Electronic Colloquium on Computational Complexity, Report No. 25 (2000)

Super-linear time-space tradeoff lower bounds for randomized computation

Paul Beame*
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
beame@cs.washington.edu

Michael Saks†
Dept. of Mathematics
Rutgers University
New Brunswick, NJ
saks@math.rutgers.edu

Xiaodong Sun†
Dept. of Mathematics
Rutgers University
New Brunswick, NJ
sunxd@math.rutgers.edu

Erik Vee
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
env@cs.washington.edu

May 19, 2000

Abstract

We prove the first time-space lower bound tradeoffs for randomized computation of decision problems. The bounds hold even in the case that the computation is allowed to have arbitrary probability of error on a small fraction of inputs. Our techniques are an extension of those used by Ajtai [Ajt99a, Ajt99b] in his time-space tradeoffs for deterministic RAM algorithms computing element distinctness and for Boolean branching programs computing a natural quadratic form.

Ajtai's bounds were of the following form: if the machine uses at most kn time for some constant k it requires space at least $\epsilon_k n$ for some constant ϵ_k. In this paper we provide an explicit relationship between ϵ_k and k that also achieves larger lower bounds than those of [Ajt99a, Ajt99b]. In particular, we obtain time-space tradeoff lower bounds of the form $T = \Omega(n\sqrt{\log / \log\log(n/S)})$, which implies that if the space is $n^{1-\epsilon}$ then the time used is $\Omega(n\sqrt{\log / \log\log(n)})$.

1 Introduction

The study of time-space tradeoffs for computational problems is fundamental to complexity theory. These tradeoffs were considered early in the history of complexity [Cob66], and have continued to be an important area of research [Bor93]. An important motivation for these investigations was the observation that for some natural problems such as sorting, algorithms were known that were extremely space efficient, and other algorithms were known that were very time efficient, but

* Research supported by NSF grant CCR-9800124.
†Research supported by NSF grant CCR-9700239 and by DIMACS

ISSN 1433-8092

no known algorithm could simultaneously achieve the optimal time and space efficiency. Another motivation comes from the general study of lower bounds where time-space tradeoff lower bounds can be viewed as milestones towards proving nontrivial (superlogarithmic) space lower bounds for problems in P (or even NP).

As with most lower bound problems in complexity theory, work on the problem divides into "uniform" and "nonuniform" models. In the uniform setting, a series of recent papers have established limitations on Turing machines computing SAT. The first work along these lines was by Fortnow [For97], which was followed by [LV99] and [FvM00]. The latter gives the best current result: any algorithm for SAT that runs in space $n^{o(1)}$ requires time at least $\Omega(n^{\phi-\epsilon})$ where $\phi = (\sqrt{5}-1)/2$ and ϵ is any positive constant. Although some of these lower bounds apply even to co-nondeterministic computation, none of them give any results for randomized algorithms.

In the nonuniform setting, the standard model is the *branching program*. In this model, a program for computing a function $f(x_1, \ldots, x_n)$ (where the variables take values in some finite domain D) is represented as a DAG with a unique start node. Each non-sink node is labeled by a variable and the arcs out of a node correspond to the possible values of the variable. The sink nodes are labeled by outputs 0 or 1. Executing the program on a given input corresponds to following a path from the start node using the values of the input variables to determine the arcs to follow. The maximum length of a path corresponds to time and the logarithm of the number of nodes corresponds to space. This model is often called the D-way branching program model; in the case that the domain D is $\{0,1\}$ is referred to as *Boolean branching program* model.

In this model (or more precisely an extension which permits outputs during the course of computation), there was considerable success in proving time-space tradeoff lower bounds for *multi-output functions* such as sorting, pattern matching, matrix-vector product and hashing [BC82, Bea91, Abr90, Abr91, MNT93]. However, these techniques fundamentally break down when considering decision problems. In the *comparison branching program model* (where the inputs are numbers and the tests at nodes are pairwise comparisons of inputs) there were strong results obtained for element distinctness [BFMadH$^+$87, Yao88]. In the Boolean model, there is an extensive literature on various restricted *read-k* models ([BRS93]) which have strict limitations on the number of times that any one variable may appear on any path in the branching program.

Recently, the first results have been obtained for decision problems on unrestricted branching programs using time more than n. In the D-way model, [BST98] exhibited a problem in P, where the domain D grows with the number of variables n, for which any subexponential size nondeterministic branching program has depth $\Omega(n \log\log n)$. In the boolean case, they obtained the first (barely) nontrivial bound by exhibiting a problem in P and a constant $\epsilon > 0$ for which any subexponential size branching program requires depth at least $(1+\epsilon)n$. Extending techniques of [BRS93] for bilinear forms, the lower bounds in [BST98] were shown for functions based on quadratic forms over finite fields.

In a remarkable breakthrough, Ajtai [Ajt99b] exhibited a P-time computable Boolean function (also based on quadratic forms) for which any subexponential size deterministic branching program requires superlinear depth. Much of the technical argument for this result was actually contained in Ajtai's earlier paper [Ajt99a, Ajt98] which developed the main tools for analyzing the branching programs. The earlier paper gave similar lower bounds for two non-boolean problems whose input is a list of n binary strings, each of length $b = O(\log n)$ bits long:(1) determine whether the list contains a pair of strings within hamming distance δb for some fixed δ, and (2) determine whether the strings are all distinct.

The basic approach of all of these papers was to show that for any function computed by a "small" branching program, there must be a "large" *embedded rectangle* of inputs accepted by the function, and then to demonstrate that some particular functions don't have large embedded rectangles. (We will define embedded rectangle in section 2.1; for now it suffices for the reader to know that it is a highly structured subset of D^n.) This was done for syntactic read-k branching programs in [BRS93] and for general branching programs in [BST98]. The hamming distance result of [Ajt99a] can be deduced from the techniques of [BST98]. The lower bounds on embedded rectangle size proved in [BST98] are not strong enough to obtain the element distinctness and boolean function lower bounds. Ajtai obtained these bounds by proving an amazing combinatorial lemma that gave much stronger lower bound on embedded rectangle size. This directly gave his tradeoff results for element distinctness and was the basis for the subsequent Boolean branching program lower bound.

1.1 Our results

In this paper, we extend Ajtai's approach for deterministic branching programs in order to obtain the first time-space tradeoff results for (two-sided error) randomized branching programs, and also for deterministic branching programs that are allowed to err on a small fraction of inputs. Previously, there were no known time-space tradeoffs even in the uniform setting for these modes of computation. Our results apply to randomized RAM algorithms as well.

We also obtain substantial quantitative improvement over the previous results. More specifically, we show that, for the functions considered by Ajtai, any branching program of subexponential size must have depth at least $\Omega(n\sqrt{\frac{\log n}{\log\log n}})$. Ajtai does not explicitly give the functional form of his depth bounds, but analyzing his argument gives at most an $\Omega(n\frac{\log\log n}{\log\log\log n})$ bound.

Finally, while our argument is heavily based on Ajtai's, our version is considerably simpler.

One of the key aspects of both our extension and our simplification is to apply the basic approach developed in [BST98] of breaking up branching programs into collections of decision trees called decision forests and then analyzing the resulting decision forests. This has the effect of applying the space restriction only once, early in the argument, rather than delaying the application of the space restriction until the end of the argument which complicates the analysis without fundamentally changing its ideas.

Our extension of Ajtai's lemma shows that for a small deterministic branching program not only is there a large embedded rectangle of accepted inputs, but there is a set of large embedded rectangles of accepted inputs that cover almost all such inputs without covering any one input too many times. From this we show that if the given branching program agrees with a given target function f on all but a small fraction of inputs then there is a large embedded rectangle almost all of whose inputs are ones of f. We obtain our lower bounds for random algorithms by strengthening Ajtai's arguments about element distinctness and the quadratic forms to show that, not only do the functions not accept any relatively large embedded rectangle, they reject a large fraction of inputs in any such rectangle.

2 Notation

2.1 Boolean functions

Throughout, X is a set of variable indices (usually $X = \{1, \ldots, n\}$). An *input* x is a point in D^X, the set of mappings from X to D, and we identify this set, in the usual way, with D^n. If $A \subseteq X$, a point $\sigma \in D^A$ is a *partial input* to X. Given partial inputs σ and π defined on disjoint subsets A, A' of X, we write $\sigma\pi$ for the partial inputs defined on $A \cup A'$ whose restriction on A, A' is σ and π respectively.

Given $x \in D^X$ and $A \subseteq X$ define $x_A \in D^A$ to be the partial input that is the projection of x on A. We also find it convenient to define the modification of an input $x \in D^X$ by a partial input $\rho \in D^A$. Let x^ρ be the input which agrees with x on $X - A$ and with ρ on A, i.e., if $x' = x^\rho$ then $x'_A = \rho$ and $x'_{X-A} = x_{X-A}$.

A *Boolean function* f over X is a function mapping D^n to $\{0, 1\}$. Given a partial input $\rho \in D^A$, the *restriction* of f by ρ, $f\lceil_\rho$ is the function on $X' = X \setminus A$ such that for $\sigma \in D^{X'}$, $f\lceil_\rho(\sigma) = f(\sigma\rho)$.

Recall that a subset R of a set $U \times V$ is a *(combinatorial) rectangle* if and only if it is of the form $S \times T$ with $S \subseteq U$ and $T \subseteq V$ and that a function f is a rectangle if and only if it is the characteristic function of some rectangle R. We can extend this notion by saying that a subset R of D^X is a A-rectangle if and only if there is a set A such that R is a rectangle in $D^A \times D^{X-A}$. This notion of rectangle has been used, for example, in the study of communication complexity in the "best-partition" model and in the study of read-once branching programs.

For disjoint sets A_1, A_2 in X we say that a subset R of D^X is an *embedded* (A_1, A_2)*-rectangle* (we usually omit the word "embedded" throughout this paper for simplicity) if and only if there exist and sets $Y_1 \subseteq D^{A_1}$ and $Y_2 \subseteq D^{A_2}$ and a partial input ρ on $X - A_1 - A_2$ such that $R = \{x \in D^n : x_{X-A_1-A_2} = \rho, x_{A_1} \in Y_1, x_{A_2} \in Y_2\}$. We call the partial input ρ the *stem* of R. Further, we say that R is an (m, α)*-rectangle* if $|A_1|$, $|A_2| \geq m$ and $|Y_1| \geq \alpha|D^{A_1}|$ and $|Y_2| \geq \alpha|D^{A_2}|$. If sets A_1 and A_2 have size precisely m then we call R an *exact* (m, α)*-rectangle.*

Note that every (A_1, A_2)-rectangle is also an A-rectangle for any set $A \supseteq A_1$ with $A \cap A_2 = \emptyset$ (or dually). Also, note that $R_{A_1 \cup A_2} = \{x_{A_1 \cup A_2} : x \in R\}$ is an A_1-rectangle in $D^{A_1 \cup A_2}$ that we can identify with $Y_1 \times Y_2$.

Lemma 1. *If R is an (m, α)-rectangle with associated sets A_1, A_2, Y_1, Y_2 and partial assignment ρ then there is an $R' \subseteq R$ with defining sets $A'_1, A'_2, Y'_1, , Y'_2$ and partial assignment ρ' such that R' is also an (m, α)-rectangle and $|A'_1| = |A'_2| = m$; i.e. R' is an exact (m, α)-rectangle.*

Proof. Let A'_1, A'_2 consist of an arbitrary set of m elements of A_1, A_2, respectively. Among the partial assignments τ_i to $A_i - A'_i$, choose the one which maximizes the size of $(Y_i)\lceil_{\tau_i} = \{\sigma \in D^{A'_i} : \sigma\tau_i \in Y_i\}$ and set $Y'_i = (Y_i)\lceil_{\tau_i}$ for $i = 1, 2$. Clearly, by averaging, $|Y'_i| \geq \alpha$ and setting $\rho' = \rho\tau_1\tau_2$ we can see that the rectangle R' defined by A'_1, A'_2, Y'_1, Y'_2 and ρ' is an exact (m, α)-rectangle. $\square$

2.2 Branching programs, decision trees, and decision forests

Since we are only interested in the computation of Boolean (single output) functions here, we present our definitions of branching programs only for this case. A *branching program* B on a D-valued variable set X is an acyclic directed graph with the following properties:

- There is a unique source node, denoted $start_B$.
- Every sink node v has a label $output(v)$, which is 0 or 1.
- Each non-sink node v is labeled by a variable index $i(v) \in X$
- There are exactly $|D|$ arcs out of each non-sink node, each with a different element $value(a)$ of D.

Intuitively, a branching program is executed on input x by starting at $start_B$, reading the variable $x_{i(start_B)}$ and following the unique arc labeled by $x_{i(start_B)}$. This process is continued until a sink is reached and the output of the computation is the output value of the sink.

We say that B *accepts* the input x if the sink reached on input x is labeled 1. We view B as a boolean function from D^n by defining $B(x) = 1$ if and only if B accepts x.

Two measures associated with B are *size* which equals the number of nodes, and *length* which is the length of the longest path.

A branching program of length d is *leveled* if the nodes can be partitioned into d sets $V_0, V_1, \ldots, V_d$ where V_0 is the source, V_d is the set of sink nodes and every arc out of V_i goes to V_{i+1}, for $0 \leq i < d$. It is well known[Pip79] that every branching program P of size s and length d, can be converted into a leveled branching program P' of length d that has at most s nodes in each of its levels and computes the same function as P (and is deterministic if P is).

A *decision tree* is a branching program B whose underlying graph is a tree rooted at $start_B$. In particular, a decision tree is leveled. Every function on n variables is computable by a deterministic decision tree of length n. Following common practice, the length of a decision tree is referred to as its *height*.

A *decision program* or *decision forest* is an $\bigwedge$ of decision trees. More precisely, an (r, ϵ)-decision forest P over D is a collection $T_1, \ldots, T_r$ of decision trees on D^n such that each tree has height at most $\lceil \epsilon n \rceil$. The function computed by P is $\wedge_{i=1}^{r} T_i$.

3 Main Decomposition Theorems

The main approach taken in [BST98, Ajt99a, Ajt99b] for proving time-space tradeoff lower bounds is to show that if f can be computed by a branching program then f must evaluate to 1 on some (m, α)-rectangle where m and α are large as functions of the time T and space S used by the branching program. Roughly speaking, large m means that $m = \beta n$ where β is a function of T/n, and large α means $\alpha > |D|^{-w(m)}$ where $w(m)$ is small enough compared to m.

The first step in showing the existence of these large (m, α)-rectangles is to view the branching program as divided into a certain number, r, of layers, each of which is of length a small fraction of n, and only enforce the space limitation on the boundaries between these layers. This is similar to the approach that has been used in all of the time-space tradeoffs for multi-output problems, starting with [BFK+81, BC82]. In the multi-output arguments it was possible to argue that on each input there must be a single layer that produces a $1/r$ fraction of the outputs and argue about the computation in the layer beginning at each boundary node separately. In the case of decision problems, we must maintain the connection between the different layers. We express this by our decomposition of any branching program into a disjunction of decision forests (Lemma 2).

Paper	β	r	$\lambda(\beta)$	Program type	$\frac{\lvert R\cap f^{-1}(0)\rvert}{\lvert R\rvert}$	Applicability
[BST98]	2^{-k}	$\theta(k^2 2^k)$	$4(k+1)\beta$	non-determ.	0	$\log\log\lvert D\rvert = \Omega(k), k = O(\frac{\log n}{\log\log n})$
[Ajt99a]	$2^{-k^{\theta(k)}}$	$2^{k^{\theta(k)}}$	$\beta^{1+1/(50k)}$	determ.	0	all D, $k = O(\frac{\log\log n}{\log\log\log n})$
This paper	$k^{-\theta(k^2)}$	$k^{\theta(k^2)}$	$\beta^{1+1/(8k^2)}$	determ.	0	all D, $k = O(\sqrt{\frac{\log n}{\log\log n}})$
This paper	$k^{-\theta(k^2)}$	$k^{\theta(k^2)}$	$\beta^{1+1/(8k^2)}$	random 2-sided err. ϵ	$O(k^2\epsilon)$	all D, $k = O(\sqrt{\frac{\log n}{\log\log n}})$

Figure 1: $(m = \beta n, \alpha)$-rectangle R for branching programs with time $T = kn$ and space S that accept a δ fraction of D^n. Here $\alpha = \delta 2^{-\lambda(\beta)n - Sr}$.

There are two main differences between our results and previous results. First of all, we obtain substantially better values for m and α as functions of the time and space of the branching program. Secondly, we show that not only is there one large (m, α)-rectangle in $f^{-1}(1)$ but there is a small number of partial partitions of $f^{-1}(1)$ into large (m, α)-rectangles that together cover *almost all* the inputs on which f outputs 1. This latter aspect permits us to prove lower bounds for randomized and distributional as well as deterministic branching program complexity.

We summarize the relationships between the different results in Figure 1. In each case, we assume that we begin with a D-way branching program that accepts a δ fraction of all inputs in D^n. Each result shows that a branching program for some function f running in time $T = kn$ and space S admits a large $(\beta n, \alpha)$-rectangle, for some $\beta = \beta(k)$ and $\alpha = 2^{-\lambda(\beta)n - Sr}$ where λ is a nonnegative function of β and r is a function of k. In the first three lines of the table, the guaranteed rectangle consists entirely of points in $f^{-1}(1)$. In the last case, where the branching program is randomized with 2-sided error ϵ, the guaranteed rectangle may contain a small fraction of 0's. The last column gives the size of D and range of k for which the rectangle bounds can be used to get linear space bounds for explicit functions. The restriction on D in the first line comes from the fact that to get a nontrivial tradeoffs we need that $2^{\lambda(\beta)n}$ is "sufficiently small" compared to $D^{\beta n}$. The upper bound on k in each row is because the lower bound on space is roughly $\frac{\beta}{r}n$.

In the section 3.1 we describe the decomposition into decision forests and in section 3.2 we show how to find the appropriate partitions into (m, α)-rectangles within each decision forest. We derive the key theorems concerning branching programs in section 3.3.

3.1 Decomposition into Decision Forests

The following lemma is proved in [BST98]. We include its proof here for completeness.

Lemma 2. *Let $k \in \mathbf{R}$ and $n, s \in \mathbf{N}$. Let f be a boolean function on D^n for some finite set D. If f can be computed by a deterministic branching program of length kn and size s then for any integer $r \geq k$, f can be expressed as:*

$$f = \bigvee_{i=1}^{u} P_i,$$

where $u \leq s^{r-1}$, each P_i is a $(r, \frac{k}{r})$-decision forest, and the sets $P_i^{-1}(1)$ are pairwise disjoint sets of inputs.

Proof. Let B be any branching program of size s computing f of length $d \leq kn$. Let B' be an equivalent leveled program of length d with at most s nodes per level. For distinct nodes v and w of the branching program, let $f_{v,w}$ denote the function on D^n which is 1 on input σ if, starting from v, there is a path consistent with σ that leads to w. It is easy to see that if v is at level i and w is level $j > i$, then $f_{v,w}$ can be computed by a decision tree of height $j - i$. For $1 \leq i \leq r-1$, define $l_i = \lceil \frac{id}{r} \rceil$. Note that $l_1 < \cdots < l_{r-1} < d$ divides the interval $[0, d]$ into r intervals each of size at most $\lceil \frac{d}{r} \rceil \leq \lceil \frac{kn}{r} \rceil$. An input is accepted by P if and only there is a sequence of nodes $v_0, v_1, v_2, \ldots, v_{r-1}, v_r$, where v_0 is the start node, v_r is the accepting node and for $i \in [r-1]$, v_i is at level l_i, such that $f_{v_{i-1},v_i}(\sigma) = 1$ for each $i \in [r]$. Therefore

$$f = \bigvee_{v_1,\ldots v_{r-1}} \bigwedge_{i=0}^{r-1} f_{v_i,v_{i+1}}.$$

There are at most s^{r-1} terms in the $\bigvee$, and each term is a $(r, \frac{k}{r})$ decision forest.

Finally, each input follows a unique path, and so is accepted by at most one of the decision forests.

□

3.2 Finding (m, α)-rectangles in decision forests

Let $P = (T_1, \ldots, T_r)$ be a D-way $(r, k/r)$-decision program over X. For $x \in D^X$ and $i \in X$, let $\texttt{trees}_P(i, x) = \{j \in [1, ..., r] \ : \ i \text{ is read in } T_j \text{ on input } x\}$ and $\texttt{acc}_P(x, \ell) = \{i \ : \ |\texttt{trees}_P(i, x)| = \ell\}$. If T is a decision tree, let $\texttt{read}(x, T)$ to be the set of elements i read in T on input x. If F is a set of decision trees, let $\texttt{read}(x, F) = \bigcup_{T \in F} \texttt{read}(x, T)$.

For a subset $F \subseteq P = (T_1, \ldots, T_r)$, let $\texttt{core}_P(x, F)$ be the set of elements of X that are read only in trees in F on input x. Let $\texttt{stem}_P(x, F)$ be the partial assignment that is the restriction of x to $X - \texttt{core}_P(x, F)$. If $F_1, F_2 \subseteq P$ are disjoint then let $\texttt{stem}_P(x, F_1, F_2)$ be the partial assignment that is the restriction of x to $X - \texttt{core}_P(x, F_1) - \texttt{core}_P(x, F_2)$.

Call $x \in D^X$ *P-visible* if and only if for all $i \in X$, $\texttt{trees}_P(i, x) \neq \emptyset$. Note that if x is P-visible, $\texttt{core}_P(x, F) = X - \texttt{read}(x, P - F)$. Call a decision forest P over D^X *inquisitive* if every $x \in D^X$ is P-visible. Note that any $(r, k/r)$-decision forest may be made inquisitive by adding at most r/k decision trees, each of which has all inputs follow the same path with leaf label 1, corresponding to queries that read the inputs but ignore them. All results of this section pertain to inquisitive decision forest. This assumption yields the nice characterization of $\texttt{core}_P(x, F)$ above, but it is purely a matter of convenience and could be avoided.

The goal of the argument is to find partitions of inputs into large (m, α)-rectangles on which the decision forest accepts. Each (m, α)-rectangle will be an (A_1, A_2)-rectangle where A_1 and A_2 consist of variables that are read exclusively by one of two small sets of decision trees in the decision forest.

Let F_1 and F_2 be disjoint sets of decision trees in decision forest P. Let $P^{-1}(1) \subseteq D^X$ be the set of inputs accepted by P. We say that inputs $x, y \in P^{-1}(1)$ are (F_1, F_2)-*equivalent* if and only if

- $\texttt{core}_P(x, F_1) = \texttt{core}_P(y, F_1)$,
- $\texttt{core}_P(x, F_2) = \texttt{core}_P(y, F_2)$, and

- $\texttt{stem}_P(x, F_1, F_2) = \texttt{stem}_P(y, F_1, F_2)$.

Let $\mathcal{R}_P(F_1, F_2)$ be the set of (F_1, F_2)-equivalence classes. Clearly, these partition $P^{-1}(1)$.

Lemma 3. *Let P be an inquisitive decision forest and let $F_1, F_2 \subset P$ be disjoint sets of decision trees in P. Let R be an (F_1, F_2)-equivalence class in $\mathcal{R}_P(F_1, F_2)$ and let $x \in R$. Then R is a $(\texttt{core}_P(x, F_1), \texttt{core}_P(x, F_2))$-rectangle with stem $\texttt{stem}_P(x, F_1, F_2)$.*

Proof. Let $A_1 = \texttt{core}_P(x, F_1)$ and $A_2 = \texttt{core}_P(x, F_2)$. By definition, A_1 and A_2 are disjoint. Let $A_0 = X - A_1 - A_2$ and $\rho = \texttt{stem}_P(x, F_1, F_2)$. Let $Y_1 = \{y_{A_1} \ : \ y \in R\}$ and $Y_2 = \{y_{A_2} \ : \ y \in R\}$. Consider the (A_1, A_2)-rectangle Q defined by A_1, A_2, Y_1, Y_2 and ρ. Clearly, $R \subseteq Q$, so it suffices to show that $Q \subseteq R$.

Let $z \in Q$. By definition of Q, there must exist some $y^1 \in R$ such that $z_{A_1} = y^1_{A_1}$ and $y^2 \in R$ such that $z_{A_2} = y^2_{A_2}$. Furthermore $z_{A_0} = y^1_{A_0} = y^2_{A_0} = \rho$. Since the trees of $P - F_2$ do not depend on the variables in $A_2 = X - A_0 - A_1$, every tree $T \in P - F_2$ behaves the same on input z as it does on input y^1. Therefore $\texttt{read}(z, P - F_2) = \texttt{read}(y^1, P - F_2)$, and thus $\texttt{core}_P(z, F_2) = \texttt{core}_P(y^1, F_2) = A_2$ since P is inquisitive, and every tree in $P - F_2$ accepts z since $R \subseteq P^{-1}(1)$. Similarly, every tree $T \in P - F_1$ behaves the same on input z as it does on input y^2, $\texttt{core}_P(z, F_1) = \texttt{core}_P(y^2, F_1) = A_1$, and every tree in $P - F_1$ accepts z. Thus $z \in P^{-1}(1)$, $\texttt{core}_P(z, F_1) = A_1$, $\texttt{core}_P(z, F_2) = A_2$, and $\texttt{stem}_P(z, F_1, F_2) = \rho$ and thus $z \in R$ as required. □

Let $R_x(F_1, F_2)$ be the equivalence class in $\mathcal{R}_P(F_1, F_2)$ containing x. and $Y_j(x, F_1, F_2) = \{y_{\texttt{core}_P(x, F_j)} \ : \ y \in R_x(F_1, F_2)\}$ for $j = 1, 2$. By Lemma 3, $R_x(F_1, F_2)$ is an embedded rectangle defined by sets $(\texttt{core}_P(x, F_1), \texttt{core}_P(x, F_2), Y_1(x, F_1, F_2), Y_2(x, F_1, F_2))$ and partial assignment $\texttt{stem}_P(x, F_1, F_2)$. Defining

$$\begin{aligned} m_x(F_1, F_2) &= \min(|\texttt{core}_P(x, F_1)|, |\texttt{core}_P(x, F_2)|) \text{ and} \\ \alpha_x(F_1, F_2) &= \min(|Y_1(x, F_1, F_2)/|D^{\texttt{core}_P(x, F_1)}|, |Y_2(x, F_1, F_2)|/|D^{\texttt{core}_P(x, F_2)}|), \end{aligned}$$

we see that $R_x(F_1, F_2)$ is an $(m_x(F_1, F_2), \alpha_x(F_1, F_2))$-rectangle.

However we have no guarantee that either $m_x(F_1, F_2)$ or $\alpha_x(F_1, F_2)$ is large. In fact, we will not be able to show that any one fixed pair (F_1, F_2) will guarantee that $m_x(F_1, F_2)$ and $\alpha_x(F_1, F_2)$ are large. Instead, we will show that almost all accepted inputs x can be associated with one of a very small number of pairs (F_1, F_2) for which both of these are large.

Given an accepted input x, to make $m_x(F_1, F_2)$ and $\alpha_x(F_1, F_2)$ large, we would like to find a pair (F_1, F_2) so that both $\texttt{core}_P(x, F_1)$ and $\texttt{core}_P(x, F_2)$ are large and both $Y_1(x, F_1, F_2)$ and $Y_2(x, F_1, F_2)$ are large.

Let us first take a closer look at the $Y_j(x, F_1, F_2)$ using the property that $R_x(F_1, F_2)$ is an embedded rectangle. Clearly

$$\begin{aligned} Y_j(x, F_1, F_2) &= \{\sigma \in D^{\texttt{core}_P(x, F_j)} \ : \ x^\sigma \in R_x(F_1, F_2)\} \\ &= \{\sigma \in D^{\texttt{core}_P(x, F_j)} \ : \ \texttt{core}_P(x^\sigma, F_1) = \texttt{core}_P(x, F_1), \\ &\qquad \texttt{core}_P(x^\sigma, F_2) = \texttt{core}_P(x, F_2), \text{ and } x^\sigma \in P^{-1}(1)\} \end{aligned}$$

The following lemma shows that we can simplify this definition still further by removing the condition $\texttt{core}_P(x^\sigma, F_j) = \texttt{core}_P(x, F_j)$.

Lemma 4. *Let P be a decision forest and let $F \subset P$. If $x, y \in D^X$ are P-visible inputs such that x and y agree on all elements in $X - \mathtt{core}_P(x, F)$ then $\mathtt{core}_P(x, F) = \mathtt{core}_P(y, F)$ and $\mathtt{stem}_P(x, F) = \mathtt{stem}_P(y, F)$.*

Proof. Since x and y agree on all elements of $X - \mathtt{core}_P(x, F)$, the computations of all trees of P outside of F are the same on x and y. Thus, in particular, $\mathtt{read}(x, P - F) = \mathtt{read}(y, P - F)$ and so $\mathtt{core}_P(y, F) = X - \mathtt{read}(y, P - F) = X - \mathtt{read}(x, P - F) = \mathtt{core}_P(x, F)$. It then follows that $\mathtt{stem}_P(x, F) = \mathtt{stem}_P(y, F)$ since they agree on $X - \mathtt{core}_P(x, F) = X - \mathtt{core}_P(y, F)$. $\square$

Thus $Y_1(x, F_1, F_2) = \{\sigma \in D^{\mathtt{core}_P(x,F_1)} : \mathtt{core}_P(x^\sigma, F_2) = \mathtt{core}_P(x, F_2) \text{ and } x^\sigma \in P^{-1}(1)\}$ and $Y_2(x, F_1, F_2) = \{\sigma \in D^{\mathtt{core}_P(x,F_2)} : \mathtt{core}_P(x^\sigma, F_1) = \mathtt{core}_P(x, F_1) \text{ and } x^\sigma \in P^{-1}(1)\}$. In particular, having a large $Y_1(x, F_1, F_2)$ requires that there are many ways of varying the values of variables in $\mathtt{core}_P(x, F_1)$ to get a new assignment y with $\mathtt{core}_P(y, F_2) = \mathtt{core}_P(x, F_2)$, and having a large $Y_2(x, F_1, F_2)$ requires its dual. (Having large $Y_j(x, P, F_1, F_2)$ requires more than this; it also requires that each modified input y is also accepted by P but we will see that this will follow frequently enough if there are sufficiently many accepted inputs overall.)

Thus for an input x we would like to find (F_1, F_2) such that

(i) each $\mathtt{core}_P(x, F_j)$ is large for $j = 1, 2$, and

(ii) many ways of varying x on $\mathtt{core}_P(x, F_j)$ generate inputs y that have $\mathtt{core}_P(y, F_{3-j}) = \mathtt{core}_P(x, F_{3-j})$ for $j = 1, 2$.

We will see that if the pair (F_1, F_2) is chosen randomly of roughly the right size based on some simple characteristics of x, then it is very likely that it will have both property (i) and something close to property (ii). More precisely, we will show that almost all changes to x on $\mathtt{core}_P(x, F_j)$ will not affect $\mathtt{core}_P(\cdot, F_{3-j})$ too much. This will allow us to group the inputs y that differ from x only on $\mathtt{core}_P(x, F_j)$ based on their value of $\mathtt{core}_P(\cdot, F_{3-j})$ and, since the variation from $\mathtt{core}_P(x, F_{3-j})$ is small, there aren't too many groups. These inputs y will all have the same value of $\mathtt{core}_P(\cdot, F_{3-j})$. Applying this grouping for both $j = 1$ and 2 will give the large sets of inputs y with the desired properties.

More precisely, we choose (F_1, F_2) by independently assigning each decision tree $T \in P$ according the following probability distribution $\mathcal{F}(q)$ for some $q \in (0, \frac{1}{2}]$:

$$T \in \begin{cases} F_1 & \text{with probability } q \\ F_2 & \text{with probability } q \\ P - F_1 - F_2 & \text{with probability } 1 - 2q. \end{cases}$$

The probability that an element $i \in X$ is in $\mathtt{core}_P(x, F_j)$ depends only on q and the number of different decision trees in which i is read on input x. If i is read in precisely ℓ decision trees on input x, the probability that $i \in \mathtt{core}_P(x, F_j)$ is precisely q^ℓ. Therefore, because P is not too large, $\mathtt{acc}_P(x, \ell)$ will be large for a relatively small ℓ, and thus the expected size of $\mathtt{core}_P(x, F_j)$ will be reasonably large, at least $q^\ell |\mathtt{acc}_P(x, \ell)|$. The following lemma makes this more precise and shows that the sizes of the cores do not vary too much from their expected values.

Lemma 5. *Let P be an inquisitive $(r, k/r)$-decision forest and $\epsilon = k/r$. Let x be any input and $j \in \{1, 2\}$. Let (F_1, F_2) be chosen according to probability distribution $\mathcal{F}(q)$. Then for $j = 1, 2$:*

(a) $\mathrm{E}[|\mathtt{core}_P(x, F_j)|] \stackrel{def}{=} \mu_x \geq q^k n$.

(b) $\mathrm{Var}(|\mathtt{core}_P(x, F_j)|) \leq \epsilon n k \mu_x$

(c) $\Pr\left[||\mathtt{core}_P(x, F_j)| - \mu_x| \geq \frac{1}{2}\mu_x\right] \leq 4\epsilon k/q^k$

Proof. It suffices to fix $j = 1$.

Let $t(i) = |\mathtt{trees}(i)|$ be the number of trees in which variable i is accessed. We have $\mathrm{E}[|\mathtt{core}_P(x, F_1)|] = \sum_{i \in X} q^{t(i)}$. Since P makes a total of at most kn accesses to variables on input x, $\frac{1}{n}\sum_{i \in X} t(i) \leq k$. By the arithmetic-geometric mean inequality, $\mathrm{E}[|\mathtt{core}_P(x, F_1)|] = \sum_i q^{t(i)} \geq nq^{\frac{1}{n}\sum_i \tau(i)} \geq q^k n$.

Let $M(i)$ be the event that $i \in \mathtt{core}_P(x, F_1)$. For $1 \leq i, i' \leq n$, we say $i \sim i'$ if x_i and $x_{i'}$ are both accessed on some decision tree. Now

$$\mathrm{Var}(|\mathtt{core}_P(x, F_1)|) = \sum_{i,i'} (\Pr[M(i) \wedge M(i')] - \Pr[M(i)] \cdot \Pr[M(i')]).$$

If $\neg(i \sim i')$ then the events $M(i)$ and $M(i')$ are independent and the corresponding term in the sum is 0. If $i \sim i'$ then we upper bound $\Pr[M(i) \wedge M(i')] - \Pr[M(i)] \cdot \Pr[M(i')]$ crudely by $\Pr[M(i)]$. Since on input x, each tree reads at most ϵn variables, for each i the number of i' such that $i \sim i'$ is at most $t(i)\epsilon n$. Thus,

$$\mathrm{Var}(|\mathtt{core}_P(x, F_1)|) \leq \epsilon n \sum_{i=1}^{n} t(i) q^{t(i)} \leq \epsilon \sum_{i=1}^{n} t(i) \sum_{i=1}^{n} q^{t(i)} \leq \epsilon n k \mathrm{E}[|\mathtt{core}_P(x, F_1)|] = \epsilon n k \mu_x.$$

For the last part of the Lemma we use Chebyshev's inequality: for any random variable Z, $\Pr[|Z - \mathrm{E}[Z]| \geq \lambda] \leq \mathrm{Var}[Z]/\lambda^2$.

$$\begin{aligned} \Pr\left[||\mathtt{core}_P(x, F_1)| - \mu_x|]| \geq \frac{1}{2}\mu_x\right] &\leq 4\mathrm{Var}[|\mathtt{core}_P(x, F_1)|]/\mu_x^2 \\ &\leq 4\epsilon n k/\mu_x \\ &\leq 4\epsilon k/q^k. \end{aligned}$$

□

In order to show that changing the assignment to the variables in $\mathtt{core}_P(x, F_j)$ does not change the core on F_{3-j} too much, we need to bound the symmetric difference of $\mathtt{core}_P(x, F_{3-j})$ and $\mathtt{core}_P(y, F_{3-j})$ for inputs x, y such that $\mathtt{stem}_P(x, F_j) = \mathtt{stem}_P(y, F_j)$. To do this, it will be useful to consider following subsets of X. For $j \in \{1, 2\}$, let

$$\begin{aligned} B_j(x, \ell) &= \mathtt{core}_P(x, F_j) - \mathtt{acc}_P(x, \ell), \\ B'_j(x, \ell) &= \{i \in X \ : \ \text{on input } x,\ i \text{ is read in exactly } \ell \text{ trees in } F_j, \\ &\qquad \text{in at least one tree in } F_{3-j}, \\ &\qquad \text{and in no trees in } P - F_1 - F_2. \ \}. \end{aligned}$$

The following lemma gives a combinatorial description of the variables in the symmetric difference of cores.

Lemma 6. *Let P be an inquisitive decision forest. Let ℓ be some positive integer. For $j \in \{1,2\}$ and inputs $x, y \in D^X$ such that $\mathtt{stem}_P(x, F_{3-j}) = \mathtt{stem}_P(y, F_{3-j})$ we have*

$$\mathtt{core}_P(x, F_j) \Delta \mathtt{core}_P(y, F_j) \subseteq B_j(x,\ell) \cup B_j(y,\ell) \cup B'_j(x,\ell) \cup B'_j(y,\ell)$$

Proof. Without loss of generality, let $j = 1$.

Let $i \in \mathtt{core}_P(x, F_1) - \mathtt{core}_P(y, F_1)$. If $i \notin \mathtt{acc}_P(x,\ell)$, then $i \in B_1(x,\ell)$. If $i \in \mathtt{acc}_P(x,\ell)$, then since x and y agree outside of F_2 and i is in $\mathtt{core}_P(x, F_1)$ but not in $\mathtt{core}_P(y, F_1)$, on input y, i must be read by exactly ℓ trees in F_1 and by at least one tree in F_2 and by no trees elsewhere, i.e., i is in $B'_1(y,\ell)$. Therefore $i \in B_1(x,\ell) \cup B'_1(y,\ell)$.

By symmetry of x and y, we have

$$\mathtt{core}_P(x, F_1) \Delta \mathtt{core}_P(y, F_1) \subseteq B_1(x,\ell) \cup B_1(y,\ell) \cup B'_1(x,\ell) \cup B'_1(y,\ell).$$

□

The following lemma shows that we can choose ℓ, q from a small number of possible values such that the expectations of the sizes of $B_j(x,\ell)$ and $B'_j(x,\ell)$ $(j = 1,2)$ are substantially smaller than the expectation of the size of the core. The major improvement of our bounds over those implicit in [Ajt99a, Ajt99b] is due to our more precise description of these quantities and our sharper calculation of the expected sizes of these four sets. Roughly speaking, in each case, the structure of the analysis in [Ajt99a] only makes use of the randomness of one of the forests in the pair (F_1, F_2) while holding the other fixed. We restructure the analysis so that we can take advantage of the fact that both forests are random.

Lemma 7. *Let $P = (T_1, \ldots, T_r)$ be an inquisitive $(r, k/r)$-decision forest. Let $q_1 \le 1/(4k)$. For every P-visible $x \in D^X$, there is a pair $(\ell_P(x), q_P(x)) = (\ell, q)$ with $1 \le \ell \le 2k$ and $q = q_b = q_1^b$ for some integer b, $1 < b \le 4k$, such that for (F_1, F_2) chosen according to distribution $\mathcal{F}(q)$ and $j \in \{1,2\}$,*

(a) $\mathrm{E}[B_j(x,\ell)] \le 4q_1 \cdot \mathrm{E}[|\mathtt{core}_P(x, F_j)|]$.

(b) $\mathrm{E}[B'_j(x,\ell)] \le 4kq_1 \cdot \mathrm{E}[|\mathtt{core}_P(x, F_j)|]$.

Proof. Let $\nu_h = |\mathtt{acc}_P(x,h)|$ for $h = 1, \ldots, r$. It is easy to see that

$$\mathrm{E}[|\mathtt{core}_P(x, F_1)|] = \mathrm{E}[|\mathtt{core}_P(x, F_2)|] = \sum_{h=1}^{r} \nu_h q^h.$$

We will choose an ℓ and a q such that the term $\nu_\ell q^\ell$ dominates all the other terms in the sum.

Let $q_a = q_1^a$. For each $1 \le a \le 4k+1$, there is an $h(a)$ such that the largest term in the sum comes from inputs accessed precisely $h(a)$ times, i.e., $\nu_{h(a)} q_a^{h(a)} = \max_h \nu_h q_a^h$. We have the following claims.

(1) $h(1) < 2k$.

(2) $h(a)$ is decreasing with respect to a.

(3) $h(a) \geq 1$.

By definition of k, P makes at most kn queries on input x. Therefore at least $n/2$ elements of X are queried in fewer than $2k$ trees on input x. Therefore, there is some $\ell_1 < 2k$ such that more than $n/(4k)$ elements of X are queried in precisely ℓ_1 trees on input x. This means $\nu_{\ell_1} \geq n/(4k)$ for some $\ell_1 < 2k$. Thus for $h \geq 2k$,

$$\nu_h q_1^h \leq n q_1^h \leq n q_1^{\ell_1+1} < \nu_{\ell_1} q_1^{\ell_1}$$

since $q_1 \leq 1/(4k)$. Therefore, the first claim is true. The second claim is true simply due to the fact that if $h < h'$ then $\nu_h q_a^h \geq \nu_{h'} q_a^{h'}$ implies $\nu_h q_{a+1}^h > \nu_{h'} q_{a+1}^{h'}$. The third claim follows immediately from the definition.

Thus by the pigeonhole principle, there must exist some $2 \leq b \leq 4k$ such that $h(b-1) = h(b) = h(b+1)$. Now we choose $\ell = h(b)$ and $q = q_b = q_1^b$. We have for $h \neq \ell$,

$$\nu_h q_a^h \leq \nu_\ell q_a^\ell \qquad \text{for } a = b-1, b, b+1.$$

This implies that for $h \neq \ell$,

$$\nu_h q_b^h \leq \nu_\ell q_b^\ell \cdot q_1^{|h-\ell|}.$$

Let $q = q_b$. By definition,

$$\begin{aligned}
\mathrm{E}[B_j(x,\ell)] &= \sum_{h \neq \ell} \nu_h q^h \\
&\leq \sum_{h \neq \ell} \nu_\ell q^\ell \cdot q_1^{|\ell-h|} \\
&\leq 2\nu_\ell q^\ell \sum_{p=1}^{\infty} q_1^p \\
&\leq 4\nu_\ell q^\ell \cdot q_1 \\
&\leq 4q_1 \cdot \mathrm{E}[|\mathtt{core}_P(x, F_j)|].
\end{aligned}$$

Now any $i \in B_j'(x,\ell)$ must be in $\mathtt{acc}_P(x,h)$ for some $h \geq \ell+1$. If we fix such an $i \in \mathtt{acc}_P(x,h)$ there are precisely $\binom{h}{\ell}$ ways to allocate ℓ of these trees to F_j. Furthermore, the probability that a

fixed set of ℓ trees are in F_j and a disjoint fixed set of $h-\ell$ trees are in F_{3-j} is precisely q^h.

$$\begin{aligned}
\mathrm{E}[B'_j(x,\ell)] &= \sum_{h=\ell+1}^{r} \nu_h q^h \binom{h}{\ell} \\
&= \sum_{h=\ell+1}^{r} \nu_h q^h \cdot \frac{(\ell+1)}{1}\frac{(\ell+2)}{2}\cdots\frac{(\ell+h-\ell)}{(h-\ell)} \\
&\le \sum_{h=\ell+1}^{r} \nu_h q^h (\ell+1)^{h-\ell} \\
&\le \sum_{h=\ell+1}^{r} \nu_h q^h (2k)^{h-\ell} \\
&\le \sum_{h=\ell+1}^{r} \nu_\ell q^\ell (2k)^{h-\ell} q_1^{h-\ell} \\
&\le \nu_\ell q^\ell \sum_{p=1}^{\infty} (2kq_1)^p \\
&\le 4kq_1 \cdot \mathrm{E}[|\mathtt{core}_P(x,F_j)|].
\end{aligned}$$

since $q_1 \le 1/(4k)$. □

We classify the inputs by two parameters ℓ and q in Lemma 7. Since there are fewer than $2k$ different ℓ and $4k$ different q, there are fewer than $8k^2$ classes. Lemma 6 and Lemma 7 show that the expectation of $|\mathtt{core}_P(x,F_j)\Delta\mathtt{core}_P(y,F_j)|$ ($j \in \{1,2\}$) is substantially smaller than that of $|\mathtt{core}_P(x,F_j)|$. In the following lemma, we use Markov's inequality to prove that this is actually the case with high probability.

Lemma 8. *Let $k \ge 2$ be an integer, $q_1 \le 1/(4k)$, and $\gamma \in (0, \frac{1}{6})$ be given. Let $r \ge 4k^2/(\gamma q_1^{4k^2})$. Let $P = (T_1, \ldots, T_r)$ be an inquisitive $(r, k/r)$-decision forest with $\epsilon = k/r$. Let $1 \le \ell < 2k$, $1 < b \le 4k$, $q = q_1^b$ and $I \subseteq D^X$ be a set of inputs z such that $\ell_P(z) = \ell$, and $q_P(z) = q$. There exists a pair of subforests (F_1, F_2) and a subset I' of I such that $|I'| \ge (1-6\gamma)|I|$ and for $j \in \{1,2\}$, any $x \in I'$ satisfies*

$$|\mathtt{core}_P(x,F_j)| \ge \frac{1}{2}q^k n \qquad \textit{and} \qquad |\mathtt{core}_P(x,F_j)| \le 3|\mathtt{core}_P(x,F_{3-j})|,$$

and any $x, y \in I'$ such that $\mathtt{stem}_P(x,F_{3-j}) = \mathtt{stem}_P(y,F_{3-j})$ satisfy

$$|\mathtt{core}_P(x,F_j)\Delta\mathtt{core}_P(y,F_j)| \le 40kq_1 \cdot \min(|\mathtt{core}_P(x,F_1)|, |\mathtt{core}_P(x,F_2)|)/\gamma.$$

Proof. Observe that for this value of r and $\epsilon = k/r$, we have $4\epsilon k/q^k \le \gamma$. For any fixed $z \in I$ and $j \in \{1,2\}$, Lemma 5, Lemma 7 and Markov's inequality imply that the each of the following statements fails with probability at most γ.

(i) $||\mathtt{core}_P(z,F_j)| - E[|\mathtt{core}_P(z,F_j)|]| \le \frac{1}{2}\mathrm{E}[|\mathtt{core}_P(z,F_j)|]$,

(ii) $|B_j(z,\ell)| \le 4q_1 \cdot E[|\mathtt{core}_P(z,F_i)|]/\gamma$.

(iii) $|B'_j(z,\ell)| \le 4kq_1 \cdot E[|\mathtt{core}_P(z,F_i)|]/\gamma$.

Therefore there is some fixed pair of disjoint subforests F_1, F_2 such that all of (i), (ii) and (iii) hold for at least a $1-6\gamma$ fraction of I for both $j=1$ and 2. Fix this F_1 and F_2 and let $I' \subset I$ be the set of those z for which these conditions hold.

Fix $j=1$. Due to (i) and the simple lower bound $\mathrm{E}[|\mathtt{core}_P(x,F_1)|] \geq q^k n$ from Lemma 5, $|\mathtt{core}_P(x,F_1)| \geq \frac{1}{2}q^k n$. Since $E[|\mathtt{core}_P(x,F_1)|] = E[|\mathtt{core}_P(x,F_2)|]$, (i) implies $|\mathtt{core}_P(x,F_1)| \leq 3|\mathtt{core}_P(x,F_2)|$.

Without loss of generality, we consider $x, y \in I'$ such that $\mathtt{stem}_P(x,F_2) = \mathtt{stem}_P(y,F_2)$. By Lemma 6 we have

$$\begin{aligned} |\mathtt{core}_P(x,F_1)\Delta\mathtt{core}_P(y,F_1)| &\leq |B_1(x,\ell)| + |B_1(y,\ell)| + |B_1'(x,\ell)| + |B_1'(y,\ell)| \\ &\leq 5kq_1 \cdot (\mathrm{E}[|\mathtt{core}_P(x,F_1)|] + \mathrm{E}[|\mathtt{core}_P(y,F_1)|])/\gamma. \end{aligned}$$

Since $\mathrm{E}[|\mathtt{core}_P(x,F_1)|] = \mathrm{E}[|\mathtt{core}_P(x,F_2)|]$ and $\mathrm{E}[|\mathtt{core}_P(y,F_1)|] = \mathrm{E}[|\mathtt{core}_P(y,F_2)|]$ and using the fact that $|\mathtt{core}_P(z,F_j)| \geq \frac{1}{2}\mathrm{E}[|\mathtt{core}_P(z,F_j)|]$ for $z \in I'$,

$$\begin{aligned} &\mathrm{E}[|\mathtt{core}_P(x,F_1)| + \mathrm{E}[|\mathtt{core}_P(y,F_1)|] \\ &\leq 2[\min(|\mathtt{core}_P(x,F_1)|, |\mathtt{core}_P(x,F_2)|) + \min(|\mathtt{core}_P(y,F_1)|, |\mathtt{core}_P(y,F_2))|] \\ &\leq 2\min(|\mathtt{core}_P(x,F_1)|, |\mathtt{core}_P(x,F_2))| + 2|\mathtt{core}_P(x,F_2)| \qquad \text{since } \mathtt{core}_P(x,F_2) = \mathtt{core}_P(y,F_2) \\ &\leq \min(4|\mathtt{core}_P(x,F_2)|, 8|\mathtt{core}_P(x,F_1)|) \qquad \text{since } |\mathtt{core}_P(x,F_2)| \leq 3|\mathtt{core}_P(x,F_1)|. \end{aligned}$$

Plugging this in, we obtain, as required,

$$|\mathtt{core}_P(x,F_1)\Delta\mathtt{core}_P(y,F_1)| \leq 40kq_1 \cdot \min(|\mathtt{core}_P(x,F_1)|, |\mathtt{core}_P(x,F_2)|)/\gamma.$$

□

We now are fairly close to getting our desired rectangles. Let $S(n,d)$ be the number of subsets of X of size at most d. Note that $S(n,d) \leq 2^{H_2(d/n)n}$ where H_2 is the binary entropy function.

Lemma 9. *Let $k \geq 2$ be an integer, $\gamma \in (0, \frac{1}{8})$ be given and $q_1 \leq \gamma/(120k)$. Let $r \geq 4k^2/(\gamma q_1^{4k^2})$. Let $P = (T_1, \ldots, T_r)$ be an inquisitive $(r, k/r)$-decision forest. Let $I \subseteq D^X$ be a set of inputs accepted by P. There is an $I' \subseteq I$ with $|I'| \geq (1-8\gamma)|I|$ that is covered by fewer than $8k^2$ families $\mathcal{R}_{\ell,b}$ for $1 \leq \ell < 2k$, $1 < b \leq 4k$ of disjoint (m,α)-rectangles, where each (m,α)-rectangle satisfies $m \geq q_1^{4k^2} n/2$ and $\alpha \geq \gamma|I||D|^{-n}/(8k^2 S(n, 120kq_1 m/\gamma))$.*

Proof. Let $\alpha_{\mathrm{low}}(m) = \gamma|I||D|^{-n}/(8k^2 S(n, 120kq_1 m/\gamma))$.

By Lemma 8, we can partition I into sets $I_{\ell,b} = \{x \in I \,:\, \ell_P(x) = \ell, q_P(x) = q_1^b\}$, for $1 \leq \ell < 2k$ and $1 < b \leq 4k$, and for each such pair (ℓ, b) there is a pair of disjoint subforests $F_1 = F_1^{\ell,b}$ and $F_2 = F_2^{\ell,b} \subseteq P$ and a subset $I^*_{\ell,b} \subseteq I_{\ell,b}$ with $|I^*_{\ell,b}| \geq (1-6\gamma)|I_{\ell,b}|$ satisfying

(i) Any $x \in I^*_{\ell,b}$ has $|\mathtt{core}_P(x,F_j)| \geq \frac{1}{2}q^k n \geq q_1^{4k^2} n/2$ and $|\mathtt{core}_P(x,F_j)| \leq 3|\mathtt{core}_P(x,F_{3-j})|$ for $j \in \{1,2\}$.

(ii) Any $x, y \in I^*_{(\ell,b)}$ such that $\mathtt{stem}_P(x,F_{3-j}) = \mathtt{stem}_P(y,F_{3-j})$ have $|\mathtt{core}_P(x,F_j)\Delta\mathtt{core}_P(y,F_j)| \leq 40kq_1 \cdot \min(|\mathtt{core}_P(x,F_1)|, |\mathtt{core}_P(x,F_2)|)/\gamma$ for $j \in \{1,2\}$.

Furthermore, for each (ℓ, b) pair, letting $F_1 = F_1^{\ell,b}$ and $F_2 = F_2^{\ell,b}$ define

$$\mathcal{R}_{\ell,b} = \{R \in \mathcal{R}_P(F_1, F_2) \ : \ R \text{ is an } (m, \alpha)\text{-rectangle with } m \geq q_1^{4k^2} n/2 \text{ and } \alpha \geq \alpha_{\text{low}}(m)\}.$$

We want to show that at most an 8γ fraction of the inputs in I is not covered by rectangles in $\mathcal{R}_{\ell,b}$ for some (ℓ, b). By the argument above, at most a 6γ fraction of the inputs in I is not contained in $I^* = \bigcup_{\ell,b} I^*_{\ell,b}$. It remains to show that at most $2\gamma|I|$ inputs are in I^* but are not covered by any rectangle in some $\mathcal{R}_{\ell,b}$.

Fix (ℓ, b) with $1 \leq \ell < 2k$ and $1 < b \leq 4k$ and let $F_1 = F_1^{\ell,b}$ and $F_2 = F_2^{\ell,b}$. Consider the following equivalence relations.

For $j \in \{1, 2\}$, for $x, y \in D^X$ define $x \sim_j^{\ell,b} y$ if and only if $\texttt{stem}_P(x, F_j) = \texttt{stem}_P(y, F_j)$. For input $x \in D^X$, let $[x]_1^{\ell,b}$ and $[x]_2^{\ell,b}$ be the respective equivalence classes in D^X containing x. Observe that by Lemma 4, for any $x \in D^x$, every $y \in D^X$ such that y agrees with x on $X - \texttt{core}_P(x, F_j)$ has $\texttt{stem}_P(y, F_j) = \texttt{stem}_P(x, F_j)$ and thus $y \in [x]_j^{\ell,b}$. Since these are the only y for which $\texttt{stem}_P(y, F_j) = \texttt{stem}_P(x, F_j)$, $|[x]_j^{\ell,b}| = |D^{\texttt{core}_P(x,F_j)}|$.

For $j \in \{1, 2\}$, for $x, y \in P^{-1}(1)$, let $x \sim_{j\langle 3-j\rangle}^{\ell,b} y$ if and only if $\texttt{stem}_P(x, F_j) = \texttt{stem}_P(y, F_j)$ and $\texttt{core}_P(x, F_{3-j}) = \texttt{core}_P(y, F_{3-j})$. For input $x \in P^{-1}(1)$, let $[x]_{1\langle 2\rangle}^{\ell,b}$ and $[x]_{2\langle 1\rangle}^{\ell,b}$ be the respective equivalence classes of inputs in $P^{-1}(1)$ containing x under these equivalence relations. Note that for $x, y \in P^{-1}(1)$, $x \sim_{j\langle 3-j\rangle}^{\ell,b} y$ implies that $x \sim_j^{\ell,b} y$.

Consider $x \in I^*_{\ell,b}$. Recall the $Y_j(x, F_1, F_2)$, $m_x(F_1, F_2)$, and $\alpha_x(F_1, F_2)$ associated with rectangle $R_x(F_1, F_2)$. Observe that $[x]_{j\langle 3-j\rangle}^{\ell,b} = \{x^\sigma \ : \ \sigma \in Y_j(x, F_1, F_2)\}$ and thus $|Y_j(x, F_1, F_2)| = |[x]_{j\langle 3-j\rangle}^{\ell,b}|$.

Now $R_x(F_1, F_2)$ is an $(m_x(F_1, F_2), \alpha_x(F_1, F_2))$-rectangle where

$$\begin{aligned} m_x(F_1, F_2) &= \min(|\texttt{core}_P(x, F_1)|, |\texttt{core}_P(x, F_2)|), \text{ and} \\ \alpha_x(F_1, F_2) &= \min(|Y_1(x, F_1, F_2)|/|D^{\texttt{core}_P(x,F_1)}|, |Y_2(x, F_1, F_2)|/|D^{\texttt{core}_P(x,F_2)}|) \\ &= \min([x]_{1\langle 2\rangle}^{\ell,b}|/|[x]_1^{\ell,b}|, [x]_{2\langle 1\rangle}^{\ell,b}|/|[x]_2^{\ell,b}|). \end{aligned}$$

By construction, all $x \in I^*_{\ell,b}$ satisfy $m_x = m_x(F_1, F_2) \geq q_1^{4k^2} n/2$ so we only need to count those $x \in I^*_{\ell,b}$ such that $\alpha_x = \alpha_x(F_1, F_2) \leq \alpha_{\text{low}}(m_x)$. Let $I'_{\ell,b}$ be those $x \in I^*_{\ell,b}$ such that $\alpha_x > \alpha_{\text{low}}(m_x)$. By construction, every element of $I'_{\ell,b}$ is covered by some rectangle in $\mathcal{R}_{\ell,b}$. We now bound $|I^*_{\ell,b} - I'_{\ell,b}|$.

If $\alpha_x \leq \alpha_{\text{low}}(m_x)$ then either $\alpha_x^1 = |[x]_{1\langle 2\rangle}^{\ell,b}|/|[x]_1^{\ell,b}| \leq \alpha_{\text{low}}(m_x)$ or $\alpha_x^2 = [x]_{2\langle 1\rangle}^{\ell,b}|/|[x]_2^{\ell,b}| \leq \alpha_{\text{low}}(m_x)$. We consider the two cases separately.

Let $B_1^{\ell,b} = \{z \in I^*_{\ell,b} \ : \ \alpha_z^1 \leq \alpha_{\text{low}}(m_z)\}$. Fix some $x \in I^*_{\ell,b}$. We bound $|B_1^{\ell,b} \cap [x]_1^{\ell,b}|$. Since $x \sim_{1\langle 2\rangle}^{\ell,b} y$ implies $x \sim_1^{\ell,b} y$, we obtain a partition of $P^{-1}(1) \cap [x]_1^{\ell,b}$ into $\sim_{1\langle 2\rangle}^{\ell,b}$-equivalence classes. Let C_x be a set of representative elements z for these classes, $[z]_{1\langle 2\rangle}^{\ell,b}$, where we choose each z to be in $I^*_{\ell,b}$, if possible. Note that for all $z \in C_x$, $[z]_1^{\ell,b} = [x]_1^{\ell,b}$. Let $C'_x = \{z \in C_x \ : \ z \in I^*_{\ell,b} \text{ and } \alpha_z^1 \leq \alpha_{\text{low}}(m_z)\}$. Note that

$$B_1^{\ell,b} \cap [x]_1^{\ell,b} \subseteq \bigcup_{z \in C'_x} [z]_{1\langle 2\rangle}^{\ell,b}.$$

By Lemma 8, if $z \in C'_x$, $|\texttt{core}_P(x, F_2) \Delta \texttt{core}_P(z, F_2)| \leq 40kq_1 m_x/\gamma$ and therefore $|C'_x| \leq S(n, 40kq_1 m_x/\gamma)$ since all $z \in C_x$ have distinct values of $\texttt{core}_P(z, F_2)$. Also for $z \in C'_x$,

$|\mathtt{core}_P(z,F_1)| \le 3|\mathtt{core}_P(z,F_2)|$ and $\mathtt{stem}_P(x,F_1) = \mathtt{stem}_P(z,F_1)$. Therefore $\mathtt{core}_P(x,F_1) = \mathtt{core}_P(z,F_1)$ and $m_x \le |\mathtt{core}_P(x,F_1)| = |\mathtt{core}_P(z,F_1)| \le 3m_z$. It follows that for any $z \in C'_x$, $|C'_x| \le S(n, 40kq_1m_x/\gamma) \le S(n, 120kq_1m_z/\gamma)$.

Let $z = z^* \in C'_x$ maximize α^1_z. Then

$$\begin{aligned}
|B_1^{\ell,b} \cap [x]_1^{\ell,b}| &\le |\bigcup_{z\in C'_x} [z]^{\ell,b}_{1\langle 2\rangle}| \\
&\le \sum_{z\in C'_x} |[z]^{\ell,b}_{1\langle 2\rangle}| \\
&= \sum_{z\in C'_x} \alpha^1_z \cdot |[z]^{\ell,b}_1| \\
&= \sum_{z\in C'_x} \alpha^1_z \cdot |[x]^{\ell,b}_1| \\
&\le |C'_x| \cdot \alpha^1_{z^*} \cdot |[x]^{\ell,b}_1| \\
&\le S(n, 120kq_1m_{z^*}/\gamma) \cdot \alpha_{\text{low}}(m_{z^*}) \cdot |[x]^{\ell,b}_1| \\
&\le \gamma|I||D|^{-n}|[x]^{\ell,b}_1|/(8k^2).
\end{aligned}$$

Summing over all equivalence classes $[x]^{\ell,b}_1$, we obtain that $|B_1^{\ell,b}| \le \gamma|I|/(8k^2)$ since these classes cover D^X.

By similar reasoning we obtain the same upper bound of $\gamma|I|/(8k^2)$ on the size of the set $B_2^{\ell,b} = \{z \in I^*_{\ell,b} \ : \ \alpha^2_z \le \alpha_{\text{low}}(m_z)\}$. Therefore

$$|I^*_{\ell,b} - I'_{\ell,b}| \le |B_1^{\ell,b}| + |B_2^{\ell,b}| \le 2\gamma|I|/(8k^2).$$

Let $I' = \bigcup_{\ell,b} I'_{\ell,b}$. Summing over all $8k^2$ pairs (ℓ, b), we see that $|I^* - I'| = \bigcup_{\ell,b} |I^*_{\ell,b} - I'_{\ell,b}| \le 2\gamma|I|$. □

¿From this lemma we can obtain the following corollary which is a sharper version of lemmas proved by Ajtai.

Corollary 10. *Let $k \ge 2$ be an integer, $\gamma \in (0, \frac{1}{8})$ be given and $q_1 \le \gamma/(120k)$. Let $r \ge 4k^2/(\gamma q_1^{4k^2})$. Let $P = (T_1, \ldots, T_r)$ be an inquisitive $(r, k/r)$-decision forest. Let $I \subseteq D^X$ be a set of inputs accepted by P. There is an (m, α)-rectangle $R \subseteq D^X$, with $m \ge q_1^{4k^2} n/2$ and $\alpha \ge \gamma|I||D|^{-n}/(8k^2 S(n, 120kq_1m/\gamma))$, such that all inputs in R are accepted by P.*

Proof. Choose any rectangle in one of the families of rectangles in Lemma 9 since I' is non-empty. □

3.3 Branching Program Theorems

Theorem 11. *Let $k \ge 2$ be an integer, $q_1 \le 2^{-60}k^{-8}$, $r = 40k^2/q_1^{4k^2}$, and $|X| = n \ge r^2$. Let B be a branching program of length at most $(k-1)n$ and size 2^S. Let $I \subseteq D^X$ be a set of inputs accepted by B. There is an (m, α)-rectangle on which B outputs 1 where $m \ge q_1^{4k^2} n/2$ and $\alpha \ge 2^{-q_1^{1/2}m - Sr}|I||D|^{-n}$.*

Proof. We first break up B into at most 2^{Sr} inquisitive $(r, k/r)$-decision forests each accepting disjoint portions of I, using Lemma 2. We can do this either

- by adding n dummy queries at the start of B to make its length kn and then applying Lemma 2 to the resulting branching program, or
- by first applying Lemma 2 with $r' = r(k-1)/k$ and then adding r/k decision trees to each decision forest, each of which has all inputs follow the same path with leaf label 1, as described at the beginning of section 3.2.

Thus, there is an inquisitive $(r, k/r)$-decision forest P that accepts at least a 2^{-Sr} fraction of I. Let $\gamma = 3/25 < 1/8$ and observe that $r \geq 4k^2/(\gamma q_1^{4k^2})$. Applying Corollary 10 to P, we obtain an ℓ and b, with $1 \leq \ell < 2k$, $1 < b \leq 4k$, and an (m, α)-rectangle $R \subseteq D^X$, with $m \geq q_1^{4k^2} n/2$ and $\alpha \geq 3 \cdot 2^{-Sr}|I||D|^{-n}/(200k^2 S(n, 1000kq_1 m))$, such that all inputs in R are accepted by P (and therefore B).

Let $\beta = m/n \geq q_1^{4k^2}/2$ and $p = 1000kq_1\beta$. Then $S(n, 1000kq_1 m) \leq 2^{H_2(p)\cdot n}$ where $H_2(p) = p\log_2(1/p) + (1-p)\log_2(1/(1-p))$ is the binary entropy function. Since $m \leq n/2$ and $q_1 \leq 1/(1000k)$, $p < 1/2$ which implies $H_2(p) \leq 2p\log_2(1/p)$. Also $p = 1000kq_1\beta \geq 500kq_1^{4k^2+1}$ implies $\log_2 1/p \leq 5k^2 \log_2 1/q_1$. Hence $H_2(p) \leq 2p\log_2(1/p) \leq ck^3 q_1 \log_2(1/q_1)\beta$, where $c = 2^{15}$. Therefore

$$200k^2 S(n, 1000kq_1 m)/3 \leq 200k^2 2^{ck^3 q_1 \log_2(1/q_1)m}/3 \leq 2^{q_1^{1/2}m}$$

for $q_1 \leq 1/(c^4k^8)$. Thus we obtain the desired lower bound on α. □

Theorem 12. *Let $k \geq 2$ be an integer, $q_1 \leq 2^{-60}k^{-8}$, $r = 40k^2/q_1^{4k^2}$, and $|X| = n \geq r^2$. Let B be a branching program of length at most $(k-1)n$ and size 2^S. Let $f : D^X \to \{0,1\}$ and $\delta = |f^{-1}(1)|/|D^X|$. Suppose that B correctly computes f on all but an ϵ fraction of D^X. Then there is an (m, α)-rectangle $R \subseteq D^X$ accepted by B with $m = q_1^{4k^2} n/2$ and $\alpha \geq 2^{-q_1^{1/2}m - Sr}(\delta - \epsilon)$ such that at most an $80k^2\epsilon/(\delta - \epsilon)$ fraction of the elements of $x \in R$ have $f(x) = 0$.*

Proof. Follow the same pattern as the beginning of the proof of Theorem 11, producing the 2^{Sr} inquisitive $(r, k/r)$-decision forests P which partition the space of inputs that B accepts.

For each of the 2^{Sr} decision forests, P, apply Lemma 9 with $\gamma = 1/10$ and I_P equal to the set of inputs in $f^{-1}(1)$ correctly accepted by P. Lemma 9 gives $8k^2$ partitions $\mathcal{R}_{j,P}$ for $j = 1, \ldots, 8k^2$ into rectangles for each of these sets I_P. Let $I = \bigcup_P I_P$. Throw out all I_P that cover less than a $\gamma/2^{Sr}$ fraction of I. This involves throwing out at most a γ fraction of I in total. By assumption, $|I||D|^{-n} \geq \delta - \epsilon$ and these sets I_P are disjoint. Thus we can derive new partitions $\mathcal{R}_j$ for $j = 1, \ldots, 8k^2$ where $\mathcal{R}_j = \bigcup_P \mathcal{R}_{j,P}$ which together cover all but a 9γ fraction of I since I' covered all but a 8γ fraction of I.

Since the rectangles together cover at least a $(1-9\gamma)(\delta-\epsilon)$ fraction of D^X, there is at least one of the $8k^2$ partitions $\mathcal{R}_j$ that covers at least a $(1-9\gamma)(\delta-\epsilon)/(8k^2)$ fraction of D^X. Call an input $x \in D^X$ a *false positive* if B accepts x but $f(x) = 0$. Since B has at most $\epsilon|D^X|$ false positives, $\mathcal{R}_j$ covers at most $\epsilon|D^X|$ false positives. Therefore the fraction of inputs covered by $\mathcal{R}_j$ that are false positives is at most $8k^2\epsilon/[(1-9\gamma)(\delta-\epsilon)]$. Therefore there is some (m, α)-rectangle R in $\mathcal{R}_j$ that has a fraction of false positives no worse than $8k^2\epsilon/[(1-9\gamma)(\delta-\epsilon)] \leq 80k^2\epsilon/(\delta-\epsilon)$. Choose such a rectangle R.

Using the same reasoning as in the proof of Theorem 11, we obtain the bounds on m and α. □

4 Lower Bounds

4.1 Element Distinctness

The element distinctness problem is the function $ED : D^X \to \{0,1\}$ which outputs 1 if and only if there is no pair $i \neq j \in X$ such that $x(i) = x(j)$.

Proposition 13. *For $N \geq n^2 - n$, $\binom{N}{n} \geq N^n/e$.*

Thus if $|D| \geq n^2 - n$, at least a $1/e$ fraction of all inputs $x \in D^n$ have $ED(x) = 1$. We first use Ajtai's argument to obtain a lower bound for deterministic branching programs computing ED.

Lemma 14. *Let $ED : D^n \to \{0,1\}$. Any (m,α)-rectangle $R \subseteq D^n$ such that $ED(x) = 1$ for all $x \in R$ has $\alpha \leq 2^{-m}$.*

Proof. Suppose that $\alpha > 2^{-m}$. Let A_1, A_2, Y_1, Y_2, and ρ witness the fact that R is an (m,α)-rectangle. For each $i \in A_1 \cup A_2$ let R_i be the set of projections of elements of R onto their i-th coordinate. Clearly, $|Y_1| \leq \prod_{i \in A_1} |R_i|$ and so there is some $i_1 \in A_1$ such that $|R_{i_1}| \geq \alpha^{1/m}|D| > |D|/2$, since $\alpha > 2^{-m}$. Similarly, there is an $i_2 \in A_2$ such that $|R_{i_2}| > |D|/2$. Therefore $R_{i_1} \cap R_{i_2} \neq \emptyset$. Let $j \in R_{i_1} \cap R_{i_2}$. By the rectangle property of R there is an element $x \in R$ such that $x(i_1) = j = x(i_2)$ and thus $ED(x) = 0$ contradicting our assumption. □

Theorem 15. *There is a constant $c > 0$ such that any $[1,n^2]$-way deterministic branching program computing $ED : [1,n^2]^n \to \{0,1\}$ in time T and size 2^S requires $T = \Omega(n\sqrt{\log / \log\log(n/S)})$.*

Proof. Suppose we have a branching program of length $(k-1)n$ and size $s = 2^S$ for ED. Apply Theorem 11 with $q_1 = 2^{-60}k^{-8}$. For $r = 40k^2/q_1^{4k^2}$ and $n \geq r^2$ we obtain an (m,α)-rectangle on which B outputs 1 such that $m \geq q_1^{4k^2}n/2$ and $\alpha \geq 2^{-q_1^{1/2}m - Sr}/e > 2^{-q_1^{1/2}m - Sr - 2}$. Using Lemma 14, this means $2^{-q_1^{1/2}m - Sr - 2} \leq 2^{-m}$ and thus $Sr \geq m(1 - q_1^{1/2}) - 2 \geq q_1^{4k^2}n/4$ or $S \geq q_1^{4k^2}n/(4r)$. Thus for some constant $c > 0$ any algorithm solving ED in time kn requires space at least $k^{-ck^2}n$. Substituting $T = (k-1)n$, and re-arranging we obtain the claimed tradeoff. □

Corollary 16. *For any $\epsilon \geq 0$ there is a constant c_ϵ, such that any RAM algorithm for element distinctness on inputs in the range $[1,n^2]$ taking at most $c_\epsilon n\sqrt{\log / \log\log(n)}$ time requires at least $n^{1-\epsilon}$ space.*

We now consider randomized branching programs for ED. As usual, we prove a lower bound on the ϵ-error randomized complexity of ED by proving a time-space tradeoff lower bound on the complexity of any deterministic algorithm that errs on at most an ϵ fraction of the input space D^X. That is we prove a lower bound on the ϵ-error distribution complexity of ED. In this case we use the uniform distribution on $[1,n^2]^n$.

Lemma 17. *Let $ED : [1,n^2]^n \to \{0,1\}$, $0 < \kappa < 1$, and $0 < \epsilon < 1/12$. If $R \subseteq [1,n^2]^n$ is an exact $(\kappa n,\alpha)$-rectangle such that at most an ϵ fraction of $x \in R$ have $ED(x) = 0$ then*

$$\alpha \leq \max\{\frac{2n}{\kappa}(2\sqrt{2}/3)^{\kappa n}, 2[(8/9)^\kappa 2^{\frac{H_2(4\epsilon)}{3\kappa}}]^n\}.$$

Proof. Essentially, we reduce this problem to proving a similar result about the set-disjointness problem on two sets of size $m = \kappa n$ from a universe of size $n' = n^2 - n + 2m$.

Let A_1, A_2, Y_1, Y_2 and ρ be the associated defining components of R. If any element τ of Y_1 or Y_2 is such that $\rho\tau$ contains two matching elements then $ED(x) = 0$ for all elements x that agree with $\rho\tau$. Assume without loss of generality that this does not occur since it only helps us.

Thus all elements of Y_1 and Y_2 have values chosen from the set $[1, n^2] - range(\rho)$ and furthermore we can associate each element of Y_1 or Y_2 with an m-subset of $[1, n^2] - range(\rho)$. An input $x \in R$ has $ED(x) = 1$ if and only if these associated sets are disjoint. To simplify notation, let F be the collection of m-subsets corresponding to Y_1 and G be the set of m-subsets corresponding to Y_2. We have $|F|, |G| \geq \alpha\binom{n'}{m}$.

We now apply an argument of Babai, Frankl, and Simon [BFS86] showing an ϵ-error communication complexity lower bound to derive a lower bound for this set-disjointness problem. (Note that we cannot apply the optimal communication complexity arguments of Kalayanasundaram and Schnitger [KS87] or Razborov [Raz90] even if we worked with a non-uniform input distribution, since these require precise linear relationships between the set and universe sizes. Surprisingly, the weaker lower bound proved in [BFS86] is sufficiently strong.)

Suppose that $\alpha \geq \frac{2n}{\kappa}(2\sqrt{2}/3)^{\kappa n}$, otherwise we are done. Let F' be the set of all $S \in F$ that intersect with less than a 2ϵ fraction of $T \in G$. Since at most an ϵ fraction of pairs (S, T) in $F \times G$ intersect, $|F'| \geq |F|/2$.

Claim: We can select $p \geq n'/(3m)$ sets $S_1, \ldots, S_p \in F'$ such that $|S_j \cap \bigcup_{i<j} S_i| \leq m/2$ for $j = 1, \ldots, p$; i.e. each S_j has at least half of its elements not occurring in earlier sets.

We construct the sets $S_1, \ldots, S_p$ inductively. Since $p \leq n'/(3m)$, for any j, $|\bigcup_{i<j} S_i| \leq n'/3$. The total number of the $\binom{n'}{m}$ possible m-subsets that have more than half their elements in a given such set of size $n'/3$ is at most $\binom{n'/3}{m/2}\binom{2n'/3}{m/2}$ and there are at most p such sets of size at most n' to consider. Thus in total, at most

$$\begin{aligned} p\binom{n'/3}{m/2}\binom{2n'/3}{m/2} &\leq \frac{n'}{3\kappa n}(1/3)^{m/2}(2/3)^{m/2}\binom{n'}{m/2}^2 \\ &\leq \frac{n'}{3\kappa n}e(8/9)^{m/2}\binom{n'}{m} \\ &< \frac{n}{\kappa}(2\sqrt{2}/3)^{\kappa n}\binom{n'}{m} \end{aligned}$$

sets in F' cannot be used at some stage of the construction (using Proposition 13 to obtain the second line from the first). If α is as large as supposed, F' is large enough that there is always a set $S_j \in F'$ that can be chosen and the claim follows.

Fix the sets $S_1, \ldots, S_p$ according to the claim. Now let G' be the set of $T \in G$ that intersect with at most a 4ϵ fraction of the sets S_j. Since each $S_j \in F'$, each intersects with at most an 2ϵ fraction of sets in G. Therefore at least half the elements of G intersect with at most twice the average number of the S_j, and thus $|G'| \geq |G|/2$.

We now upper bound the number of elements in G' and thus G. We can identify an element T of G' by identifying a collection of $p - 4\epsilon p$ of the p sets S_j that T is not permitted to intersect and then describing which m of the remaining elements of the domain that comprise T. By the claim, any collection of $p(1 - 4\epsilon)$ of the sets has a total of $mp(1 - 4\epsilon)/2 \geq n'(1 - 4\epsilon)/6 \geq n'/9$

elements since $\epsilon < 1/12$. Therefore $|G| \geq 2|G'| \geq 2\binom{p}{4\epsilon p}\binom{8n'/9}{m} < 2 \cdot 2^{H_2(4\epsilon)p}(8/9)^m\binom{n'}{m} < 2 \cdot 2^{H_2(4\epsilon)n/(3\kappa)}(8/9)^{\kappa n}\binom{n'}{m} = 2 \cdot [2^{\frac{H_2(4\epsilon)}{3\kappa}}(8/9)^{\kappa}]^n\binom{n'}{m}$ and thus $\alpha \leq 2 \cdot [2^{\frac{H_2(4\epsilon)}{3\kappa}}(8/9)^{\kappa}]^n$. □

We'd like to avoid the restriction that we only consider exact (m, α)-rectangles. Unfortunately, in the randomized case we cannot simply apply Lemma 1 to find an exact (m, α)-rectangle with small error given an (m, α)-rectangle with small error. Nonetheless we can show:

Lemma 18. *Given an (m, α)-rectangle R which has at most an ϵ fraction of its elements labelled bad, there is an exact $(m, \alpha/2)$-rectangle $R' \subseteq R$ with at most a 4ϵ fraction of its elements labelled bad.*

Proof. Let A_1, A_2, Y_1, Y_2 and ρ be the defining components of R. Choose A_1' and A_2' to be the first m elements of A_1 and A_2, respectively. For each assignment σ to $A_1 - A_1'$ let $Y_1(\sigma)$ be the set of inputs in Y_1 that extend σ. Call a σ *good* if at most a 2ϵ fraction of the elements $(\tau_1, \tau_2) \in Y_1(\sigma) \times Y_2$ have (ρ, τ_1, τ_2) labelled *bad*. Since the different $Y_1(\sigma)$ partition Y_1, at least $1/2$ of the elements of Y_1 are in classes $Y_1(\sigma)$ such that σ is good. Since there are at most $D^{|A_1|-|A_1'|}$ possible good σ and $\sum_{\sigma \text{ good}} |Y_1(\sigma)| \geq |Y_1|/2 \geq (\alpha/2)D^{|A_1|}$, choosing the σ_1 to be the good σ that achieves the maximum size of $Y_1(\sigma)$ we obtain $|Y_1(\sigma_1)| \geq (\alpha/2)D^{|A_1|}/D^{|A_1|-|A_1'|} = (\alpha/2)D^{|A_1'|}$.

We now apply a similar argument considering each assignment σ to $A_2 - A_2'$, and select $Y_2(\sigma_2)$ to be the largest of the sets $Y_2(\sigma)$ such that at most a 4ϵ fraction of the elements in $Y_1(\sigma_1) \times Y_2(\sigma)$ are labelled *bad*. Clearly, the resulting set $T_2(\sigma_2)$ satisfies $|Y_2(\sigma_2)| \geq (\alpha/2)D^{|A_2'|}$. Finally, let $\rho' = \rho\sigma_1\sigma_2$ and Y_i' be the set of assignments to A_i' consisting of the restriction of $Y_i(\sigma_i)$ to A_i'. It is clear that the result is an exact $(m, \alpha/2)$-rectangle having at most a 4ϵ fraction of elements labelled *bad*. □

Theorem 19. *There is a constant $c > 0$ such that if $\epsilon \leq (T/n)^{-c(T/n)^2}$ any ϵ-error randomized $[1, n^2]$-way branching program computing $ED : [1, n^2]^n \to \{0, 1\}$ in time T and size 2^S requires $T = \Omega(n\sqrt{\log / \log\log(n/S)})$.*

Proof. Let $T = (k-1)n$ and without loss of generality that k is an integer and at least 3. We choose $q_1 = 2^{-60}k^{-8}$. Then there is a constant $c > 0$ such that all positive $\epsilon \leq (k-1)^{-c(k-1)^2}$ satisfy $4\epsilon' < 1/2$ and $H_2(4\epsilon') \leq 3\kappa^3$ for all $\kappa \geq q_1^{4k^2}/2$ where $\epsilon' = 240k^2\epsilon$. Suppose that ϵ is at most this value.

Let $r = 40k^2/q_1^{4k^2}$ and $n \geq r^2$. By Theorem 12, if a deterministic branching program disagrees with ED on all but at most an ϵ fraction of inputs then there is an (m, α)-rectangle R with $m = \kappa n$, $\kappa \geq q_1^{4k^2}/2$ and $\alpha \geq 2^{-q_1^{1/2}m - Sr}(1/e - \epsilon) \geq 2^{-q_1^{1/2}m - Sr - 2}$ such that ED is 0 on at most an $\epsilon' = 80k^2\epsilon/(1/e - \epsilon) \leq 240k^2\epsilon$ fraction of R.

Applying Lemma 18 we obtain an exact $(\kappa n, \alpha/2)$-rectangle R' with error at most $4\epsilon'$. By Lemma 17,

$$\alpha/2 \leq \max\{\frac{2n}{\kappa}(2\sqrt{2}/3)^{\kappa n}, 2[(8/9)^{\kappa}2^{\frac{H_2(4\epsilon')}{3\kappa}}]^n\}$$

and thus

$$2^{-q_1^{1/2}\kappa n - Sr - 3} \leq \max\{\frac{2n}{\kappa}(2\sqrt{2}/3)^{\kappa n}, 2[(8/9)^{\kappa}2^{\frac{H_2(4\epsilon')}{3\kappa}}]^n\}.$$

By our condition on ϵ, the second quantity in the maximum is at most $(8/9)^{\kappa n}2^{\kappa^2 n+1} \leq (9/10)^{\kappa n}$ provided that k is sufficiently large. Since $n \geq r^2$, the first quantity is at most $(19/20)^{\kappa n}$. Thus, $q_1^{1/2}\kappa n + Sr \geq c''\kappa n$ for some constant $c'' > 1/20$ independent of k. Therefore $Sr \geq c''\kappa n/2$ which

implies $k^{c'''k^2} \geq n/S$ for some constant $c''' > 0$ independent of k, n and S. The claimed lower bound follows immediately. $\square$

Corollary 20. *There is a constant $c > 0$ so that for any $\delta \geq 0$ there is a constant c_δ such that any randomized RAM algorithm for element distinctness on inputs in the range $[1, n^2]$ taking at most $c_\delta n\sqrt{\log n/\log\log n}$ time and having at most $c\frac{\log\log n}{\log n}$ error requires at least $n^{1-\delta}$ space.*

Boolean Branching Programs Computing Quadratic Forms

Following [BST98], Ajtai looks at quadratic forms $x^T Mx$ for certain matrices M. Using ideas of Borodin, Razborov, Smolensky [BRS93], Beame, Saks, and Thathachar had shown a relationship between the approximation of $x^T Mx$ on (m, α)-rectangles and the minimal rank of sub-matrices (known as the rigidity) of M. The Sylvester matrices considered in [BST98] do not have sufficiently strong rigidity properties to obtain a non-trivial time-space tradeoff for the specific parameters m and α for which Ajtai needs them.

Instead, Ajtai looks at Hankel matrices, matrices whose every anti-diagonal is constant. Although an $n \times n$ Hankel matrix has n^2 entries, it only takes $2n - 1$ values to specify a Hankel matrix and therefore one can get explicit functions even by looking at random Hankel matrices. Ajtai shows the following lemma concerning the rigidity properties of random Hankel matrices over $GF(2)$.

Lemma 21. *Assume that n, s, R, t are positive integers, $t^2 < s < n$, $R < Q = \lfloor s/t^2 \rfloor$. If M is a random $n \times n$ Hankel matrix over $GF(q)$, the probability that there is some $s \times s$ sub-matrix of M of rank less than R is at most*

$$\binom{n}{Qt}^2 \binom{Q}{Q-R+1} q^{-\frac{1}{4}(Q-R+1)t^2}.$$

We restate this in a more convenient form, making no attempt to optimize constants.

Corollary 22. *Let M be a random $n \times n$ Hankel matrix over $GF(2)$. Let δ satisfy $\delta n \geq 24$ and $1/\delta \geq 2^{16}\log_2^2(1/\delta)$. With probability at least 1/2 a random $\delta n \times \delta n$ minor of M has rank at least $\delta n/(256\log_2(1/\delta))^2$.*

Proof. Let $s = \delta n$, $t = 128\log_2(1/\delta)$, $Q = \lfloor s/t^2 \rfloor$, and $R = Q/2$. The failure probability from Lemma 21 is at most

$$\binom{n}{Qt}^2 \binom{Q}{Q-R+1} q^{-\frac{1}{4}(Q-R+1)t^2} \leq 2^{[2H_2(\delta/t)+(\delta/t^2)n-\frac{1}{9}\delta]n}$$

Now

$$\begin{aligned} H_2(\delta/t) &= -(\delta/t)\log_2(\delta/t) - (1-\delta/t)\log_2(1-\delta/t) \\ &\leq -2(\delta/t)\log_2(\delta/t) \\ &\leq -4(\delta/t)\log_2(\delta) \\ &\leq \delta/32 \end{aligned}$$

Therefore the failure probability is at most $2^{[1/16+(1/t^2)-1/9]\delta n} \leq 2^{-\frac{1}{24}\delta n} \leq 1/2$ since $\delta n \geq 24$. $\square$

The following lemma relates the rank of a matrix to the fraction of inputs on which a bilinear form based on the matrix outputs given values. Note that part (a), which was used by Ajtai, was already shown in [BRS93] and can be easily proved more directly by a rank argument. We need the stronger results of either part (b) or (c) to obtain our bounds.

Lemma 23. *Let M be an $m \times m$ matrix over $GF(2)$ of rank r. Let $b \in \{0,1\}$. Let $U, V \subseteq GF(2)^m$, $\alpha = |U|/2^m$, and $\beta = |V|/2^m$. Then*

(a) if $\alpha\beta > 2^{-r}$ then $u^T M v$ is not constant for all $(u,v) \in U \times V$

(b) if $\alpha\beta \geq m^2 2^{2-r}$ then $u^T M v$ evaluates to b for at least a $1/(16m^2)$ fraction of $(u,v) \in U \times V$.

(c) if $\alpha\beta \geq m^2 2^{3-r}$ then $u^T M v$ evaluates to b for at least a $1/32$ fraction of $(u,v) \in U \times V$.

Proof. We can write $M = KL$ where K is an $m \times r$ matrix over $GF(2)$ of rank r and L is an $r \times m$ matrix over $GF(2)$ of rank r. Therefore $u^T M v = u^T K L v = (K^T u)^T L v = x^T y$ for $x = K^T u$ and $y = Lv$. Observe that both K^T and L are $m \times r$ matrices of rank r that map 2^{m-r} points of $GF(2)^m$ onto each point of $GF(2)^r$. Therefore, the sets $X = K^T U$ and $Y = LV$ satisfy $|X| \geq \alpha 2^r$ and $|Y| \geq \beta 2^r$. Now, by well-known properties of inner product (see, e.g., [BFS86]), the difference between the number of 0's and 1's of $x^T y$ on $X \times Y$ is at most $\sqrt{\alpha\beta} 2^{3r/2}$. Now for $\alpha\beta > 2^{-r}$, the number of points in $X \times Y$ is at least $2^r > \sqrt{\alpha\beta} 2^{3r/2}$, and so $x^T y$ is non-constant on $X \times Y$. Thus $u^T M v$ is non-constant on $U \times V$.

Also, for $\alpha\beta \geq 2^{2-r}$, the difference between the number of 0's and 1's in $X \times Y$ is at most 2^{r+1} but the number of elements in $X \times Y$ is at least 2^{r+2}. Therefore, at least $1/4$ of the elements of $X \times Y$ take on value 1 and $1/4$ of the elements of $X \times Y$ take on value 0. However, we would like to show that $u^T M v$ takes on a significant fraction of both 0's and 1's on $U \times V$. The complication is that each value $x \in X$ and $y \in Y$ is not the image of the same number of elements of U and V respectively.

We handle this by grouping elements of $U \times V$ based on the approximate sizes of these pre-images. For $i = 0, \ldots, m-r$, let U_i be the set of points $u \in U$ such that $2^i \leq |\{u' \in U \mid K^T u' = K^T u\}| < 2^{i+1}$ and for $j = 0, \ldots, m-r$, let V_i be the set of points $v \in V$ such that $2^j \leq |\{v' \in V \mid Lv' = Lv\}| < 2^{j+1}$. Let $\alpha_i = |U_i|/2^m$ and $\beta_j = |V_j|/2^m$, $X_i = K^T U_i$ and $Y_j = LV_j$. Clearly, $|X_i| \geq \alpha_i 2^r$ and $|Y_j| \geq \beta_j 2^r$. Also, observe that any two points in $X_i \times Y_j$ differ by at most a factor of 4 in the number of pre-images in $U \times V$ (which are all in $U_i \times V_j$).

Now if we choose the largest U_i and V_j, we have $\alpha_i \geq \alpha/m$ and $\beta_j \geq \beta/m$ and thus, since $\alpha\beta/m^2 \geq 2^{2-r}$, at least $1/4$ of the elements of $X_i \times Y_j$ take on each value. Therefore, at least $1/16$ of the elements of $U_i \times V_j$ take on each value. By choice of U_i and V_j, therefore at least $1/(16m^2)$ of the elements of $U \times V$ take on each value.

We can do better by a subtler argument. There are at most m^2 sets $U_i \times V_j$ and these partition $U \times V$. At most $1/2$ of the elements in $U \times V$ are in sets $U_i \times V_j$ such that $|U_i \times V_j| = \alpha_i \beta_j 2^{2m} < \alpha\beta 2^{2m}/(2m^2)$. For each of the remaining sets $U_i \times V_j$, we have $|X_i \times Y_j| \geq \alpha\beta 2^{2r}/(2m^2) \geq 2^{2+r}$, and so at least $1/4$ of the elements of $X_i \times Y_j$ take on each value. Therefore, at least $1/16$ of the elements of each these $U_i \times V_j$ take on each value. Since together these $U_i \times V_j$ cover at least $1/2$ of $U \times V$, $u^T M v$ on $U \times V$ takes on each value at least $1/32$ of the time. □

The relationship between the rank of sub-matrices and the approximation of quadratic forms given in [BST98] works only for $GF(q)$ with $char(q) \neq 2$ and needs to be modified to handle $GF(2)$.

To handle $GF(2)$, instead of working with the Hankel matrix directly, Ajtai zeroes out the entries above the diagonal to get a matrix M' based on the Hankel matrix M, otherwise, except for the diagonal, all the submatrix products of $x^T Mx$ would appear twice, once above and once below the diagonal, and therefore cancel each other out modulo 2. We first will want to show that our (m, α)-rectangle can be chosen whose defining sets A_1, A_2 correspond to a large block below the diagonal of M'.

Lemma 24. *Let $X = \{1, \ldots, n\}$ and consider the space D^X. Given an (m, α)-rectangle R which has at most an ϵ fraction of its elements labelled bad, there is an exact $(\lceil m/2 \rceil, \alpha/2)$-rectangle $R' \subseteq R$ with at most a 4ϵ fraction of its elements labelled bad and with the defining sets A'_1 and A'_2 of R' having the property that every element of A'_1 is smaller than every element of A'_2.*

Proof. Let A_1, A_2 Y_1, Y_2, and ρ be the defining components of R. Let a_1, a_2 be the median elements of A_1 and A_2, respectively. If $a_1 < a_2$, then let A'_1 be the first $\lceil m/2 \rceil$ elements of A_1 and A'_2 be the last $\lceil m/2 \rceil$ elements of A_2. Otherwise, we can reverse the roles of A_1 and A_2 so we assume $a_1 < a_2$ without loss of generality. We now apply the same argument as in the proof of Lemma 18 to obtain the desired conclusion. Note that in the case there are no inputs labelled *bad*, we can apply the argument of Lemma 1 to obtain an exact $(\lceil m/2 \rceil, \alpha)$-rectangle. □

Theorem 25. *Let $F(x, y)$ be the function of $3n-1$ bits that defines $x^T M' x$ for the modified Hankel matrix given by y. Any deterministic Boolean branching program computing $F(x, y)$ in time T and size 2^S requires $T = \Omega(n\sqrt{\log / \log\log(n/S)})$. Furthermore, there is a constant $c' > 0$ such that any randomized Boolean branching program computing $F(x, y)$ in time T and size 2^S with at most $\epsilon < c'n^2/T^2$ error requires $T = \Omega(n\sqrt{\log / \log\log(n/S)})$.*

Proof. Fix the value of y to create a matrix M that satisfies the conditions of Corollary 22. We obtain our lower bound by considering the distribution of inputs x uniformly chosen from $GF(2)^n$ and show a lower bound on the time and space for a deterministic branching program B computing $f(x) = x^T M' x$ with at most ϵ error. Let $b \in \{0, 1\}$ be a value of f that at least $1/2$ the time. Without loss of generality we assume that $b = 1$; if not we swap the sinks of B to obtain a program computing $\bar{f}$ with at most ϵ error.

We first show the case when $\epsilon = 0$, i.e., deterministic lower bounds. Let $q_1 = 2^{-60}k^{-10}$. Applying Theorem 11 to B with I equal to the set of inputs on which B outputs 1 we obtain that for $r = 40k^2/q_1^{4k^2}$ and $n \geq r^2$ there is a $(\kappa n, 2^{-q_1^{1/2}\kappa n - Sr - 1})$-rectangle on which B outputs 1, where $\kappa \geq q_1^{4k^2}/2$. Apply Lemma 24 to get two sets A_1, A_2 of size precisely $\lceil \kappa n/2 \rceil$ with all elements of A_1 less than all elements of A_2 and an exact $(\lceil \kappa n/2 \rceil, \alpha)$-rectangle R based on (A_1, A_2) for $\alpha \geq 2^{-q_1^{1/2}\kappa n - Sr - 1}$. For any $x \in R$, with $u \in Y_1$ and $v \in Y_2$ equal to the values of x on A_1 and A_2 respectively,

$$\begin{aligned}
x^T M' x &= (\rho + u + v)^T M' (\rho + u + v) \\
&= \rho^T M' \rho + u^T M'_{A_1, A_2} v + v^T M'_{A_2, A_1} u \\
&\quad + (u^T M'_{A_1, A_1} u + \rho^T M' u + u^T M' \rho) + (v^T M'_{A_2, A_2} v + \rho^T M' v + v^T M' \rho), \\
&= \rho^T M' \rho + u^T M'_{A_1, A_2} v + f_1(u) + f_2(v)
\end{aligned}$$

where $f_1(u) = u^T M'_{A_1, A_1} u + \rho^T M' u + u^T M' \rho$ and $f_2(v) = v^T M'_{A_2, A_2} v + \rho^T M' v + v^T M' \rho$, since M'_{A_2, A_1} is a 0-matrix being entirely above the diagonal. Also, $\rho^T M' \rho$ is constant. At least $1/2$ of all elements in Y_1 have one of the two values of $f_1(u)$; fix this more popular value to get Y'_1. Do

the same for Y_2 to obtain a $(\lceil \kappa n/2 \rceil, \alpha/2)$-rectangle R on which these three terms are constant. Thus $x^T M' y$ is constant on R' if and only if $u^T M'_{A_1,A_2} v = u^T M_{A_1,A_2} v$ is constant on $Y_1' \times Y_2'$. By Lemma 23(a) and Corollary 22 applied to M, we obtain $\frac{1}{2}\kappa/(256\log_2(2/\kappa))^2 \leq q_1^{1/2}\kappa + (Sr+2)/n$. Now $q_1^{1/2} \ll 0.5/(256\log_2(2/\kappa))^2$ for k sufficiently large, therefore $S \geq \frac{1}{4r}\kappa n/(256\log_2(2/\kappa))^2 \geq n/k^{ck^2}$ for some constant $c > 0$. Using $k = T/n$ and re-arranging, we obtain the desired result.

We now consider the case that $0 < \epsilon < 1/(2^9 \cdot 200k^2)$. Let $q_1 = 2^{-60}k^{-10}$. Applying Theorem 12 to B we obtain that for $r = 40k^2/q_1^{4k^2}$ and $n \geq r^2$ there is a $(m = \kappa n, \alpha = 2^{-q_1^{1/2}\kappa n - Sr - 2})$-rectangle R with $\kappa \geq q_1^{4k^2}/2$ on which B outputs 1 but for which f is 0 for at most an $80k^2\epsilon/(1/2-\epsilon) \leq 200k^2\epsilon$ fraction of R. Apply Lemma 24 to get two sets A_1, A_2 of size precisely $\lceil \kappa n/2 \rceil$ with all elements of A_1 less than all elements of A_2 and an exact $(\lceil \kappa n/2 \rceil, \alpha/2)$-rectangle R' based on (A_1, A_2) on which f is 0 on at most a $2^2 \cdot 200k^2\epsilon$ fraction of elements. As above for $x \in R'$ we get

$$x^T M' x = \rho^T M' \rho + u^T M'_{A_1,A_2} v + f_1(u) + f_2(v),$$

for $u \in Y_1$ and $v \in Y_2$ where Y_1 and Y_2 are the sets of assignments to A_1 and A_2 defining R'. At least $1/2$ of all elements in Y_1 have one of the two values of $f_1(u)$; fix this more popular value to get Y_1'. Do the same for Y_2 to obtain a $(\lceil \kappa n/2 \rceil, \alpha/4)$-rectangle R'' on which the terms other than $u^T M_{A_1,A_2} v$ are constant. Observe that since R'' contains at least $1/4$ of the elements of R, the fraction of elements of R'' on which f is 0 is less than 4 times that on R', or $2^4 \cdot 200k^2\epsilon < 1/32$ by the bound on ϵ. However f is 0 on R'' if and only if such that $u^T M'_{A_1,A_2} v = u^T M_{A_1,A_2} v = b'$ for some $b' \in \{0,1\}$. By Lemma 23(c), this implies that $(\alpha/4)^2 \leq \lceil \kappa n/2 \rceil^2 2^{3-rank(M_{A_1,A_2})}$. Therefore by Corollary 22, we obtain

$$2^{-2q_1^{1/2}\kappa n - 2Sr - 8} \leq \lceil \kappa n/2 \rceil^2 2^{3-\lceil \kappa n/2 \rceil/(256\log_2(2/\kappa))^2}.$$

Therefore,

$$(2Sr + 11) \geq \frac{1}{2}\kappa n/(256\log_2(2/\kappa))^2 - 2q_1^{1/2}\kappa n - 2\log_2\lceil \kappa n/2 \rceil.$$

Now $2q_1^{1/2} \ll 0.5/(256\log_2(2/\kappa))^2$ for k sufficiently large and $2\log_2\lceil \kappa n/2 \rceil < n$ so $Sr \geq \frac{1}{8}\kappa n/(256\log_2(2/\kappa))^2$ for n sufficiently large. Therefore $S \geq n/k^{ck^2}$ for some constant $c > 0$. Again, the desired result follows immediately.

□

We note that the key rigidity property of the Hankel matrix that we needed for the lower bound was that any of its $\kappa n \times \kappa n$-minors has rank significantly larger than $\kappa n/(\log(1/\kappa))^{2+\epsilon}$ for any $\epsilon > 0$. The Sylvester matrices considered in [BST98] only have a rank guarantee of $\kappa^2 n$.

Acknowledgements

Thanks to Jayram Thathachar for many discussions about these branching program problems.

References

[Abr90] Karl R. Abrahamson. A time-space tradeoff for Boolean matrix multiplication. In *Proceedings 31st Annual Symposium on Foundations of Computer Science*, pages 412–419, St. Louis, MO, October 1990. IEEE.

[Abr91] Karl R. Abrahamson. Time–space tradeoffs for algebraic problems on general sequential models. *Journal of Computer and System Sciences*, 43(2):269–289, October 1991.

[Ajt98] M. Ajtai. Determinism versus non-determinism for linear time RAMs with memory restrictions. Technical Report TR98-077, Electronic Colloquium in Computation Complexity, `http://www.eccc.uni-trier.de/eccc/`, 1998. Revision 1.

[Ajt99a] M. Ajtai. Determinism versus non-determinism for linear time RAMs with memory restrictions. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, 1999.

[Ajt99b] M. Ajtai. A non-linear time lower bound for boolean branching programs. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*. IEEE, 1999.

[BC82] Allan Borodin and Stephen A. Cook. A time-space tradeoff for sorting on a general sequential model of computation. *SIAM Journal on Computing*, 11(2):287–297, May 1982.

[Bea91] Paul W. Beame. A general time-space tradeoff for finding unique elements. *SIAM Journal on Computing*, 20(2):270–277, 1991.

[BFK+81] Allan Borodin, Michael J. Fischer, David G. Kirkpatrick, Nancy A. Lynch, and Martin Tompa. A time-space tradeoff for sorting on non-oblivious machines. *Journal of Computer and System Sciences*, 22(3):351–364, June 1981.

[BFMadH+87] Allan Borodin, Faith E. Fich, Friedhelm Meyer auf der Heide, Eli Upfal, and Avi Wigderson. A time-space tradeoff for element distinctness. *SIAM Journal on Computing*, 16(1):97–99, February 1987.

[BFS86] László Babai, P. Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, Toronto, Ontario, October 1986. IEEE.

[Bor93] Allan Borodin. Time space tradeoffs (getting closer to the barrier?). In *4th International Symposium on Algorithms and Computation*, pages 209–229, Hong Kong, December 1993.

[BRS93] Allan Borodin, A. A. Razborov, and Roman Smolensky. On lower bounds for read-k times branching programs. *Computational Complexity*, 3:1–18, October 1993.

[BST98] Paul W. Beame, Michael Saks, and Jayram S. Thathachar. Time-space tradeoffs for branching programs. In *Proceedings 39th Annual Symposium on Foundations of Computer Science*, pages 254–263, Palo Alto, CA, November 1998. IEEE.

[Cob66] Alan Cobham. The recognition problem for the set of perfect squares. Research Paper RC-1704, IBM Watson Research Center, 1966.

[For97] Lance Fortnow. Nondeterministic polynomial time versus nondeterministic logarithmic space: Time-space tradeoffs for satisfiability. In *Proceedings, Twelfth Annual IEEE Conference on Computational Complexity*, pages 52–60, Ulm, Germany, 24–27 June 1997. IEEE Computer Society Press.

[FvM00] L. Fortnow and D. van Melkebeek. Time-space tradeoffs for nondeterministic computation. In *Proceedings, Fifteenth Annual IEEE Conference on Computational Complexity*. IEEE Computer Society Press, July 2000. To appear.

[KS87] B. Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings, Structure in Complexity Theory, Second Annual Conference*, pages 41–49, Cornell University, Ithaca, NY, June 1987. IEEE.

[LV99] R. Lipton and A. Viglas. Time-space tradeoffs for sat. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*. IEEE, 1999.

[MNT93] Y. Mansour, N. Nisan, and P. Tiwari. The computational complexity of universal hashing. *Theoretical Computer Science*, 107:121–133, 1993.

[Pip79] Nicholas J. Pippenger. On simultaneous resource bounds. In *20th Annual Symposium on Foundations of Computer Science*, pages 307–311, San Juan, Puerto Rico, October 1979. IEEE.

[Raz90] A. A. Razborov. On the distributional complexity of disjointness. In Michael S. Paterson, editor, *Automata, Languages, and Programming: 17th International Colloquium*, volume 443 of *Lecture Notes in Computer Science*, pages 249–253, Warwick University, England, July 1990. Springer-Verlag.

[Yao88] A. C. Yao. Near-optimal time-space tradeoff for element distinctness. In *29th Annual Symposium on Foundations of Computer Science*, pages 91–97, White Plains, NY, October 1988. IEEE.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'