n v^{(i)}$. Recall from Section 2 that for $b \in \{0,1\}$, v_b stands for the child of v corresponding to the query outcome being b. Let μ_0 and μ_1 be any two probability distributions supported on $g^{-1}(0)$ and $g^{-1}(1)$ respectively. Let $z = (z_1, \ldots, z_n) \in \{0,1\}^n$ be arbitrary. Now consider the probabilistic process $\mathcal{P}$ given by Algorithm 1. Note that $\mathcal{P}$ can be thought of as a randomized query algorithm on input $z \in \{0,1\}^n$, where a query to z_i corresponds to an assignment of 0 to $\mathsf{NOQUERY}_i$ in line 14. This view of $\mathcal{P}$ will be adopted in Section 5.

We now prove an important structural result about $\mathcal{P}$ which will be used many times in our proofs. Consider the following distribution γ_z on $(\{0,1\}^m)^n$: For each i, sample x_i independently from μ_{z_i}.

Let v be a vertex of $\mathcal{B}$. Let $A_{\mathcal{B}}(v)$ be the event that process $\mathcal{P}$ reaches node v, and $B_{\mathcal{B}}(v)$ be the event that for a random input x sampled from γ_z, the computation of $\mathcal{B}$ reaches node v.

Claim 3. For each vertex v of $\mathcal{B}$,

$$\Pr[A_{\mathcal{B}}(v)] = \Pr[B_{\mathcal{B}}(v)].$$

Proof. We will prove by induction on the depth t of v, i.e., the number of vertices on the unique path from the root to v in $\mathcal{B}$.

Base case: $t = 1$. v is the root of $\mathcal{B}$. Thus $\Pr[A_{\mathcal{B}}(v)] = \Pr[B_{\mathcal{B}}(v)] = 1$.

Inductive step: Assume that $t \geq 2$, and that the statement is true for all vertices at depth at most $t-1$. Since $t \geq 2$, v is not the root of $\mathcal{B}$. Let u be the ancestor of v, and variable $x_i^{(j)}$ be queried at u. without loss of generality assume that v is the child of u corresponding to $x_i^{(j)} = 0$. We split the proof into the following two cases.

Algorithm 1: $\mathcal{P}$ on $\mathcal{B}, \mu_0, \mu_1, z$

for $1 \le k \le n$ **do**
  $\mathsf{NOQUERY}_k \leftarrow 1$.
  $\mathsf{N}_k \leftarrow 0$.
$v \leftarrow$Root of $\mathcal{B}$ // Corresponds to $\{0,1\}^m$
while *v is not a leaf of* $\mathcal{B}$ **do**
  Let $\mathcal{B}$ query $x_i^{(j)}$ at v.
  if $\mathsf{NOQUERY}_i = 1$ **then**
    Sample a fresh real number $r \sim [0,1]$ uniformly at random.
    if $r \le \min_b \Pr_{x_i \sim \mu_b}[x_i^{(j)} = 0 \mid x_i \in v_i]$ **then**
      $v \leftarrow v_0$.
    else if $r \ge \max_b \Pr_{x_i \sim \mu_b}[x_i^{(j)} = 0 \mid x_i \in v^{(i)}]$ **then**
      $v \leftarrow v_1$.
    else
      $\mathsf{NOQUERY}_i \leftarrow 0$.
      if $r \le \Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in v^{(i)}]$ **then**
        $v \leftarrow v_0$.
      else
        $v \leftarrow v_1$.
    $\mathsf{N}_i \leftarrow \mathsf{N}_i + 1$.
  else
    Sample b from the distribution μ_{z_i} conditioned on the event $x_i \in v^{(i)}$.
    $v \leftarrow v_b$.

- **Case 1:** $\Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i] \le \Pr_{x_i \sim \mu_{\overline{z_i}}}[x_i^{(j)} = 0 \mid x_i \in u_i]$.

  Condition on $A_{\mathcal{B}}(u)$ and $\mathsf{NOQUERY}_i = 0$. The probability that $\mathcal{P}$ reaches v is $\Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i]$. Now, condition on $A_{\mathcal{B}}(u)$ and $\mathsf{NOQUERY}_i = 1$. The probability that $\mathcal{P}$ reaches v is exactly equal to the probability that the real number r sampled at v lies in $[0, \Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i]]$, which is equal to $\Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i]$. Thus,

$$\begin{aligned}\Pr[A_{\mathcal{B}}(v] &= \Pr[A_{\mathcal{B}}(u)].\Pr[A_{\mathcal{B}}(v) \mid A_{\mathcal{B}}(u)] \\ &= \Pr[A_{\mathcal{B}}(u)] \cdot \Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i]. \end{aligned} \tag{5}$$

  Now condition on $B_{\mathcal{B}}(u)$. The probability that $\mathcal{B}$ reaches v is exactly equal to the probability that $x_i^{(j)} = 0$ when x is sampled according to the distribution γ_z conditioned on the event that $x \in u$. Note that in the distribution γ_z, the x_k's are independently distributed. Thus,

$$\begin{aligned}\Pr[B_{\mathcal{B}}(v)] &= \Pr[B_{\mathcal{B}}(u)].\Pr[B_{\mathcal{B}}(v) \mid B_{\mathcal{B}}(u)] \\ &= \Pr[B_{\mathcal{B}}(u)] \cdot \Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i]. \end{aligned} \tag{6}$$

  By the inductive hypothesis, $\Pr[A_{\mathcal{B}}(u)] = \Pr[B_{\mathcal{B}}(u)]$. The claim follows from (5) and (6).

- **Case 2:** $\Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in u_i] > \Pr_{x_i \sim \mu_{\overline{z_i}}}[x_i^{(j)} = 0 \mid x_i \in u_i]$. Let v' be the child of u corresponding to $x_i^{(j)} = 1$. By an argument similar to Case 1, we have that

$$\Pr[A_{\mathcal{B}}(v')] = \Pr[B_{\mathcal{B}}(v')]. \tag{7}$$

Now,

$$\begin{aligned}\Pr[A_{\mathcal{B}}(v)] &= \Pr[A_{\mathcal{B}}(u)] - \Pr[A_{\mathcal{B}}(v')] \\ &= \Pr[B_{\mathcal{B}}(u)] - \Pr[A_{\mathcal{B}}(v')] \quad \text{(By inductive hypothesis)} \\ &= \Pr[B_{\mathcal{B}}(u)] - \Pr[B_{\mathcal{B}}(v')] \quad \text{(By (7))} \\ &= \Pr[B_{\mathcal{B}}(v].\end{aligned}$$

□

Let $n = 1, z \in \{0,1\}$, and $\mathcal{B}$ be a decision tree that computes g. Consider process $\mathcal{P}$ on $\mathcal{B}, \mu_0, \mu_1, z$. Note that $\mathsf{NOQUERY}_1$ is set to 0 with probability 1. To see this observe that as long as $\mathsf{NOQUERY}_1 = 1$, the current subcube v contains strings from the supports of both μ_0 and μ_1, and hence from both $g^{-1}(0)$ and $g^{-1}(1)$. If $\mathsf{NOQUERY}_1$ is not set to 0 for the entire run of $\mathcal{P}$, then there exist inputs $x \in g^{-1}(0), x' \in g^{-1}(1)$ which belong to the same leaf of $\mathcal{B}$, contradicting the hypothesis that $\mathcal{B}$ computes g. Let the random variable $\mathcal{N}$ stand for the value of the variable N_1 after the termination of $\mathcal{P}$. Note that $\mathcal{N}$ is equal to the the index of the iteration of the while loop in which $\mathsf{NOQUERY}_1$ is set to 0. The distribution of $\mathcal{N}$ depends on μ_0, μ_1 and $\mathcal{B}$, which in our applications will either be clear from the context, or clearly specified. Note that the distribution of $\mathcal{N}$ is independent of the value of z.

Definition 6. The *conflict complexity* of a partial function g with respect to distributions μ_0 and μ_1 supported on $g^{-1}(0)$ and $g^{-1}(1)$ respectively, and decision tree $\mathcal{B}$ computing g, is defined as:

$$\chi(\mu_0, \mu_1, \mathcal{B}) = \mathsf{E}[\mathcal{N}].^{1}$$

The conflict complexity of g is defined as:

$$\chi(g) = \max_{\mu_0,\mu_1} \min_{\mathcal{B}} \chi(\mu_0, \mu_1, \mathcal{B}).$$

Where the maximum is over distributions μ_0 and μ_1 supported on $g^{-1}(0)$ and $g^{-1}(1)$ respectively, and the minimum is over decision trees $\mathcal{B}$ computing g.

For a pair (μ_0, μ_1) of distributions, let $\mathcal{B}$ be the decision tree computing g such that $\mathsf{E}[\mathcal{N}]$ is minimized. We call such a decision tree an *optimal* decision tree for μ_0, μ_1. We conclude this section by making an important observation about the structure of optimal decision trees. Let v be any node of $\mathcal{B}$. Let $\mu_0' := \mu_0 \mid v$ and $\mu_1' := \mu_1 \mid v$. Let $\mathcal{B}_v$ denote the subtree of $\mathcal{B}$ rooted at v. We observe that $\mathcal{B}_v$ is an optimal tree for μ_0' and μ_1'; if it is not then we could replace it by an optimal tree for μ_0' and μ_1', and for the resultant tree, the expected value of $\mathcal{N}$ with respect to μ_0 and μ_1 will be smaller than that in $\mathcal{B}$. This will contradict the optimality of $\mathcal{B}$. This recursive sub-structure property of optimal trees will be helpful to us.

4 Conflict Complexity and Randomized Query Complexity

In this section, we will prove Theorem 2 (restated below).

Theorem 2. *For any partial Boolean function $g \subseteq \{0,1\}^n \times \{0,1\}$,*

$$\chi(g) = \Omega\left(\sqrt{\mathsf{R}_{1/3}(g)}\right).$$

[1] As observed before, the choices of μ_0, μ_1 and $\mathcal{B}$ are built into the definition of $\mathcal{N}$.

Proof. We will bound the distributional query complexity of g for each input distribution μ with rspect to error $47/95 < 1/2$, $\mathsf{D}^{\mu}_{47/95}(g)$, from above by $O(\chi(g)^2)$. Theorem 2 will follow from the *minimax principle* (Fact 1), and the observation that the error can be brought down to $1/3$ by constantly many independent repetitions followed by a selection of the majority of the answers. It is enough to consider distributions μ supported on valid inputs of g. To this end, fix a distribution μ supported only on $g^{-1}(0) \cup g^{-1}(1)$.

Let $\chi(g) = d$. Let μ_b be the distribution obtained by conditioning μ on the event $g(x) = b$. Let $\mathcal{B}$ be an optimal decision tree for distributions μ_0 and μ_1. Clearly $\mathsf{E}[\mathcal{N}] \leq \chi(g) = d$.

We first prove some structural results about $\mathcal{B}$. Let $\mathcal{B}$ be run on a random input x sampled according to μ. Let v_t be the random vertex at which the t-th query is made; If $\mathcal{B}$ terminates before making t queries, define $v_t := \perp$. Let $\mathcal{E}$ be any event which is a collection of possible transcripts of $\mathcal{B}$, such that $\Pr[\mathcal{E}] \geq \frac{3}{4}$. Recall from Section 2 that for any vertex v of $\mathcal{B}$, $\Delta^{(v)}$ is assumed to be with respect to the probability distribution $\mu \mid v$.

Claim 4.

$$\sum_{t=1}^{10d} \mathsf{E}[\Delta^{(v_t)} \mid \mathcal{E}] \geq \frac{13}{20}.$$

Proof. Let us sample vertices u_t of $\mathcal{B}$ as follows:

1. Set $z = \begin{cases} 1 & \text{with probability } \Pr_{x\sim\mu}[g(x) = 1], \\ 0 & \text{with probability } \Pr_{x\sim\mu}[g(x) = 0] \end{cases}$
2. Run process $\mathcal{P}$ for $\mathcal{B}, \mu_0, \mu_1, z$.
3. Let u_t be the vertex v in the beginning of the t-th iteration of the *while* loop of Algorithm 1. Return $(u_t)_{t=1,\ldots}$. If the simulation stops after i iterations, set $u_t := \perp$ for all $t > i$.

By Claim 3, and since z has the same distribution as that of $g(x)$ where x is sampled from μ, the vertices u_t and v_t have the same distribution. In the above sampling process for each $t = 1, \ldots, 10d$, let E_t be the event that $\mathsf{NOQUERY}_1 = 1$ in the beginning of the t-th iteration of the *while* loop of Algorithm 1. Conditioned on $\mathcal{E}$, the probability that $\mathsf{NOQUERY}_1$ is set to 0 in the t-th iteration is $\Pr[E_t \mid \mathcal{E}] \cdot \mathsf{E}[\Delta^{(u_t)} \mid E_t, \mathcal{E}]$[2]. By union bound we have that,

$$\begin{aligned}
\sum_{t=1}^{10d} \mathsf{E}[\Delta^{(v_t)} \mid \mathcal{E}] &= \sum_{t=1}^{10d} \mathsf{E}[\Delta^{(u_t)} \mid \mathcal{E}] \\
&\geq \sum_{t=1}^{10d} \Pr[E_t \mid \mathcal{E}] \cdot \mathsf{E}[\Delta^{(u_t)} \mid E_t, \mathcal{E}] \\
&\geq \Pr\left[\overline{\bigcap_{t=1}^{10d} E_t} \mid \mathcal{E}\right] \\
&\geq \Pr\left[\overline{\bigcap_{t=1}^{10d} E_t}\right] - \Pr[\overline{\mathcal{E}}]. \qquad (8)
\end{aligned}$$

Now, since $\mathsf{E}[\mathcal{N}] \leq \chi(g) = d$, we have by *Markov's inequality* that the probability that the process $\mathcal{P}$, when run for $\mathcal{B}, \mu_0, \mu_1$ and the random bit z generated as above[3], sets $\mathsf{NOQUERY}_1$ to 0 within first

[2]Note that conditioned on E_t, $u_t \neq \perp$.
[3]Recall that the distribution of $\mathcal{N}$ is independent of z.

$10d$ iterations of the *while* loop, is at least $9/10$. Thus we have that,

$$\Pr\left[\bigcap_{t=1}^{10d} E_t\right]^c \geq \frac{9}{10}. \tag{9}$$

The claim follows from (8), (9) and the hypothesis $\Pr[\mathcal{E}] \geq \frac{3}{4}$. □

The next Lemma follows from Claim 4 and the recursive sub-structure property of optimal trees discussed in the last paragraph of Section 3.

Lemma 5. *Let i be any positive integer. Then,*

$$\sum_{t=1}^{10di} \mathsf{E}[\Delta^{(v_t)} \mid \mathcal{E}] \geq \frac{13i}{20}.$$

Notice that if $\mathcal{B}$ terminates before making t queries, $v_t = \perp$ and $\Delta^{(v_t)} = 1$.

Proof of Lemma 5. For $j = 1, \ldots, i$, let w be any vertex at depth $10jd + 1$. Consider the subtree T of $\mathcal{B}$ rooted at w. By the recursive sub-structure property of $\mathcal{B}$, T is an optimal tree for distributions $\mu'_0 := \mu_0 \mid w, \mu'_1 := \mu_1 \mid w$. Let w_t be the random vertex at depth t of T, when T is run on a random input from $\mu \mid w$. By Claim 4, we have that,

$$\sum_{t=1}^{10d} \mathsf{E}[\Delta^{(w_t)} \mid \mathcal{E}] \geq \frac{13}{20}. \tag{10}$$

In (10), $\Delta^{(w_t)}$ is with respect to distributions $\mu'_0 \mid w_t = \mu_0 \mid w_t, \mu'_1 \mid w_t = \mu_1 \mid w_t$. Now, when w is the random vertex v_{10jd+1}, w_t is the random vertex v_{10jd+t}. Thus from (10) we have that,

$$\sum_{t=10jd+1}^{10(j+1)d} \mathsf{E}[\Delta^{(v_t)} \mid \mathcal{E}] \geq \frac{13}{20}. \tag{11}$$

The claim follows by adding (11) over $j = 0, \ldots, i-1$. □

We now finish the proof of Theorem 2 by showing that $D^{\mu}(g) = O(d^2)$. Let x be distributed according to μ, and $\mathcal{B}$ be run on x. Let BIASED denote the event that in at most $10d^2$ queries, the computation of $\mathcal{B}$ reaches a vertex v for which $\Pr_{x\sim\mu}[g(x) = 0 \mid x \in v] \cdot \Pr_{x\sim\mu}[g(x) = 1 \mid x \in v] \leq \frac{1}{9}$. Let STOP denote the event that $\mathcal{B}$ terminates after making at most $10d^2$ queries. Let $\mathcal{E} := \overline{\mathsf{BIASED} \vee \mathsf{STOP}}$.

Consider the following decision tree $\mathcal{B}'$: Start simulating $\mathcal{B}$. Terminate the simulation if one of the following events occurs. The outputs in each case is specified below.

1. (*Event* STOP) If $\mathcal{B}$ terminates, terminate and output what $\mathcal{B}$ outputs.
2. If $10d^2$ queries have been made and the computation is at a vertex v, terminate and output $\arg\max_b \Pr[g(x) = b \mid x \in v]$.

By construction, $\mathcal{B}'$ makes at most $10d^2$ queries in the worst case. We shall show that $\Pr_{x\sim\mu}[\mathcal{B}'(x) \neq g(x)] \leq \frac{47}{95} < \frac{1}{2}$. This will prove Theorem 2.

We split the proof into the following two cases.

Case 1: $\Pr[\overline{\mathcal{E}}] \geq \frac{1}{4}$.

First, condition on the event that the computation reaches a vertex v for which $\Pr_{x\sim\mu}[g(x)=0 \mid x \in v]\cdot\Pr_{x\sim\mu}[g(x)=1 \mid x\in v] \leq \frac{1}{9}$ holds. Thus one of $\Pr_{x\sim\mu}[g(x)=0 \mid x\in v]$ and $\Pr_{x\sim\mu}[g(x)=1 \mid x\in v]$ is at most $1/3$. Hence, $|\Pr_{x\sim\mu}[g(x)=0 \mid x\in v]-\Pr_{x\sim\mu}[g(x)=1 \mid x\in v]| \geq 2/3$. Let m be the random leaf of the subtree of $\mathcal{B}'$ rooted at v at which the computation ends. The probability that $\mathcal{B}'$ errs is at most

$$\begin{aligned}
&\mathsf{E}_{x\sim\mu|v}\left[\frac{1}{2}-\frac{1}{2}\left|\Pr_{x\sim\mu}[g(x)=0 \mid x\in m]-\Pr_{x\sim\mu}[g(x)=1 \mid x\in m]\right|\right].\\
&\leq \frac{1}{2}-\frac{1}{2}\left|\mathsf{E}_{x\sim\mu|v}\Pr_{x\sim\mu}[g(x)=0 \mid x\in m]-\mathsf{E}_{x\sim\mu|v}\Pr_{x\sim\mu}[g(x)=1 \mid x\in m]\right|\\
&\qquad\qquad\text{(By Jensen's inequality)}\\
&=\frac{1}{2}-\frac{1}{2}\left|\Pr_{x\sim\mu}[g(x)=0 \mid x\in v]-\Pr_{x\sim\mu}[g(x)=1 \mid x\in v]\right| \leq \frac{1}{3}.
\end{aligned}$$

Then, condition on the event STOP. The probability that $\mathcal{B}'$ errs is $0 \leq \frac{1}{3}$.

Thus we have shown that conditioned on $\overline{\mathcal{E}}$ the probability that $\mathcal{B}'$ errs is at most $\frac{1}{3}$. Thus the probability that $\mathcal{B}'$ errs is at most $\frac{1}{4}\cdot\frac{1}{3}+\frac{3}{4}\cdot\frac{1}{2}=\frac{11}{24}<\frac{47}{95}$.

Case 2: $\Pr[\overline{\mathcal{E}}] < \frac{1}{4}$.

By Claim 5 we have that

$$\sum_{t=1}^{10d^2}\mathsf{E}[\Delta^{v^{(t)}} \mid \mathcal{E}] \geq \frac{13d}{20}. \tag{12}$$

Let $a_i := (x_i, b_i)$ be the tuple formed by the random input variable x_i queried at the i-th step by $\mathcal{B}'$, and the outcome b_i of the query; if $\mathcal{B}'$ terminates before i-th step, $a_i := \perp$. Notice that the vertex v_i at which the i-th query is made is determined by $(a_1,\ldots,a_{i-1})$ and vice versa. We have,

$$\begin{aligned}
&\mathsf{I}(a_1,\ldots,a_{10d^2}:g(x))\\
&=\sum_{i=1}^{10d^2}\mathsf{I}(a_i:g(x)\mid a_1,\ldots,a_{i-1}) \qquad \text{(Chain rule of mutual information)}\\
&=\sum_{i=1}^{10d^2}\mathsf{I}(b_i:g(x)\mid v_i)\\
&\geq 32\sum_{i=1}^{10d^2}\mathsf{E}\left[\mathbf{1}_{v_i\neq\perp}\cdot\left[\Pr[g(x)=0 \mid x\in v_i]\cdot\Pr[g(x)=1 \mid x\in v_i]\cdot\Delta^{(v_i)}\right]^2\right]\\
&\qquad\qquad\text{(From Claim 2)}\\
&\geq 32\sum_{i=1}^{10d^2}\Pr[\mathcal{E}]\cdot\mathsf{E}\left[\left[\Pr[g(x)=0 \mid x\in v_{i-1}]\cdot\Pr[g(x)=1 \mid x\in v_{i-1}]\cdot\Delta^{(v_i)}\right]^2 \mid \mathcal{E}\right]\\
&\qquad\qquad\text{(Conditioned on } \mathcal{E}, v_i\neq\perp)\\
&\geq 32\sum_{i=1}^{10d^2}\frac{3}{4}\cdot\frac{1}{9}\cdot\mathsf{E}[{\Delta^{(v_i)}}^2 \mid \mathcal{E}]\\
&=\frac{8}{3}\sum_{i=1}^{10d^2}\mathsf{E}[{\Delta^{(v_i)}}^2 \mid \mathcal{E}] \qquad \text{(By the assumption } \Pr[\overline{\mathcal{E}}]\leq\tfrac{1}{4}\text{)}
\end{aligned}$$

$$\geq \frac{8}{3} \cdot \frac{1}{10d^2} \left(\sum_{i=1}^{10d^2} \mathsf{E}[\Delta^{(v_i)} \mid \mathcal{E}] \right)^2 \text{ (By Cauchy-Schwarz inequality)}$$

$$\geq \frac{1}{10}. \quad \text{(From (12))} \tag{13}$$

Hence, from (13) we have

$$\mathsf{H}(g(x) \mid a_1, \ldots a_{v_{10d^2}}) \leq 1 - \frac{1}{10} = \frac{9}{10}. \tag{14}$$

Let $\mathcal{L}$ be the set of leaves ℓ of $\mathcal{B}'$ such that $\mathsf{H}(g(x) \mid \ell) \leq \frac{19}{20}$. For each $\ell \in \mathcal{L}$, $\min_b \Pr_{x \sim \mu}[g(x) = b \mid x \in \ell] \leq \frac{2}{5}$. Conditioned on $(a_1, \ldots, a_{10d^2}) \in \mathcal{L}$, the probability that $\mathcal{B}'$ errs is at most $\frac{2}{5}$. By *Markov's inequality* and (14), it follows that $\Pr[(a_1, \ldots, a_{10d^2}) \in \mathcal{L}] \geq \frac{1}{19}$. Thus $\mathcal{B}'$ errs with probability at most $\frac{1}{19} \cdot \frac{2}{5} + \frac{18}{19} \cdot \frac{1}{2} = \frac{47}{95}$.

□

5 The Composition Theorem

In this section we prove Theorem 3 (restated below).

Theorem 3. *Let $\mathcal{S}$ be an arbitrary set, $f \subseteq \{0,1\}^n \times \mathcal{S}$ be a relation and $g \subseteq \{0,1\}^m \times \{0,1\}$ be a partial Boolean function. Then,*

$$\mathsf{R}_{1/3}(f \circ g^n) = \Omega(\mathsf{R}_{4/9}(f) \cdot \chi(g)).$$

Proof. We shall prove that for each distribution η on the inputs to f, there is a query algorithm $\mathcal{A}$ making $O(\mathsf{R}(f \circ g^n)/\chi(g))$ queries in the worst case, for which $\Pr_{z \in \nu}[(z, \mathcal{A}(z)) \in f] \geq \frac{5}{9}$ holds. This will imply the theorem by *Yao's minimax principle*. To this end let us fix a distribution η over $\{0,1\}^n$.

Let $\chi(g) = d$. Thus, there is a *hard* pair of distributions μ_0, μ_1, supported on $g^{-1}(0)$ and $g^{-1}(1)$ respectively, such that for every decision tree $\mathcal{B}$ that computes g, $\chi(\mu_0, \mu_1, g) \geq d$. We will use distributions η, μ_0 and μ_1 to set up a distribution γ_η over the input space of $f \circ g^n$. For a fixed $z = (z_1, \ldots, z_n) \in \{0,1\}^n$, We recall the distribution γ_z over $(\{0,1\}^m)^n$ from Section 3. γ_z is given by the following sampling procedure:

1. For $i = 1, \ldots, n$, sample $x_i = (x_i^{(j)})_{j=1,\ldots,m}$ from μ_{z_i} independently for each i.
2. return $x = (x_i)_{i=1,\ldots,n}$.

Now, let γ_η be the distribution over $(\{0,1\}^m)^n$ that is given by the following sampling procedure:

1. Sample $z = (z_1, \ldots, z_n)$ from η.
2. Sample $x = (x_i)_{i=1,\ldots,n}$ from γ_z. Return x.

Observe that for each z, x sampled as above, for each $s \in \mathcal{S}$, $(z, s) \in f$ *if and only if* $(x, s) \in f \circ g^n$.

Assume that $\mathsf{R}_{1/3}(f \circ g^n) = t$. Yao's mimimax principle implies that there is a deterministic query algorithm $\mathcal{A}'$ for inputs from $(\{0,1\}^m)^n$, that makes at most t queries in the wors case, such that $\Pr_{x \in \gamma_\nu}[(x, \mathcal{A}'(x)) \in f \circ g^n] \geq \frac{2}{3}$. We will first use $\mathcal{A}'$ to construct a randomized algorithm T for f, whose accuracy is as desired, and for which the expected number of queries made is small.

Algorithm 2: T on z

for $1 \leq k \leq n$ **do**
  $\mathsf{NOQUERY}_k \leftarrow 1$.
  $\mathsf{N}_k \leftarrow 0$.
$v \leftarrow$Root of $\mathcal{A}'$ // Corresponds to $\{0,1\}^m$
while *v is not a leaf of $\mathcal{A}'$* **do**
  Let $\mathcal{A}'$ query $x_i^{(j)}$ at v.
  if $\mathsf{NOQUERY}_i = 1$ **then**
    Sample a fresh real number $r \sim [0,1]$ uniformly at random.
    if $r \leq \min_b \Pr_{x_i \sim \mu_b}[x_i^{(j)} = 0 \mid x_i \in v_i]$ **then**
      $v \leftarrow v_0$.
    else if $r \geq \max_b \Pr_{x_i \sim \mu_b}[x_i^{(j)} = 0 \mid x_i \in v^{(i)}]$ **then**
      $v \leftarrow v_1$.
    else
      $\mathsf{NOQUERY}_i \leftarrow 0$.
      Query z_i.
      if $r \leq \Pr_{x_i \sim \mu_{z_i}}[x_i^{(j)} = 0 \mid x_i \in v^{(i)}]$ **then**
        $v \leftarrow v_0$.
      else
        $v \leftarrow v_1$.
    $\mathsf{N}_i \leftarrow \mathsf{N}_i + 1$.
    else
      Sample b from the distribution μ_{z_i} conditioned on the event $x_i \in v^{(i)}$.
      $v \leftarrow v_b$.

T, described formally in Algorithm 2, is essentially viewing the process $\mathcal{P}$ for z, μ_0, μ_1, A' as a query algorithm runnng on input z; an assignment of 0 to $\mathsf{NOQUERY}_i$ corresponds to a query to z_i. By Claim 3, we have that for each $z \in \{0,1\}^n$, $\Pr[(z, T(z)) \in f] = \Pr_{x \sim \gamma_z}[(x, A'(x)) \in f \circ g^n]$. Thus, $\Pr_{z \sim \eta}[(z, T(z)) \in f] = \Pr_{x \sim \gamma_\eta}[(x, A'(x)) \in f \circ g^n] \geq \frac{2}{3}$.

We now bound the expected number of queries made by T on each z. For doing that we consider the following randomized process Q that acts on z. Let $\mathcal{B}$ be an optimal tree for distributions μ_0, μ_1. Q is described formally in Algorithm 3. Since $\mathcal{B}$ computes g, process Q is guaranteed to set $\mathsf{NOQUERY}_i$

Algorithm 3: Q on z

Run T on z.
for $1 \leq i \leq n$ **do**
  if $\mathsf{NOQUERY}_i = 1$ **then**
    Run process $\mathcal{P}$ on $\mathcal{B}, \mu_0, \mu_1, x_i$ until $\mathsf{NOQUERY}_i$ is set to 0.

to 0 for each i. In steps 1 and 4, the process $\mathcal{P}$ is run with trees A' and $\mathcal{B}$, and the trees make queries inside the for loop of $\mathcal{P}$. These queries can be thought of as being made to an mn bit string $(x_i^{(j)})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$. Let the random variable X_i stand for the total number of queries made by these trees in x_i. $X = \sum_{i=1}^n X_i$ is the total number of queries in Q, i.e., the total number of iterations of the for loop of $\mathcal{P}$ in all the runs of $\mathcal{P}$ in Q. The next claim bounds $\mathsf{E}X$ from below.

Claim 6.

$$\mathsf{E}X \geq nd.$$

Proof. Towards a contradiction assume that $\mathsf{E}X < nd$. Thus there exists an i such that $\mathsf{E}X_i < d$. Notice that this expectation is over the random real numbers sampled in the for loop of $\mathcal{P}$. Thus, there exists a fixing of those real numbers r that are sampled in those iterations of the for loop of $\mathcal{P}$ that correspond to queries into x_j for $j \neq i$, such that conditioned on that fixing, $\mathsf{E}X_i < d$. However, under that fixing, process Q is equivalent to process $\mathcal{P}$ for some deterministic decision tree T' that computes $g(x_i)$ (since $\mathsf{NOQUERY}_i$ is set to 0 with probability 1), μ_0, μ_1 and z_i. Thus $\mathsf{E}X_i < d$ conditioned on the above-mentioned fixing of randomness contradicts the assumption that $\min_{\mathcal{B}} \chi(\mathcal{B}, \mu_0, \mu_1) = \chi(g) = d$, where the minimum is taken over all deterministic decision tree β that computes g. □

Now, let Y denote the size of the random set $\{i \mid \mathsf{NOQUERY}_i \text{ is set to 0 in step 1 in } Q\}$. Now, conditoned on the event $Y = b$, the expected number of queries made in step 4 of Q is $(n-b)d = nd - bd$. So under this conditioning the total number of queries X made by Q is at most $t + nd - bd$. Taking expectation over b, and using Claim 6 we have that

$$t + nd - d \cdot \mathsf{E}Y \geq nd \Longrightarrow \mathsf{E}Y \leq \frac{t}{d}.$$

Observing that for each z, Y has the same distribution as the number of queries made by T when run on z, we conclude that for each z, T makes at most t/d queries on expectation. By Markov's inequality, the probability that T makes more than $9t/d$ queries is at most $1/9$. Thus the probabilistic algorithm $\mathcal{A}''$ obtained by terminating T after $10t/d$ queries computes f with probability at least $2/3 - 1/9 = 5/9 > 1/2$ on a random input from η. By fixing the randomness of $\mathcal{A}'$ appropriately we get a deterministic algorithm $\mathcal{A}$ of complexity $O(t/d) = O(\mathsf{R}(f \circ g)/\chi(g))$ such that $\Pr_{z\sim\eta}[(z, \mathcal{A}(z)) \in f] \geq \frac{5}{9}$. □

Acknowledgements. I thank Rahul Jain for helpful discussions.

This material is based on research supported by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

References

[1] Anurag Anshu, Dmitry Gavinsky, Rahul Jain, Srijita Kundu, Troy Lee, Priyanka Mukhopadhyay, Miklos Santha, and Swagato Sanyal. A composition theorem for randomized query complexity. In *FSTTCS*, 2017.

[2] Shalev Ben-David and Robin Kothari. Randomized query complexity of sabotaged and composed functions. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 60:1–60:14, 2016.

[3] Dmitry Gavinsky, Troy Lee, and Miklos Santha. On the randomised query complexity of composition. *CoRR*, abs/1801.02226, 2018.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il