

Testing Juntas Nearly Optimally

Eric Blais^{*}
Carnegie Mellon University
School of Computer Science
Pittsburgh, PA
eblais@cs.cmu.edu

ABSTRACT

A function on n variables is called a k -junta if it depends on at most k of its variables. In this article, we show that it is possible to test whether a function is a k -junta or is “far” from being a k -junta with $O(k/\epsilon + k \log k)$ queries, where ϵ is the approximation parameter. This result improves on the previous best upper bound of $\tilde{O}(k^{3/2})/\epsilon$ queries and is asymptotically optimal, up to a logarithmic factor.

We obtain the improved upper bound by introducing a new algorithm with one-sided error for testing juntas. Notably, the algorithm is a valid junta tester under very general conditions: it holds for functions with arbitrary finite domains and ranges, and it holds under any product distribution over the domain.

A key component of the analysis of the new algorithm is a new structural result on juntas: roughly, we show that if a function f is “far” from being a k -junta, then f is “far” from being determined by k parts in a random partition of the variables. The structural lemma is proved using the *Efron-Stein decomposition* method.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: Probabilistic algorithms;
F.2.m [Analysis of Algorithms and Problem Complexity]: Miscellaneous

General Terms

Algorithms, Theory

Keywords

Property Testing, Juntas, Efron-Stein decomposition

^{*}Research supported in part by a scholarship from the Fonds québécois de recherche sur la nature et les technologies (FQRNT).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

STOC’09, May 31–June 2, 2009, Bethesda, Maryland, USA.
Copyright 2009 ACM 978-1-60558-506-2/09/05 ...\$5.00.

1. INTRODUCTION

In many areas of science, data collection methods are rapidly becoming more sophisticated. As a result, datasets obtained from experiments contain an increasing number of features. For example, until recently biologists were only able to measure the expression of a handful of genes at a time; now they can measure the expression of tens of thousands of genes simultaneously [12].

Datasets with large numbers of features provide new opportunities, but they also introduce new challenges. For the task of learning a target function, a large number of features causes naïve learning algorithms to overfit the data, and can lead to the formulation of hypotheses that are hard to interpret. The *feature subset selection* method is commonly used to avoid both those challenges [11, §3.4].

The feature subset selection method is appropriate when the target function depends on only k of the n features in a dataset, for some $k \ll n$. We call such target functions *k-juntas*. In many cases, the target function may be quite far from being a k -junta. When this is the case, the feature subset selection method is bound to fail. It is therefore preferable to *test* a target function to see if it is a k -junta before attempting to learn it via the feature subset selection method.

In this article, we study the problem of testing k -juntas in the *property testing* framework. Informally, we seek to determine the minimum number of queries to a function required to distinguish k -juntas from functions that are “far” from being k -juntas, for some appropriate notion of distance. (See Section 2 for formal definitions.)

1.1 Previous work

The first result explicitly related to testing juntas was obtained by Parnas, Ron, and Samorodnitsky [17], who generalized a result of Bellare, Goldreich, and Sudan [2] on testing long codes to obtain an algorithm for testing 1-juntas (i.e., dictators) with only $O(1/\epsilon)$ queries.

Soon afterwards, Fischer *et al.* [9] introduced algorithms for testing k -juntas with $\tilde{O}(k^2)/\epsilon$ queries. The original analysis of Fischer *et al.* only applied to functions with a boolean range; Diakonikolas *et al.* [7] extended the analysis to handle functions with arbitrary finite ranges.

The junta-testing algorithms of Fischer *et al.* remained the most query-efficient ways to test juntas until very recently, when the current author introduced an algorithm for testing boolean functions for the property of being k -juntas with $\tilde{O}(k^{3/2})/\epsilon$ queries [3].

The first non-trivial lower bound on the query complexity of the testing juntas problem was provided by Fischer *et al.* [9], who showed that $\Omega(\log k)$ queries are necessary to test k -juntas.¹ That lower bound was subsequently improved to $\Omega(k)$ by Chockler and Gutfreund [6].

1.2 Our results

The research presented in this article was motivated by the desire to close the gap between the upper and lower bounds on the query complexity of the junta testing problem. Our main result is a new algorithm for testing juntas that significantly improves the upper bound.

THEOREM 1.1. *The number of queries required to ϵ -test k -juntas is bounded above by $O(k/\epsilon + k \log k)$. Furthermore, this result holds for testing functions that have arbitrary finite product domains and arbitrary finite ranges, and it also holds under any product distribution over the domain.*

Combined with the lower bound of Chockler and Gutfreund [6], this completely characterizes the asymptotic query complexity for ϵ -testing k -juntas (up to a logarithmic factor) for constant values of ϵ .

The new algorithm for testing juntas, presented in Section 3.1, is surprisingly simple. It is, however, quite general. As Theorem 1.1 indicates, it can test functions with arbitrary finite product domains and arbitrary finite ranges for the property of being a k -junta, and it also is a valid tester under the general property testing framework where distance is measured by any product distribution over the input. Furthermore, the algorithm has *one-sided error*: it always accepts k -juntas.

The analysis of the algorithm constitutes the main technical contribution of this current research. At the heart of the analysis lies a fundamental structural lemma about juntas: roughly, the lemma states that if a function is “far” from being a k -junta, then it will also be “far” from being determined by the coordinates in k parts in a (sufficiently fine) random partition of the coordinates. The lemma is presented in Section 3.2 and its proof is presented in Section 4.

1.3 Our techniques

The analysis of the junta testing algorithm and the proof of our main structural lemma rely on the analysis of the *influence* of coordinates in a function. The main tool we use to do this is the *Efron-Stein decomposition* method.

The Efron-Stein decomposition of a function is a coarser version of the Fourier decomposition of a function. While both decompositions share many similarities, it can be more convenient to work with the Efron-Stein decomposition when the range of the function is not boolean. In particular, the Efron-Stein decomposition provides a much simpler analysis of junta tests for functions with non-boolean ranges than the approach of Diakonikolas *et al.* [7].

The Efron-Stein decomposition method has found numerous applications in statistics [8, 13, 19], hardness of approximation [1, 14, 15], learning theory [4], and social choice theory [15]. As we see below, the method is also particularly well suited for the analysis of juntas.

¹In fact, Fischer *et al.* proved the stronger statement that $\Omega(\sqrt{k})$ queries are required to test k -juntas *non-adaptively*.

2. PRELIMINARIES

Throughout this article, we consider functions of the form $f : \mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ is a finite set and \mathcal{Y} is an arbitrary finite set. We define $\Omega = \Omega_1 \times \dots \times \Omega_n$ to be a product probability space over \mathcal{X} , where $\Omega_i = (\mathcal{X}_i, \mu_i)$ is defined by an arbitrary probability measure μ_i on \mathcal{X}_i .

For the elements $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathcal{X}$ and the set $S \subseteq [n]$, we let x_S represent the ordered list $(x_i : i \in S)$ and use the notation $x_S y_{\bar{S}}$ to represent the element $z = (z_1, \dots, z_n) \in \mathcal{X}$ where $z_S = x_S$ and $z_{\bar{S}} = y_{\bar{S}}$.

Throughout the rest of this article, $\mathcal{I} = \{I_1, \dots, I_s\}$ denotes a *random partition* of the coordinates in $[n]$ obtained by uniformly and independently assigning each coordinate at random to one of the parts I_1, \dots, I_s .

2.1 Juntas

DEFINITION 2.1 (INFLUENCE). *The influence of the set $S \subseteq [n]$ of coordinates in the function $f : \mathcal{X} \rightarrow \mathcal{Y}$ under the probability space Ω is*

$$\text{Inf}_f(S) \stackrel{\text{def}}{=} \Pr_{x, y \sim \Omega} [f(x) \neq f(y_S x_{\bar{S}})].$$

When $\text{Inf}_f(S) > 0$, we say that the set S of coordinates is *relevant* to f , or alternatively that f depends on the coordinates in S .

DEFINITION 2.2 (JUNTAS). *The function $f : \mathcal{X} \rightarrow \mathcal{Y}$ is a k -junta if it has at most k relevant coordinates.*

Conversely, f is ϵ -far from being a k -junta under Ω if for every k -junta $g : \mathcal{X} \rightarrow \mathcal{Y}$, $\Pr_{x \sim \Omega} [f(x) \neq g(x)] \geq \epsilon$.

The analysis of the junta testing algorithm relies on the following characterization of non-juntas.

PROPOSITION 2.3. *If $f : \mathcal{X} \rightarrow \mathcal{Y}$ is ϵ -far from being a k -junta, then for every set $J \subseteq [n]$ of size $|J| \leq k$,*

$$\text{Inf}_f([n] \setminus J) \geq \epsilon.$$

In our analysis of the junta testing algorithm, we also show that there is a close connection between juntas and *partition juntas*.

DEFINITION 2.4 (PARTITION JUNTAS). *Let \mathcal{I} be a partition of $[n]$. The function $f : \mathcal{X} \rightarrow \mathcal{Y}$ is a k -part junta with respect to \mathcal{I} if the relevant coordinates in f are all contained in at most k parts of \mathcal{I} .*

Conversely, f is ϵ -far from being a k -part junta with respect to \mathcal{I} under Ω if for every set J formed by taking the union of k parts in \mathcal{I} , $\text{Inf}_f([n] \setminus J) \geq \epsilon$.

The proof of Proposition 2.3 is included in Appendix A.2.

We examine the problem of testing juntas in the property testing model introduced by Goldreich, Goldwasser, and Ron [10], which is a generalized version of the property testing model of Rubinfeld and Sudan [18].

DEFINITION 2.5 (JUNTA TESTERS). *A randomized algorithm \mathcal{A} that queries a given function on a small number of inputs is an ϵ -tester for k -juntas under the distribution Ω if it accepts k -juntas with probability at least $2/3$, and rejects functions that are ϵ -far from being k -juntas under Ω with probability at least $2/3$.*

The query complexity of the algorithm \mathcal{A} is the number of queries it makes to the function before accepting or rejecting it.

If a junta tester accepts k -juntas with probability 1, then we say that it has *one-sided error*.

2.2 Efron-Stein decomposition

Let $\mathbb{R}^{\mathcal{Y}}$ be the vector space generated by the set of all formal linear combinations of elements in \mathcal{Y} . Given two vectors $v = \sum_{\mathbf{y} \in \mathcal{Y}} v_{\mathbf{y}} \mathbf{y}$ and $w = \sum_{\mathbf{y} \in \mathcal{Y}} w_{\mathbf{y}} \mathbf{y}$ in $\mathbb{R}^{\mathcal{Y}}$, we define their inner product to be $\langle v, w \rangle_{\mathbb{R}^{\mathcal{Y}}} = \sum_{\mathbf{y} \in \mathcal{Y}} v_{\mathbf{y}} w_{\mathbf{y}}$.

The set of all functions of the form $f : \mathcal{X} \rightarrow \mathbb{R}^{\mathcal{Y}}$ forms the inner product space $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ under the inner product

$$\langle f, g \rangle = \mathbf{E}_{x \sim \Omega} [\langle f(x), g(x) \rangle_{\mathbb{R}^{\mathcal{Y}}}] .$$

The norm of a function $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ is defined by

$$\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathbf{E}_{x \sim \Omega} [\langle f(x), f(x) \rangle_{\mathbb{R}^{\mathcal{Y}}}] .}$$

By identifying elements in \mathcal{Y} with elements in $\mathbb{R}^{\mathcal{Y}}$ in the natural way (i.e., by identifying $\mathbf{y} \in \mathcal{Y}$ with the formal linear combination $1 \cdot \mathbf{y} \in \mathbb{R}^{\mathcal{Y}}$), we observe that the set of functions of the form $f : \mathcal{X} \rightarrow \mathcal{Y}$ forms a subset of $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$. We call such functions *pure-valued functions*. The norm of a pure-valued function f is $\|f\|_2 = 1$.

THEOREM 2.6 (EFRON-STEIN [8]). *Every function f in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ has a unique decomposition of the form*

$$f(x) = \sum_{S \subseteq [n]} f^S(x)$$

where for every $S \subseteq [n]$, $S' \not\supseteq S$, and $y \in \mathcal{X}$:

1. f^S depends only on the coordinates in S , and
2. $\mathbf{E}_{x \sim \Omega} [f^S(x) \mid x_{S'} = y_{S'}] = 0$.

The Efron-Stein decomposition is an *orthogonal* decomposition of functions.² As a result, Parseval's identity holds in this context.

THEOREM 2.7 (PARSEVAL'S IDENTITY). *For every function $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$,*

$$\sum_{S \subseteq [n]} \|f^S\|_2^2 = \|f\|_2^2 .$$

In particular, when f is a pure-valued function,

$$\sum_{S \subseteq [n]} \|f^S\|_2^2 = 1 .$$

REMARK 2.8. *When $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is a boolean function, the Efron-Stein decomposition of f is the same as its Fourier decomposition (i.e., for every set $S \subseteq [n]$, $f^S = \hat{f}(S) \chi_S$). This can be easily verified by noting that the functions $\hat{f}(S) \chi_S$ satisfy the two conditions of Theorem 2.6.*

2.3 Influence

There is a natural connection between the influence of coordinates in a function and the Efron-Stein decomposition of that function.

²That is, the sets of projections $\mathcal{H}^S = \{f^S : f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})\}$ form orthogonal subspaces of $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$.

PROPOSITION 2.9. *For every pure-valued function f in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ and every set $S \subseteq [n]$,*

$$\text{Inf}_f(S) = \sum_{T : S \cap T \neq \emptyset} \|f^T\|_2^2 .$$

For completeness, we include the proof of Proposition 2.9 in Appendix A.1.

The monotonicity and subadditivity properties of influence follow directly from the proposition.

COROLLARY 2.10 (MONOTONICITY & SUBADDITIVITY). *For any pure-valued function $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ and any sets $S, T \subseteq [n]$,*

$$\text{Inf}_f(S) \leq \text{Inf}_f(S \cup T) \leq \text{Inf}_f(S) + \text{Inf}_f(T) .$$

Another key contribution of Proposition 2.9 to the proof of our main lemma is that it suggests two natural extensions to the definition of influence: *low-order* and *high-order* influence.

DEFINITION 2.11 (LOW- & HIGH-ORDER INFLUENCE). *The influence of order at most k of a set $S \subseteq [n]$ of coordinates in the pure-valued function $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ is*

$$\text{Inf}_f^{\leq k}(S) = \sum_{|T| \leq k : S \cap T \neq \emptyset} \|f^T\|_2^2$$

and the influence of order greater than k of the set S on f is

$$\text{Inf}_f^{> k}(S) = \sum_{|T| > k : S \cap T \neq \emptyset} \|f^T\|_2^2 .$$

Proposition 2.9, along with Parseval's identity, makes it easy to show that the sum of the low-order influence of each coordinate in a pure-valued function cannot be too large, and that only a few coordinates can have significant low-order influence in a pure-valued function.

PROPOSITION 2.12. *For every pure-valued function f in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ and any $k \leq n$, the sum of the low-order influence of each coordinate in f is bounded above by*

$$\sum_{i \in [n]} \text{Inf}_f^{\leq k}(i) \leq k .$$

COROLLARY 2.13. *For every pure-valued function f in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$, any $k \leq n$, and any $\theta > 0$,*

$$\left| \left\{ i \in [n] : \text{Inf}_f^{\leq k}(i) \geq \theta \right\} \right| \leq \frac{k}{\theta} .$$

3. MAIN RESULT

3.1 The algorithm

The JUNTA TEST algorithm is based on a simple but useful observation of Blum, Hellerstein, and Littlestone [5]: if we have two inputs $x, y \in \mathcal{X}$ such that $f(x) \neq f(y)$, then the set S of coordinates in which x and y disagree contains a coordinate that is relevant in f . Furthermore, by performing a binary search over the hybrid inputs formed from x and y , we can identify the relevant coordinate with $O(\log |S|)$ queries.

We build on this observation by noting that if we have a partition of the coordinates into s parts and only care to

JUNTATEST(f, k, ϵ)
Additional parameters: $s = 10^{20}k^9/\epsilon^5$, $r = 12(k+1)/\epsilon$

1. Randomly partition the coordinates in $[n]$ into s sets I_1, \dots, I_s .
2. Initialize $S \leftarrow [n]$, $\ell \leftarrow 0$.
3. For each of r rounds,
 - 3.1. Generate a pair $(x, y) \in \Omega \times \Omega$.
 - 3.2. If $f(x) \neq f(y_S x_{\bar{S}})$, then
 - 3.2.1. Use binary search to find a set I_j that contains a relevant variable.
 - 3.2.2. Update $S \leftarrow S \setminus I_j$.
 - 3.2.3. Set $\ell \leftarrow \ell + 1$.
 - 3.2.4. If $\ell > k$, then **reject** the function.
4. **Accept** the function.

Figure 1: The algorithm for ϵ -testing k -juntas.

identify a part that contains a relevant coordinate (rather than the coordinate itself), then we can optimize the binary search to only take $O(\log s)$ queries.

The JUNTATEST algorithm applies the above observation in the obvious way. It maintains a set S of coordinates that may or may not be relevant to the function, generates pairs of inputs $x, y \in \mathcal{X}$ at random, and checks if $f(x) \neq f(x_{\bar{S}}y_S)$. When such a pair is found, the algorithm identifies a part that contains a relevant coordinate, and removes all the coordinates in that part from S . If the algorithm identifies $k+1$ different parts with relevant coordinates, it rejects the function; otherwise, it accepts the function. The details of the algorithm are presented in Figure 1.

3.2 Main lemma

To establish the correctness of the JUNTATEST algorithm, we want to show that the algorithm does not lose too much accuracy by identifying parts that contain relevant coordinates instead of identifying the relevant coordinates individually.

In other words, we want to show that under a random partition \mathcal{I} , (1) a function that is a k -junta is also a k -part junta with respect to \mathcal{I} , and (2) a function that is ϵ -far from being a k -junta is $\Theta(\epsilon)$ -far from being a k -part junta with respect to \mathcal{I} . The former statement is clearly always true; the following lemma shows that the latter statement holds with high probability.

LEMMA 3.1. *Let \mathcal{I} be a random partition of $[n]$ with $s = 10^{20}k^9/\epsilon^5$ parts obtained by uniformly and independently assigning each coordinate to a part. With probability at least $5/6$, a function $f: \mathcal{X} \rightarrow \mathcal{Y}$ that is ϵ -far from being a k -junta is also $\frac{\epsilon}{2}$ -far from being a k -part junta with respect to \mathcal{I} .*

The proof of Lemma 3.1 is presented in Section 4. Before proceeding with that proof, we first show how the rest of the proof of Theorem 1.1 is constructed.

3.3 Proof of Theorem 1.1

We prove Theorem 1.1 by showing that the JUNTATEST algorithm ϵ -tests k -juntas with only $O(k/\epsilon + k \log k)$ queries.

THEOREM 1.1 (RESTATED). *The number of queries required to ϵ -test k -juntas is bounded above by $O(k/\epsilon + k \log k)$. Furthermore, this result holds for testing functions that have arbitrary finite product domains and arbitrary finite ranges,*

and it also holds under any product distribution over the domain.

PROOF. We begin by determining the query complexity of the JUNTATEST algorithm. At most $2r = 24(k+1)/\epsilon$ queries are made in the execution of line 3.2 of the algorithm, and at most $(k+1) \log s = O(k \log(k/\epsilon))$ queries are made in line 3.2.1 of the algorithm. So the algorithm makes a total of $O(k/\epsilon + k \log k)$ queries to the input function.

The completeness of the JUNTATEST algorithm is easy to establish: when the input function is a k -junta, it contains at most k parts with relevant coordinates, so the algorithm must accept the function. Therefore, the JUNTATEST algorithm has one-sided error.

Finally, we analyze the soundness of the JUNTATEST algorithm. By Lemma 3.1, with probability at least $5/6$ a function f that is ϵ -far from being a k -junta is also $\epsilon/2$ -far from being a k -part junta with respect to the random partition of the coordinates. When this is the case, the influence of S is at least $\epsilon/2$ until $k+1$ parts with relevant coordinates are identified. So the expected number of rounds required to identify $k+1$ parts with relevant variables is $2(k+1)/\epsilon$. By Markov's Inequality, the probability that the algorithm does not identify $k+1$ relevant parts in $12(k+1)/\epsilon$ rounds is at most $1/6$, and the overall probability that the JUNTATEST algorithm fails to reject f is at most $1/3$. \square

4. PROOF OF THE MAIN LEMMA

4.1 Overview of the proof

To establish Lemma 3.1, we want to show that with high probability every set J formed by taking the union of k parts in a random partition of the coordinates satisfies

$$\sum_{S \subseteq J} \|f^S\|_2^2 \leq 1 - \epsilon. \quad (1)$$

We show this with a combination of three arguments.

First, In Section 4.2, we examine the Efron-Stein coefficients $\|f^S\|_2^2$ for sets of size $|S| > 2k$. Under a random partition of the coordinates, most of these sets have elements distributed over more than k parts. Therefore, with high probability the contribution of those sets to the sum in (1) is small.

We then examine the coordinates with large low-order influence. In Section 4.3, we show that for a sufficiently fine

random partition, with high probability those coordinates are completely separated by the partition and therefore provide a limited contribution to the sum in (1).

Lastly, we examine the coordinates with small low-order influence. In Section 4.4, we use a Hoeffding bound argument to show that their contribution to the sum in (1) is also negligible.

We combine the above arguments to complete the proof of Lemma 3.1 in Section 4.5.

4.2 High-order coefficients

DEFINITION 4.1 (COVERED SETS). *Let $\mathcal{I} = \{I_1, \dots, I_s\}$ be a partition of $[n]$. For any subset $S \subseteq [n]$, we say that S is k -covered by the partition \mathcal{I} , denoted by $S \preceq_k \mathcal{I}$, if there exist k indices i_1, \dots, i_k such that $S \subseteq I_{i_1} \cup \dots \cup I_{i_k}$.*

PROPOSITION 4.2. *Let $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ be a pure-valued function, let $s \geq 72ek/\epsilon$, and let $\mathcal{I} = \{I_1, \dots, I_s\}$ be a random partition of $[n]$. Then with probability at least $17/18$, the high-level influence of f contained in sets that are k -covered by \mathcal{I} is at most*

$$\sum_{S \preceq_k \mathcal{I} : |S| > 2k} \|f^S\|_2^2 \leq \epsilon/4.$$

PROOF. Let $S \subseteq [n]$ be a subset of size $|S| > 2k$. The probability that all the elements in S are sent to k or fewer parts in \mathcal{I} is

$$\begin{aligned} \Pr[S \preceq_k \mathcal{I}] &\leq \binom{s}{k} \left(\frac{k}{s}\right)^{2k+1} \\ &\leq \left(\frac{es}{k}\right)^k \left(\frac{k}{s}\right)^{2k+1} = e^k \left(\frac{k}{s}\right)^{k+1} \leq \frac{\epsilon}{72}. \end{aligned}$$

The expected weight of all large sets that are covered by \mathcal{I} is

$$\mathbf{E} \left[\sum_{S \preceq_k \mathcal{I} : |S| > 2k} \|f^S\|_2^2 \right] = \sum_{|S| > 2k} \|f^S\|_2^2 \cdot \Pr[S \preceq_k \mathcal{I}] \leq \frac{\epsilon}{72},$$

where the last inequality uses the above upper bound on $\Pr[S \preceq_k \mathcal{I}]$ and Parseval's identity. The proposition then follows from Markov's Inequality. \square

4.3 Coords. with large low-order influence

For any pure-valued function f in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$, let us define

$$H_f \stackrel{\text{def}}{=} \{i \in [n] : \text{Inf}_f^{\leq 2k}(i) \geq \theta\}.$$

to be the set of coordinates with large low-order influence in f .

With high probability, a random partition of the coordinates completely separates the set H_f .

PROPOSITION 4.3. *Let $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ be a pure-valued function, let $\theta > 0$, and let $s \geq 72k^2/\theta^2$. Then with probability at least $17/18$, a random partition $\mathcal{I} = \{I_1, \dots, I_s\}$ satisfies the condition that for all $i \in [s]$,*

$$|H_f \cap I_i| \leq 1.$$

PROOF. By Corollary 2.13, $|H_f| \leq 2k/\theta$. So the probability that there exists a part I_i that contains at least 2 elements from H_f is

$$\binom{|H_f|}{2} \binom{s}{1} (1/s)^2 \leq \left(\frac{2k}{\theta}\right)^2 \cdot \frac{1}{s} \leq \frac{1}{18}.$$

\square

4.4 Coords. with small low-order influence

Since every coordinate in $[n] \setminus H_f$ has small low-order influence, we can expect these coordinates to have little impact on the total low-order influence of each part. Indeed, this is what the next proposition shows.

PROPOSITION 4.4. *Let $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ be a pure-valued function, let $s \geq 16k^2/\epsilon$, let $\theta \leq \epsilon^2/64k^3 \log(18s)$, and let $\mathcal{I} = \{I_1, \dots, I_s\}$ be a random partition of $[n]$. Then with probability at least $17/18$,*

$$\text{Inf}_f^{\leq 2k}(I_i \setminus H_f) \leq \frac{\epsilon}{4k}$$

for every $i \in [s]$.

PROOF. Fix $i \in [s]$. For every $j \in [n]$, define X_j to be a random variable that takes the value

$$X_j = \begin{cases} \text{Inf}_f^{\leq 2k}(j) & , \text{ if } j \in I_i \setminus H_f, \\ 0 & , \text{ otherwise.} \end{cases}$$

By the subadditivity of influence,

$$\text{Inf}_f^{\leq 2k}(I_i \setminus H_f) \leq \sum_{j \in I_i \setminus H_f} \text{Inf}_f^{\leq 2k}(j) = \sum_{j \in [n]} X_j.$$

By Proposition 2.12, $\sum_{j \in [n]} \text{Inf}_f^{\leq 2k}(j) \leq 2k$. Furthermore, $\Pr[j \in I_i] = 1/s$ for every $j \in [n]$. So

$$\mathbf{E} \left[\sum_{j \in [n]} X_j \right] = \sum_{j \in [n] \setminus H_f} \text{Inf}_f^{\leq 2k}(j) \cdot \Pr[j \in I_i] \leq \frac{2k}{s}.$$

By our choice of s , we have that $2k/s \leq \epsilon/8k$. We can apply Hoeffding's inequality to obtain

$$\Pr \left[\sum_{j \in [n]} X_j \geq \frac{\epsilon}{8k} + t \right] \leq \exp \left(-\frac{2t^2}{\sum_{j \in [n] \setminus H_f} \text{Inf}_f^{\leq 2k}(j)^2} \right).$$

Applying the elementary inequality $\sum_i x_i^2 \leq \max_i x_i \cdot \sum_i x_i$ to the summation on the right-hand side of the equation and recalling that $\max_{j \in [n] \setminus H_f} \text{Inf}_f^{\leq 2k}(j) < \theta$, we get that

$$\sum_{j \in [n] \setminus H_f} \text{Inf}_f^{\leq 2k}(j)^2 \leq 2k\theta.$$

So

$$\begin{aligned} \Pr \left[\text{Inf}_f^{\leq 2k}(I_i \setminus H_f) > \frac{\epsilon}{4k} \right] &\leq \Pr \left[\sum_{j \in [n]} X_j \geq \frac{\epsilon}{8k} + \frac{\epsilon}{8k} \right] \\ &\leq e^{-\log(18s)} = \frac{1}{18s}. \end{aligned}$$

Applying the union bound over all $i \in [s]$ completes the proof of the proposition. \square

4.5 Proof of Lemma 3.1

LEMMA 3.1 (RESTATED). *Let \mathcal{I} be a random partition of $[n]$ with $s = 10^{20}k^9/\epsilon^5$ parts obtained by uniformly and independently assigning each coordinate to a part. With probability at least $5/6$, a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that is ϵ -far from being a k -junta is also $\frac{\epsilon}{2}$ -far from being a k -part junta with respect to \mathcal{I} .*

PROOF. Let J be the union of any k parts in \mathcal{I} . By Proposition 2.9,

$$\text{Inf}_f([n] \setminus J) = \sum_{S \subseteq [n]} \|f^S\|_2^2 - \sum_{S \subseteq J} \|f^S\|_2^2.$$

Let $\theta = \epsilon^2 \log(k/\epsilon)/10^9 k^4$. Note that the values of s and θ satisfy the conditions of Propositions 4.2–4.4. Define the following families:

$$\begin{aligned} \mathcal{H} &= \{S \subseteq J : |S| \leq 2k, S \subseteq J \cap H_f\}, \\ \mathcal{L} &= \{S \subseteq J : |S| \leq 2k, S \not\subseteq J \cap H_f\}, \text{ and} \\ \mathcal{B} &= \{S \subseteq J : |S| > 2k\} \end{aligned}$$

The sets $\mathcal{H}, \mathcal{L}, \mathcal{B}$ form a partition of the subsets of J , so

$$\begin{aligned} \text{Inf}_f([n] \setminus J) &= \sum_{S \subseteq [n]} \|f^S\|_2^2 - \sum_{S \in \mathcal{H}} \|f^S\|_2^2 \\ &\quad - \sum_{S \in \mathcal{L}} \|f^S\|_2^2 - \sum_{S \in \mathcal{B}} \|f^S\|_2^2 \\ &\geq \text{Inf}_f([n] \setminus (J \cap H_f)) \\ &\quad - \sum_{S \in \mathcal{L}} \|f^S\|_2^2 - \sum_{S \in \mathcal{B}} \|f^S\|_2^2. \end{aligned}$$

As Proposition 4.3 shows, with probability at least $17/18$, H_f is completely separated by the partition \mathcal{I} , so every set J formed by taking the union of at most k parts satisfies $|J \cap H_f| \leq k$. Then, since f is ϵ -far from being a k -junta,

$$\text{Inf}_f([n] \setminus (J \cap H_f)) \geq \epsilon. \quad (2)$$

By Proposition 4.4, with probability at least $17/18$, every set I_j in the partition satisfies $\sum_{i \in I_j \setminus H_f} \text{Inf}_f^{\leq 2k}(i) \leq \epsilon/4k$, so every set J formed by taking the union of k parts satisfies

$$\sum_{S \in \mathcal{L}} \|f^S\|_2^2 = \text{Inf}_f^{\leq 2k}(J \setminus H_f) \leq k \cdot \frac{\epsilon}{4k} = \frac{\epsilon}{4}. \quad (3)$$

Finally, the family \mathcal{B} contains exclusively sets S that are k -covered by \mathcal{I} , so $\mathcal{B} \subseteq \{S \subseteq [n] : |S| > 2k, S \preceq_k \mathcal{I}\}$ and we can apply Proposition 4.2 to obtain that with probability at least $17/18$, for every set J

$$\sum_{S \in \mathcal{B}} \|f^S\|_2^2 \leq \sum_{S \preceq_k \mathcal{I} : |S| > 2k} \|f^S\|_2^2 \leq \frac{\epsilon}{4}. \quad (4)$$

Combining equations (1)–(4), we obtain that with probability at least $1 - 3(1/18) = 5/6$, for every set J obtained by taking the union of at most k parts of \mathcal{I}

$$\text{Inf}_f([n] \setminus J) \geq \epsilon - \frac{\epsilon}{4} - \frac{\epsilon}{4} = \frac{\epsilon}{2}.$$

□

5. ACKNOWLEDGMENTS.

The author wishes to thank Ryan O’Donnell for much invaluable advice throughout the course of this research.

The author also thanks Kevin Matulef, Krzysztof Onak, and the anonymous referees for many helpful comments on an earlier draft of this article.

6. REFERENCES

- [1] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. In *Proc. 23rd Conf. on Computational Complexity*, 2008.
- [2] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, PCPs and non-approximability – towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998.
- [3] Eric Blais. Improved bounds for testing juntas. In *Proc. 12th Workshop RANDOM*, pages 317–330, 2008.
- [4] Eric Blais, Ryan O’Donnell, and Karl Wimmer. Polynomial regression under arbitrary product distributions. In *Proc. 21st Conf. on Learning Theory*, pages 193–204, 2008.
- [5] Avrim Blum, Lisa Hellerstein, and Nick Littlestone. Learning in the presence of finitely or infinitely many irrelevant attributes. *J. of Comp. Syst. Sci.*, 50(1):32–40, 1995.
- [6] Hana Chockler and Dan Gutfreund. A lower bound for testing juntas. *Information Processing Letters*, 90(6):301–305, 2004.
- [7] Ilias Diakonikolas, Homin K. Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A. Servedio, and Andrew Wan. Testing for concise representations. In *Proc. 48th Symposium on Foundations of Computer Science*, pages 549–558, 2007.
- [8] Brad Efron and Charles Stein. The jackknife estimate of variance. *Ann. of Stat.*, 9(3):586–596, 1981.
- [9] Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *J. Comput. Syst. Sci.*, 68(4):753–787, 2004.
- [10] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. of the ACM*, 45(4):653–750, 1998.
- [11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.
- [12] Timothy R. Hugues *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnology*, 19(4):342–347, 2001.
- [13] Samuel Karlin and Yosef Rinott. Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jack-knife estimates. *Ann. of Statistics*, 10(2):485–501, 1982.
- [14] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.
- [15] Elchanan Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49th Symp. on Foundations of Computer Science*, 2008.
- [16] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low

influences: invariance and optimality. In *Proc. 46th Symp. Foundations of Comp. Sci.*, pages 21–30, 2005.

- [17] Michal Parnas, Dana Ron, and Alex Samorodnitsky. Testing basic boolean formulae. *SIAM J. Discret. Math.*, 16(1):20–46, 2003.
- [18] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.
- [19] J. Michael Steele. An Efron-Stein inequality for non-symmetric statistics. *Ann. of Statistics*, 14(2):753–758, 1986.

APPENDIX

A. ADDITIONAL PROOFS

A.1 Proof of Proposition 2.9

It is well-known that the identity

$$\text{Inf}_f(S) = \sum_{T: S \cap T \neq \emptyset} \|f^T\|_2^2$$

holds for functions in the space $L^2(\Omega, \mathbb{R})$ (see for example [15, 16, 19, 9]). Below, we show that the same identity holds for pure-valued functions in $L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$. The proof itself is a simple generalization of known proofs of the above identity; we include it for the convenience of the reader.

PROPOSITION 2.9 (RESTATED). *For any pure-valued function $f \in L^2(\Omega, \mathbb{R}^{\mathcal{Y}})$ and set $S \subseteq [n]$,*

$$\text{Inf}_f(S) = \sum_{T: S \cap T \neq \emptyset} \|f^T\|_2^2.$$

PROOF. Since f is pure-valued, the indicator function $\mathbf{1}[f(x) = f(y)]$ is equal to the inner product $\langle f(x), f(y) \rangle_{\mathbb{R}^{\mathcal{Y}}}$, and so

$$\begin{aligned} \text{Inf}_f(S) &= \Pr_{x, y \sim \Omega} [f(x) \neq f(x_{\bar{S}} y_S)] \\ &= 1 - \mathbf{E}_{x, y \sim \Omega} [\langle f(x), f(x_{\bar{S}} y_S) \rangle_{\mathbb{R}^{\mathcal{Y}}}] . \end{aligned}$$

Taking the Efron-Stein decomposition of f and applying linearity of expectation, we get

$$\text{Inf}_f(S) = 1 - \sum_{T, U \subseteq [n]} \mathbf{E}_{x \sim \Omega} \left[\left\langle f^T(x), \mathbf{E}_{y \sim \Omega} [f^U(y) \mid y_{\bar{S}} = x_{\bar{S}}] \right\rangle_{\mathbb{R}^{\mathcal{Y}}} \right].$$

By the definition of the Efron-Stein decomposition,

$$\mathbf{E}_{y \sim \Omega} [f^U(y) \mid y_{\bar{S}} = x_{\bar{S}}] = \begin{cases} f^U(x) & , \text{ if } U \subseteq \bar{S}, \\ 0 & , \text{ otherwise.} \end{cases}$$

So

$$\begin{aligned} \text{Inf}_f(S) &= 1 - \sum_{T \subseteq [n]} \sum_{U \subseteq \bar{S}} \mathbf{E}_{x \sim \Omega} \left[\left\langle f^T(x), f^U(x) \right\rangle_{\mathbb{R}^{\mathcal{Y}}} \right] \\ &= 1 - \sum_{T \subseteq [n]} \sum_{U \subseteq \bar{S}} \left\langle f^T, f^U \right\rangle . \end{aligned}$$

By the orthogonality of the Efron-Stein decomposition, when $T \neq U$, $\langle f^T, f^U \rangle = 0$. So

$$\sum_{T \subseteq [n]} \sum_{U \subseteq \bar{S}} \left\langle f^T, f^U \right\rangle = \sum_{T \subseteq \bar{S}} \left\langle f^T, f^T \right\rangle = \sum_{T \subseteq \bar{S}} \|f^T\|_2^2.$$

By Parseval’s identity, we also have that $1 = \sum_{T \subseteq [n]} \|f^T\|_2^2$, so

$$\text{Inf}_f(S) = \sum_{T \subseteq [n]} \|f^T\|_2^2 - \sum_{T \subseteq \bar{S}} \|f^T\|_2^2 = \sum_{T: S \cap T \neq \emptyset} \|f^T\|_2^2.$$

□

A.2 Proof of Proposition 2.3

Fischer *et al.* [9] showed that in functions with boolean ranges that are far from being juntas on a set J of coordinates, the set $[n] \setminus J$ of coordinates has a significant amount of influence. A similar result was established by Diaconikolas *et al.* [7] for functions with non-boolean ranges, when a different notion of influence (“binary variation”) is considered.

We use Hölder’s Inequality to establish the analogous result with our notion of influence.

PROPOSITION 2.3 (RESTATED). *If $f : \mathcal{X} \rightarrow \mathcal{Y}$ is ϵ -far from being a k -junta, then for every set $J \subseteq [n]$ of size $|J| \leq k$,*

$$\text{Inf}_f([n] \setminus J) \geq \epsilon.$$

PROOF. For a given set J of size $|J| \leq k$, let $h : \mathcal{X} \rightarrow \mathcal{Y}$ be the function defined by

$$h(x) = \underset{y \in \mathcal{Y}}{\text{argmax}} \left\{ \Pr_z [f(x_J z_{\bar{J}}) = y] \right\},$$

where we break ties arbitrarily. Then,

$$\begin{aligned} \Pr_x [f(x) \neq h(x)] &= 1 - \mathbf{E}_x [\langle f(x), h(x) \rangle_{\mathbb{R}^{\mathcal{Y}}}] \\ &= 1 - \mathbf{E}_x \left[\left\langle \mathbf{E}_z [f(x_J z_{\bar{J}})], h(x) \right\rangle_{\mathbb{R}^{\mathcal{Y}}} \right] \\ &= 1 - \mathbf{E}_x \left[\left\| \mathbf{E}_z [f(x_J z_{\bar{J}})] \right\|_{\infty} \left\| \mathbf{E}_z [f(x_J z_{\bar{J}})] \right\|_1 \right] \\ &\leq 1 - \mathbf{E}_x \left[\left\| \mathbf{E}_z [f(x_J z_{\bar{J}})] \right\|_2^2 \right] \\ &= 1 - \sum_{S \subseteq J} \|f^S\|_2^2 \\ &= \sum_{S: S \cap ([n] \setminus J) \neq \emptyset} \|f^S\|_2^2 \end{aligned}$$

The first equality follows from the fact that f and h are pure-valued functions. The second follows from the fact that h only depends on the coordinates in J and from the linearity of expectation. The third equality uses the two easily-verified identities $\|\mathbf{E}_z [f(x_J z_{\bar{J}})]\|_{\infty} = \langle \mathbf{E}_z [f(x_J z_{\bar{J}})], h(x) \rangle_{\mathbb{R}^{\mathcal{Y}}}$ and $\|\mathbf{E}_z [f(x_J z_{\bar{J}})]\|_1 = 1$. The inequality is a special case of Hölder’s Inequality. The penultimate equality follows from the fact that $\mathbf{E}_z [f(x_J z_{\bar{J}})] = \sum_{S \subseteq J} f^S(x)$, and, finally the last equality follows from Parseval’s Theorem.

The proposition then follows from Proposition 2.9. □