Junto-symmetric functions, hypergraph isomorphism, and crunching

Sourav Chakraborty Eldar Fischer David García-Soriano Arie Matsliah

Abstract

We make a step towards characterizing the boolean functions to which isomorphism can be efficiently tested. Specifically, we prove that isomorphism to any boolean function on $\{0,1\}^n$ with a polynomial number of distinct permutations can be tested with a number of queries that is independent of n. We also show some partial results in the converse direction, and discuss related problems: testing isomorphism up to linear transformations, and testing isomorphism against a uniform (hyper)graph that is given in advance. Our results regarding the latter topic generalizes a theorem of Fischer (SICOMP 2005), and in the process we also provide a simpler proof of his original result which avoids the use of Szemerédi's regularity lemma.

1 Overview

We continue the study of property testing of boolean function isomorphism, initiated by Fischer et. al [FKR+04] and continued in the works of [AB10, BO10, CGM11b] (see also the references therein). Two boolean functions $f, g : \{0,1\}^n \to \{0,1\}$ are said to be *isomorphic* if they are equal up to relabelling of the input variables, i.e. if it is possible to permute the n input variables of f so that the resulting function is equal to g. For in-depth explanations of the motivation and the state of the art of the problem, refer to the papers cited above. Here we briefly comment that one of the reasons to study isomorphism between functions is that two functions being isomorphic means that they are "essentially the same" and have identical realizations. Also, many functional properties can be reduced to the problem of testing isomorphism or some of its variants [DLM+07]; c.f. [CGM11a]. Finally, as discussed below, this is a natural generalization to hypergraphs of the task of testing isomorphism between graphs, which is well understood [Fis05, FM08].

For reasons of space, some of the material has been moved into the Appendix. Most of our notation is standard; refer to Appendix A.1 for details.

1.1 Function isomorphism and the size of invariance groups

For a boolean function f and a permutation $\pi \in S_n$, denote by f^π the function obtained from f by permuting its inputs according to π. The *automorphism group* of f, also known as its *symmetry group* or *invariance group*, is the group of permutations that leave f invariant:

$$\mathrm{Aut(f)} \triangleq \{\pi \in \mathrm{S_n} \mid \mathrm{f}^\pi = \mathrm{f}\}.$$

Clearly Aut(f) is a subgroup of the symmetric group $S_n = Sym([n])$. Define an equivalence relation between permutations by $\pi \sim \sigma$ iff $f^\pi = f^\sigma$, and let

$$\mathrm{Isom(f)} = \{[\pi_1], \ldots, [\pi_\mathrm{t}]\}$$

be the equivalence classes formed. There is a bijection between Isom(f) and the set $S_n/\mathrm{Aut(f)}$ of cosets of Aut(f); therefore the number |Isom(f)| of distinct permutations of f is equal to the index of Aut(f) in S_n, i.e. $|\mathrm{Isom(f)}| = |\mathrm{S_n}/\mathrm{Aut(f)}| = \mathrm{n!}/|\mathrm{Aut(f)}|$. The size of Aut(f) is a rough measure of the amount of symmetry that f possesses: the larger Aut(f), the more symmetric f is. A symmetric function satisfies $\mathrm{Aut(f)} = \mathrm{S_n}$ and $|\mathrm{Isom(f)}| = 1$, whereas a random function has, with high probability, a trivial automorphism group $\mathrm{Aut(f)} = \{1\}$ and $|\mathrm{Isom(f)}| = \mathrm{n!}$.

Not every group $G \leq S_n$ can arise as the automorphism group of a boolean function on n variables; those that can be are called *2-representable*. For example, it is not hard to argue that if the alternating group A_n ($n \geq 3$) is contained in Aut(f), then Aut(f) is indeed the whole of S_n; as a result, A_n is not 2-representable. Indeed, take any $x, y \in \{0,1\}^n$ with $|x| = |y|$. Then there is a permutation $\pi \in S_n$ mapping x to y; if $n \geq 3$ then π can be assumed to be an even permutation by performing, if necessary, one additional swap between two distinct indices i, j with $y_i = y_j$. Then $\pi \in \mathrm{Aut(f)}$ and so $f(x) = f(y)$. Hence $A_n \leq \mathrm{Aut(f)}$ implies $f(x) = f(y)$ for all $|x| = |y|$, so f is actually symmetric. Groups $G \leq S_n$ that can be represented as Aut(f) for some k-valued function $f : \{0,1\}^n \to [k]$ are called k-representable; they are studied in [CK91, Kis98] (see also Chapter 3 of [CK02]).

We know that f-isomorphism can always be tested with $O(\log |\mathrm{Isom(f)}|)$ queries for constant ϵ [BO10, CGM11b], so symmetric functions are particularly easy to test isomorphism to (the query

complexity becomes constant; in fact the problem reduces to testing equality in this case). What is the smallest size Isom(f) can have for a non-symmetric function? A moment's thought reveals that there are non-symmetric functions with only n different permutations, like any dictatorship $f(x_1x_2\dots x_n) = x_i$, and indeed this can be shown to be best possible.

Even though the number of queries made by the trivial isomorphism tester is superconstant for a non-symmetric function, it is also possible to test isomorphism to dictatorships with $O(1)$ queries [PRS02], and more generally to $O(1)$-juntas [FKR$^+$04]. However these two classes do not encompass all known easy-to-test functions. For example, consider the parity function on the first $n-t$ variables out of n, $\chi_{[n-t]}$[1]. The identity $\chi_{[n-t]}(x) = \chi_{[n]}(x) \oplus \chi_{[n]\setminus[n-t]}(x)$ makes it possible to transform the response to every query made for the t-junta $\chi_{[n]\setminus[n-t]}$ into the response of a query for $\chi_{[n-t]}$. This transformation provides a reduction between the two testing problems. In particular, for constant t we can test isomorphism to $(n-t)$-parities with $O_t(1)$ queries. In the same vein, the majority function on the first $n-t$ variables $\mathrm{Maj}_{[n-t]}$ (for constant t and n large enough) is very close to the symmetric majority $\mathrm{Maj}_{[n]}$, and it is not hard to see that the constant-query test for equality between the tested function and $\mathrm{Maj}_{[n]}$ yields a tester for isomorphism to $\mathrm{Maj}_{[n]-t}$ as well.

We introduce a notion generalizing all these cases.

Definition 1.1 (Junto-Symmetric) *Let $J \subseteq [n]$. A function $f : \{0,1\}^n \to \{0,1\}$ is called* J-junto-symmetric *if it can be written in the form*

$$f(x) = \hat{f}(|x|, x\restriction_J)$$

for some $\hat{f} : \{0,\dots,n\} \times \{0,1\}^{|J|} \to \{0,1\}$. Equivalently, this means that the restriction of f to any constant-weight layer of the cube is a junta on J.

The function f is called k-junto-symmetric *if it is J-junto-symmetric on some subset J of size k.*

Let $\mathcal{JS}_J$ denote the class of J-junto-symmetric functions, and $\mathcal{JS}_k$ the k-junto-symmetric functions. Note that the definition requires the junta variables be the same on every layer, but the junta function is allowed to vary. Also variables outside J can have noticeable influence on a J-junto-symmetric function f.

Observe that any symmetric function is 0-junto-symmetric, and any k-junta is k-junto-symmetric. At the other extreme, any function at all is $(n-1)$-junto-symmetric. Additional examples of k-junto-symmetric functions are $\chi_{[n-k]}$ and $\mathrm{Maj}_{[n-k]}$; in fact, the reader may verify that any k-junta whose core function is symmetric must be $\min(k, n-k)$-junto-symmetric (see Appendix A.1 for the definition of a junta's *core*).

Definition 1.2 *Let $\mathcal{F}$ denote a sequence $f_1, f_2, \dots$ of boolean functions with $f_n : \{0,1\}^n \to \{0,1\}$ for each $n \in \mathbb{N}^+$.*

We say that $\mathcal{F}$ is an $O(1)$-junto-symmetric family *if there exists a constant k such that each f_i is k-junto-symmetric.*

We will occasionally speak of such a family as an $O(1)$-junto-symmetric (or poly-symmetric) function when the intended meaning is clear.

The size of Isom(f) for any k-junto-symmetric f is bounded by $\binom{n}{k}k!$, which is $n^{O(1)}$ for constant k. Such families were given a name in [PS10]:

[1]The symbol χ is usually reserved to a parity taking values in ± 1 so it is a character of $\mathbb{Z}_2^n$, but here we use it for $\{0,1\}$-valued functions.

Definition 1.3 *The family $\mathcal{F}$ is* poly-symmetric *if there exists a constant c such that $|\mathrm{Isom}(\mathrm{f_n})| \leq \mathrm{n^c}$ for all n.*

As it turns out, these two notions are the same.

Theorem 1.4 *Let $\mathcal{F} = \{f_n : \{0,1\}^n \to \{0,1\}\}_{n\in\mathbb{N}}$. The following are equivalent:*

(a) $\mathcal{F}$ is a poly-symmetric family;

(b) There are sets $A_n \subseteq [n]$ of constant size such that $Sym([n] \setminus A_n) \leq \mathrm{Aut(f_n)}$ for all n;

(c) $\mathcal{F}$ is an $O(1)$-junto-symmetric family;

(d) Each f_n is a boolean combination of $O(1)$-many dictators and $O(1)$-many symmetric functions (with the same constants for all n).

The proof is given in Appendix C.

One of the main results of this paper is an extension of the junta tester and the isomorphism tester for juntas:

Theorem 1.5 *Let $\epsilon > 0$ and $1/\epsilon^{1/4} < k < n^{1/12}$. Let $f : \{0,1\}^n \to \{0,1\}$ and denote $f^* \in \mathcal{JS}_k$ the k-junto-symmetric function closest to f.*

There is a $\mathrm{poly}(k/\epsilon)$-query algorithm that takes ϵ, k and an oracle for f and satisfies:

- *If $\mathrm{dist}(f, f^*) \leq 1/k^5$, the algorithm accepts with probability $\geq 2/3$.*
- *If $\mathrm{dist}(f, f^*) \geq \epsilon$, the algorithm rejects with probability $\geq 2/3$.*

See Section 3.1 for the proof.

We can also obtain an $O(1)$-query algorithm for testing isomorphism to $O(1)$-junto-symmetric functions.

Theorem 1.6 *Let k, ϵ, f as before. There is a $\mathrm{poly}(k/\epsilon)$-query ϵ-tester for testing isomorphism between f and a known function $g : \{0,1\}^n \to \{0,1\}$ that is $1/k^5$-close to k-junto-symmetric, with constant success probability.*

The proof is in Appendix F.1.

Corollary 1.7 *Isomorphism to any poly-symmetric function can be ϵ-tested with $\mathrm{poly}(1/\epsilon)$ queries.*

With a view toward obtaining a possible classification, it is best to state *tolerant* versions of these results. This is possible at the expense of an exponential blowup in the query complexities (see Appendix F.2).

Theorem 1.8 *There is a constant $0 < c < 1$ with the following property. Let k, ϵ, f as before.*

There is an $\exp(k/\epsilon)$-query algorithm that, with high probability accepts if f is $(c\epsilon)$-close to $\mathcal{JS}_k$ and rejects if it is ϵ-far from $\mathcal{JS}_k$.

Similarly, there is an $\exp(k/\epsilon)$-query algorithm to test isomorphism to a function f that is $(c\epsilon)$-close to $\mathcal{JS}_k$.

1.2 Hypergraph isomorphism

The problem of testing graph isomorphism was first raised by Alon, Fischer, Krivelevich, and Szegedy [AFKS00] (see also [Fis01]), who used a lower bound on testing isomorphism of two unknown graphs to give an example of a non-testable first-order graph property of a certain type. Later, Fischer [Fis05] studied the problem of testing isomorphism to a given graph G and characterized the class of graphs to which isomorphism can be tested with a constant number of queries. He proved that the graphs to which isomorphism can be tested with a constant number of queries are precisely those that can be approximated by an *algebra* of constantly many cliques [Fis05]; i.e. can be obtained from those cliques by applying set intersection, union and complementation operations. Subsequently Fischer and Matsliah [FM08] gave bounds (tight in most settings) on the worst-case query complexity of testing isomorphism between two graphs.

It is possible to establish a connection between function isomorphism and a generalized form of graph isomorphism. Recall that an *undirected hypergraph* is a pair $H = (V, E)$, where V is a set of vertices and $E \subseteq \mathcal{P}(V)$ is a collection of hyperedges. Isomorphism between hypergraphs is defined in the natural way. elements of V rather than unordered sets.

Now define the distance between two hypergraphs $H = (V, E)$ and $H' = (V, E')$ on the same set of vertices by $\mathrm{dist}(H, H') = |E \oplus E'|/2^n$, where $E \oplus E'$ is the symmetric difference between their edge sets. Testing function isomorphism is easily seen to be equivalent to testing isomorphism between undirected hypergraphs under this distance measure (this is the "dense hypergraphs model"). Indeed, a boolean function $f : \{0,1\}^n \to \{0,1\}$ can be identified with the hypergraph with vertex set $V = [n]$ and edge set

$$f^{-1}(1) = \{x \in \{0,1\}^n \mid f(x) = 1\},$$

where binary vectors $x \in f^{-1}(1) \subseteq \{0,1\}^n$ are themselves identified with subsets of $[n]$ in the natural way. Clearly this satisfies $f \cong g \Leftrightarrow f^{-1}(1) \cong g^{-1}(1)$ as hypergraphs, and moreover the distance between f and g coincide from both viewpoints.

Seen this way, the problem of function isomorphism becomes a natural generalization of the analogous problem for graphs. This raises the question of whether progress towards the characterization can be made by studying hypergraph isomorphism in the line of previous works on graph isomorphism. One possible line of work is the study of uniform hypergraphs. An *r-uniform hypergraph* is one in which every edge $e \in E$ has size precisely r; the number r is also said to be the *arity* of the hypergraph. The distance between two r-uniform hypergraphs $H = (V, E), H' = (V, E')$ on the same vertex set of size $|V| = n$ is defined as $|E \oplus E'|/\binom{n}{r}$. Babai and Chakraborty [BC10] studied this question and obtained worst-case query-complexity bounds for the case of uniform hypergraphs. However, a characterization of the testability of isomorphism between uniform hypergraphs remained to be found.

In this work we prove an extension of Fischer's result that resolves the problem for hypergraphs of constant arity. To state it, recall that a *homomorphism* between $H = (V, E)$ and $\hat{H} = (\hat{V}, \hat{E})$ is a mapping $\Pi : V \to \hat{V}$ such that for all $\{v_1, \ldots, v_r\} \in V$, the implication $\{v_1, \ldots, v_r\} \in E \implies \{\Pi(v_1), \ldots, \Pi(v_r)\} \in \hat{E}$ holds. The homomorphism Π is called *full* (and H is said to be *fully homomorphic* to $\hat{H}$) if this implication holds in both directions, i.e. if

$$\{v_1, \ldots, v_r\} \in E \Leftrightarrow \{\Pi(v_1), \ldots, \Pi(v_r)\} \in \hat{E}$$

Definition 1.9 *The r-uniform hypergraph H is* k-crunchable *if it is fully homomorphic to an r-uniform hypergraph with k vertices.*

The crunching number *of H is the smallest k such that H is k-crunchable.*

The ϵ-approximate crunching number *of H, denoted $CrunchNum_\epsilon(H)$, is the smallest k such that H is ϵ-close to a k-crunchable r-uniform hypergraph.*

The ϵ-testing number *of H, denoted $TestNum_\epsilon(H)$, is the minimum q for which there exists an ϵ-testing algorithm with q queries for the property of being isomorphic to H.*

For graphs, having a constant crunching number is essentially the same as being in the algebra of constantly many cliques (or close to it).

In section 2 we prove the following.

Theorem 1.10 *For every $r \in \mathbb{N}$, $\epsilon > 0$ there exists a pair of functions $L_\epsilon(n)$ and $U_\epsilon(n)$, with $\lim_{n\to\infty} L_\epsilon(t) = \infty$, such that for every r-uniform hypergraph H we have*

$$L_\epsilon(CrunchNum_\epsilon(H)) \leq TestNum_\epsilon(H) \leq U_\epsilon(CrunchNum_{\epsilon/2}(H)).$$

The original proof of Fischer for (a statement equivalent to) the special case of Theorem 1.10 when $r = 2$ applied the Szeméredi regularity lemma for the *lower* bound (which is somewhat unusual as its normal use in property testing is to obtain upper bounds). Our simpler proof shows that this can be avoided. The lower bound method, which we call *crunching*, has additional applications, as outlined in the next subsection.

1.3 Remainder of the paper

In Appendix G we study what happens when we generalize our definition of k-junto-symmetric to all functions that are k-juntas when restricted to any constant-weight layer of the cube (we call them *layered juntas*), and show that in general these functions are no longer testable for isomorphism. The proof uses "crunching" ideas similar to those in the proof of the hypergraph lower bound.

We also consider a more general notions of isomorphism. In Appendix H we consider the problem of equivalence up to transformations by an arbitrary invertible linear map over $\mathbb{F}_2^n$ (note that isomorphism in the usual sense corresponds to the linear application defined by a permutation matrix). It is shown there that functions that are far from having constant Fourier dimension are hard to test for isomorphism.

2 Hypergraph crunching: lower bound

This section contains the proof of the lower bound of Theorem 1.10. For the upper bound, consult Appendix B. The functions L_ϵ and U_ϵ can be extracted from the proofs of the lower bound and the upper bound, respectively.

Definition 2.1 (Hypergraph crunching) *Let $\Pi : V \to V$ denote a mapping from V to itself. A* Π-crunch of H *is a hypergraph $H^\Pi_{cr} = (V, E')$ where*

$$E' = \{\{v_1, \ldots, v_k\} \mid \{\Pi(v_1), \ldots, \Pi(v_k)\} \in E\}.$$

A k-crunch *of a hypergraph is a Π-crunch for some Π with an image of size $\leq k$.*

Note that every k-crunch is a k-crunchable hypergraph (as witnessed by the same mapping Π). When Π is injective, a Π-crunching of H is a hypergraph isomorphic to H. For a hypergraph $H = (V, E)$ and $k \leq |V| = n$, we show that any tester will have a hard time distinguishing non-injective crunchings from permutations. A *random k-crunching* of H is a random hypergraph on V obtained as follows:

1. pick a subset $W \subseteq V$ of size k uniformly at random;
2. pick a mapping $\Pi : V \to W$ uniformly at random and output the Π-crunch of H.

Now define the distribution $\mathcal{D}_H^k$ by drawing a random permutation of a random k-crunching of H. Also write $\mathcal{D}_H$ for the uniform distribution over all permutations of H.

Lemma 2.2 *Let H be an r-uniform hypergraph and define $\mathcal{D}_H$ and $\mathcal{D}_H^k$ as before. Then it is impossible to distinguish a random $\tilde{H} \sim \mathcal{D}_H$ from a random $\tilde{H} \sim \mathcal{D}_H^k$ with $o(\sqrt{k}/r)$ queries.*

Proof. Let $q = o(\sqrt{k}/r)$ and $e_1, \ldots, e_q$ be the (adaptive, random) edge queries made. Let $Q \subseteq V$ be the set of at most rq vertices involved in these queries. Conditioned on the event $E_Q(\Pi)$ that Π is injective on Q, the distribution of replies to queries $e_1, \ldots, e_q$ is identical for $\mathcal{D}_H$ and $\mathcal{D}_H^k$. But $E_Q(\Pi)$ occurs except with probability at most $|Q|^2/k = o(1)$ as the choice of Π is independent of Q. This means that for any sequence $e_1, \ldots, e_q$ of queries and any sequence $a_1, \ldots, a_q$ of answers, the probability of obtaining answer a_i to query e_i for all i is, up to a factor of $\Pr[E_Q(\Pi)] = 1 - o(1)$, the same when $\tilde{H}$ is drawn from $\mathcal{D}_H$ as when it is drawn from $\mathcal{D}_H^k$. We conclude by Lemma A.1 that the tester cannot distinguish $\mathcal{D}_H$ from $\mathcal{D}_H^k$ with q queries and success probability $\geq 2/3$. □

Corollary 2.3 *If an r-uniform hypergraph is ϵ-far from being k-crunchable, then ϵ-testing isomorphism to it requires $\Omega(\sqrt{k}/r)$ queries.*

Together with the upper bound in the following subsection, this provides a characterization of hypergraphs of constant arity that can be tested for isomorphism with $O(1)$ queries. To see how this generalizes Fischer's result for graphs, Appendix D shows that being $O(1)$-chunchable is equivalent to having "algebra number" $O(1)$ as well.

3 Junto-symmetric functions

We present a reduction from testing the properties of being k-junto-symmetric, or being isomorphic to a given k-junto-symmetric function, to slight generalizations of the well-studied analogous problems for k-juntas. To this end we try to approximate the "junto-symmetric" components of the tested function f, i.e. the juntas determining the behaviour of f on each constant-weight layer of the boolean cube. However, each of these juntas is defined on a very small fraction of inputs; in order to define them on the whole of $\{0,1\}^n$ we attempt use a small "ballast" set $B \subseteq [n]$ of variables to enable us to balance weights as needed.

In the main body of the paper we only show how to test for the property of being junto-symmetric. Using similar ideas one can devise isomorphism tests; the details can be found in AppendixF.

3.1 Testing junto-symmetry

Let $\ell \in \mathcal{L} \triangleq \{0, 1, \ldots, n\}$ and $x \in \{0,1\}^n$. Write x^B for the string obtained from x by flipping the bits in $B \subseteq [n]$ and consider the set of minimal changes required to turn x into a string of weight ℓ:

$$\mathcal{B}_{\ell,x} \triangleq \{B \subseteq [n] \mid |x^B| = \ell \text{ and } |B| = |\ell - |x||\}.$$

It is to be observed that for any $B \in \mathcal{B}_{\ell,x}$, either $x^B \subseteq x$ or $x \subseteq x^B$ holds, depending on whether $|x| \geq \ell$ or $|x| \leq \ell$. The set $\mathcal{B}_{\ell,x}$ is always non-empty but consists of the single element 0^n when $\ell = |x|$.

Let $\mathcal{R}$ denote the set of all possible functions $r : \mathcal{L} \times \{0,1\}^n \to \{0,1\}^n$ with $r(\ell, x) \in \mathcal{B}_{\ell,x}$ for all ℓ, x. We need a lemma concerning the probability that $B = r(\ell, x)$ happens to intersect some small set A, when (ℓ, r, x) are drawn from the product distribution $\mu \triangleq \mathcal{L} \times \mathcal{R} \times \{0,1\}^n$. Here $\mathcal{L}$ is endowed with a binomial distribution $B(n, 1/2)$ and the uniform distribution is used in $\mathcal{R}$ and $\{0,1\}^n$.

Lemma 3.1 *Let $A \subseteq [n]$. Then*

$$\Pr_{\ell,x,B}[B \cap A \neq \emptyset] \leq \frac{|A|}{\sqrt{2n}}.$$

Proof. Observe that for any ℓ, the distribution of $B = r(\ell, x) \in \mathcal{B}_{\ell,x}$ over random x is symmetric under permutations, hence for all $i \in [n]$ we have

$$\Pr[i \in B] = \frac{1}{n} \sum_{j \in [n]} \Pr[j \in B] = \frac{1}{n} \mathbb{E}\left[|B|\right].$$

On the other hand, the size of any element B of $\mathcal{B}_{\ell,x}$ is $|\ell - |x||$ by definition. We can write $\ell = |y|$ for uniformly random $y \in \{0,1\}^n$, so $\mathbb{E}\left[|B|\right] = \mathbb{E}\left[||x| - |y||\right]$. Recalling that $\mathbb{E}\left[|x|\right] = \mathbb{E}\left[|y|\right] = n/2$, $\mathbb{E}\left[|x|^2\right] = \mathbb{E}\left[|y|^2\right] = \mathrm{Var}\left[|x|\right] + \mathbb{E}\left[|x|\right]^2 = \frac{1}{4}n(n+1)$ and applying Cauchy-Schwarz,

$$(\mathbb{E}\left[||x| - |y||\right])^2 \leq \mathbb{E}\left[(|x| - |y|)^2\right] = \mathbb{E}\left[|x|^2\right] + \mathbb{E}\left[|y|^2\right] - 2\mathbb{E}\left[|x|\right]\mathbb{E}\left[|y|\right] = \frac{n}{2}.$$

Hence $\mathbb{E}\left[||x| - |y||\right] \leq \sqrt{n/2}$ and $\Pr[i \in B] \leq \sqrt{\frac{1}{2n}}$, so

$$\Pr[B \cap A \neq \emptyset] \leq \sum_{i \in A} \Pr[i \in B] \leq \frac{|A|}{\sqrt{2n}}.$$

□

Let us define a transformation T mapping each function $f : \{0,1\}^n \to \{0,1\}$ to $T(f) : \mathcal{L} \times \mathcal{R} \times \{0,1\}^n \to \{0,1\}^n$ given by

$$T(f)(\ell, r, x) = f(x^{r(\ell,x)}).$$

Thus the parameter r acts as a "random seed" selecting, for each pair (ℓ, x), one string $x^{r(\ell,x)}$ of Hamming weight ℓ with minimum distance to x; the choice is independent of all choices for any other pair when r ranges uniformly over $\mathcal{R}$.

We denote the input parameter variables of $T(f)$ by V_0, V_1 and V_2, in order; we identify V_2 with $[n]$, the input variables of f. If $g : \mathcal{L} \times \mathcal{R} \times \{0,1\}^n$ is a junta on $V_0 \cup V_2$ (that is to say, $g(\ell, r, x)$ depends only on ℓ and x, but not on r), we define the function $\psi(g) : \{0,1\}^n \to \{0,1\}$ by

$$\psi(g)(x) = g(|x|, \bullet, x),$$

where the dot emphasizes that the assignment to the second parameter is immaterial by assumption.

We show that the task of testing junto-symmetry of f is closely related to that of testing g for being a junta, where distances are measured under μ. Let $\mathsf{Jun}_{V_0}(A) = \mathsf{Jun}(V_0 \cup A)$, and $\mathsf{Jun}_k(V_0) = \cup_{|A| \le k} \mathsf{Jun}_{V_0}(A)$.

In the next lemma, the variable symbols denote functions and sets of the following kind:

- $A \subseteq [n]$, $|A| = k$;
- f, g are arbitrary functions $\{0,1\}^n \to \{0,1\}$;
- $\jmath, \jmath_1, \jmath_2 : \{0,1\}^n \to \{0,1\}$ are junto-symmetric on A;
- $\jmath' : \mathcal{L} \times \mathcal{R} \times \{0,1\}^n \to \{0,1\}$ is a member of $\mathsf{Jun}_{V_0}(A)$;
- $\pi \in 1_{V_0,V_1} \times Sym(V_2)$ (we identify π with an element of $Sym(V_2)$ as well).

Lemma 3.2 *The mappings T and ψ satisfy the following properties:*

(a) T preserves distances: $\mathrm{dist}(f,g) = \mathrm{dist}(T(f), T(g))$ *for all f, g.*

(b) For any $\jmath' \in \mathsf{Jun}_{V_0}(A)$, we have $\psi(\jmath') \in \mathcal{JS}(A)$ and

$$\mathrm{dist}(\jmath', T(\psi(\jmath'))) \le \frac{|A|}{\sqrt{2n}}.$$

(c) For any $\jmath \in \mathcal{JS}(A)$, $T(\jmath)$ is $|A|/(\sqrt{2n})$-close to some $\jmath' \in \mathsf{Jun}_{V_0}(A)$. Moreover, we can take $\jmath'$ such that $\psi(\jmath') = \jmath$.

(d) $|\mathrm{dist}(f, \mathcal{JS}_k) - \mathrm{dist}(T(f), \mathsf{Jun}_k(V_0))| \le \frac{k}{\sqrt{2n}}$.

(e) ψ preserves permutations: for any π and $\jmath'$, $\psi(\jmath')^\pi = \psi(\jmath^\pi)$. Thus

$$\mathrm{distiso}(\jmath_1, \jmath_2) = \mathrm{distiso}(\psi(\jmath_1), \psi(\jmath_2)).$$

(f) The bounds

$$|\mathrm{distiso}(f,g) - d| \le \mathrm{dist}(f, \mathcal{JS}_k) + \mathrm{dist}(g, \mathcal{JS}_k) + \frac{2k}{\sqrt{2n}}.$$

hold for

$$d \triangleq \min_{\pi \in 1_{V_0,V_1} \times Sym(V_2)} \mathrm{dist}(T(f)^\pi, T(g)).$$

Proof.

(a) For any ℓ, the distribution of $x^{r(\ell,x)}$ for random x, r is uniform over all strings of weight ℓ. Since $\ell \sim B(n, 1/2)$ is distributed as the weight of a random element of $\{0,1\}^n$, it follows that the overall distribution of $x^{r(\ell,x)}$ is uniform, hence

$$\mathrm{dist}(T(f), T(g)) = \Pr[f(x^{r(\ell,x)}) \neq g(x^{r(\ell,x)})] = \Pr[f(x) \neq g(x)] = \mathrm{dist}(f, g).$$

(b) $\psi(\jmath')(x) = \jmath'(|x|, \bullet, x)$ is a function of $|x|$ and x_A, hence junto-symmetric on A. We have

$$\begin{aligned} \mathrm{dist}(\jmath', T(\psi(\jmath'))) &= \Pr[\jmath'(\ell, r, x) \neq \psi(\jmath')(x^{r(\ell,x)}) = \jmath'(\ell, \bullet, x^{r(\ell,x)})] \\ &\leq \Pr[r(\ell, x) \cap A \neq \emptyset] \\ &\leq \frac{|A|}{\sqrt{2n}} \end{aligned}$$

by Lemma 3.1.

(c) This follows from (b) because any $\jmath \in \mathcal{JS}(A)$ can be written in the form $\psi(\jmath')$ for some (in fact, many) $\jmath' \in \mathsf{Jun}_{V_0}(A)$.

(d) Let $\jmath$ be k-junto-symmetric and $\jmath' \in \mathsf{Jun}_k(V_0)$ with $\psi(\jmath') = \jmath$. Then by parts (c) and (a),

$$\mathrm{dist}(T(f), \jmath') \leq \mathrm{dist}(T(f), T(\jmath)) + \mathrm{dist}(T(\jmath), \jmath') \leq \mathrm{dist}(f, \jmath) + \frac{k}{\sqrt{2n}},$$

so $\mathrm{dist}(T(f), \mathsf{Jun}_k(V_0)) \leq \mathrm{dist}(f, \mathcal{JS}_k) + k/(2\sqrt{n})$. Likewise, if $\jmath'$ is a junta on $V_0 \cup A$ where $|A| = k$, then

$$\begin{aligned} \mathrm{dist}(f, \psi(\jmath')) &= \mathrm{dist}(T(f), T(\psi(\jmath'))) \\ &\leq \mathrm{dist}(T(f), \jmath') + \mathrm{dist}(\jmath', T(\psi(\jmath'))) \\ &\leq \mathrm{dist}(T(f), \jmath') + \frac{k}{\sqrt{2n}}, \end{aligned}$$

which proves the inequality $\mathrm{dist}(f, \mathcal{JS}_k) \leq \mathrm{dist}(T(f), \mathsf{Jun}_k(V_0)) + k/(2\sqrt{n})$.

(e) Clear.

(f) Follows from (d), (e) and the triangle inequality for distiso.

□

Now we describe a tester for the property $\mathsf{Jun}_k(V_0)$. Let $\mu = D_1 \times \ldots \times D_m$ be a product distribution, and $T \subseteq [m]$. (For our application we could take $D_1 = \mathcal{L}, D_2 = \mathcal{R}$, $T = \{1, 2\}$ and $D_3 \times \ldots \times D_m = \{0,1\}^n$). Choose a confidence parameter $0 \leq p < 1$ and distance parameter $0 < \epsilon \leq 1$. Let $f : \mu \to \mathbb{R}$ denote a function.

Lemma 3.3 *For any constant $p < 1$, there is an algorithm* `GeneralizedJuntaTester`$_{\mu,p}(f, k, \epsilon, T)$ *that, with probability at least p,*

- *accepts if $f \in \mathsf{Jun}_k(T)$.*
- *rejects if* $\mathrm{dist}(f, \mathsf{Jun}_k(T)) \geq \epsilon$*;*

- *makes* $\Theta(k^4 \log(k+1)/\epsilon)$ *non-adaptive queries, and the marginal distribution of each query is* μ.

The proof can be found in Appendix E.

The procedure to ϵ-test the property of being k-junto-symmetric, for small enough k, is described next.

1. Let $q = \theta(k^4 \log(k+1)/\epsilon)$ bound the query complexity of Step 3.
2. Test that $\mathrm{Inf}_{\mathrm{T(f)}}(\mathrm{V}_1) < \frac{1}{18\mathrm{q}}$ with confidence $> 8/9$ (by taking $O(q)$ random pairs $(\ell, r, x), (\ell, r', x)$ and comparing $T(f)$ on them). If it isn't, reject.
3. Reject iff $\texttt{GeneralizedJuntaTester}_{\mu,8/9}(T(f), k, \epsilon, V_0 \cup V_1)$ rejects.

Theorem 3.4 *(Theorem 1.5 restated). Let* $k < (2n)^{1/12}$ *and* $\epsilon > 1/k^4$, $f : \{0,1\}^n \to \{0,1\}$. *The algorithm above*

- *(completeness) accepts when* $\mathrm{dist}(f, \mathcal{JS}_k) \leq k^{-5}$ *with probability* $\geq 2/3$;
- *(soundness) rejects when* $\mathrm{dist}(f, \mathcal{JS}_k) \geq \epsilon$ *with probability* $\geq 2/3$;
- *has query complexity* $O(k^4 \log(k+1)/\epsilon)$ *and is non-adaptive.*

Proof of Theorem 3.4. The algorithm is clearly non-adaptive and its query complexity is $\Theta(q) = \Theta(k^4 \log(k+1)/\epsilon)$. We assume that k is large enough so that $2k/\sqrt{2n} < 1/(18q) < \epsilon/5$ (small constant values for k can be dealt with separately in the tester).

The probability that the junta tester in step 3 or the influence test of step 2 give incorrect assessments is less than $2/9 < 2/3$. So if the overall test accepts with probability $\geq 2/3$, then $T(f)$ must be $\epsilon/5$-close to a junta j' on $V_0 \cup V_1 \cup A$, $|A| \leq k$. In particular $\mathrm{Inf}_{\mathrm{T(f)}}(\mathrm{V}_2 \setminus \mathrm{A}) \leq \epsilon/5$. Moreover, since the influence test succeeded we also have $\mathrm{Inf}_{\mathrm{T(f)}}(\mathrm{V}_1) < \epsilon/5$. Therefore $\mathrm{Inf}_{\mathrm{T(f)}}(\mathrm{V}_1 \cup (\mathrm{V}_2 \setminus \mathrm{A})) \leq 2\epsilon/5$, which means that $T(f)$ is in fact $4\epsilon/5$-close to a junta on $V_0 \cup A$. Consequently, f is $4\epsilon/5 + k/\sqrt{2n} < \epsilon$-close to junto-symmetric on A (Lemma 3.2), proving soundness.

On the other hand, suppose f is $1/(18q)$-close to a junto-symmetric function j. Then there is $j' \in \mathsf{Jun}_k(V_0)$ with $\mathrm{dist}(T(f), j') \leq 1/(18q) + k/\sqrt{2n} < 1/(9q)$. Recall that every query of the junta tester to $T(f)$ follows the distribution $\mathcal{L} \times \mathcal{R} \times \{0,1\}^n$, and this translates into uniform queries to f. As the tester is non-adaptive, this means that the expected number of queries exposing a difference between $T(f)$ and j' is $1/9$, so with probability $8/9$ the tester can't see the difference between $T(f)$ and j'. Hence we are effectively testing j' for the property of being a $V_0 \cup V_1 \cup A$ junta for some $|A| \leq k$, which it is indeed. Therefore step 3 accepts with probability $8/9$; and since $\mathrm{Inf}_j(\mathrm{V}_1) = 0$, we also have $\mathrm{Inf}_{\mathrm{T(f)}}(\mathrm{V}_1) \leq 2 \cdot \mathrm{dist}(\mathrm{T(f)}, j') \leq 2\mathrm{k}/\sqrt{2\mathrm{n}} < 1/(18\mathrm{q})$ and step 2 also accepts with probability $8/9$. This establishes completeness. □

References

[AB10] N. Alon and E. Blais. Testing boolean function isomorphism. In *Proceedings of the 14th International Workshop on Randomization and Computation (RANDOM)*, pages 394–405, 2010.

[AFKS00] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.

[Bab81] L. Babai. On the order of uniprimitive permutation groups. *Annals of Mathematics*, 113(3):pp. 553–568, 1981.

[BC10] L. Babai and S. Chakraborty. Property testing of equivalence under a permutation group action. *ACM Transactions on Computation Theory*, 2010. To appear.

[Bla09] E. Blais. Testing juntas nearly optimally. In *Proceedings of the 41st ACM Symposium on Theory of Computing (STOC)*, pages 151–158, New York, NY, USA, 2009. ACM.

[BO10] E. Blais and R. O'Donnell. Lower bounds for testing function isomorphism. In *Proceedings of the 25th IEEE Conference on Computational Complexity (CCC)*, pages 235–246, 2010.

[Boc89] A. Bochert. Über die Zahl verschiedener Werte, die eine Funktion gegebener Buchstaben durch Vertauschung derselben erlangen kann. *Mathematische Annalen*, pages 584–590, 1889.

[Cam81] P. J. Cameron. Finite permutation groups and finite simple groups. *Bulletin of the London Mathematical Society*, 13:1–22, 1981.

[Cam99] P. J. Cameron. *Permutation groups*, volume 45 of *London Mathematical Society Student Texts*. Cambridge University Press, New York and London, 1999.

[CGM11a] S. Chakraborty, D. García-Soriano, and A. Matsliah. Efficient sample extractors for juntas with applications. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP)*, 2011.

[CGM11b] S. Chakraborty, D. García-Soriano, and A. Matsliah. Nearly tight bounds for testing function isomorphism. In *Proceedings of the 22nd ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011.

[CK91] P. Clote and E. Kranakis. Boolean functions, invariance groups, and parallel complexity. *SIAM Journal on Computing*, 20(3):553–590, 1991.

[CK02] P. Clote and E. Kranakis. *Boolean Functions and Computation Models*. EATCS Series. Springer-Verlag, 2002.

[DLM+07] I. Diakonikolas, H. K. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. A. Servedio, and A. Wan. Testing for concise representations. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 549–558, Los Alamitos, CA, USA, 2007. IEEE Computer Society.

[Fis01] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the EATCS*, 75:97, 2001.

[Fis05] E. Fischer. The difficulty of testing for isomorphism against a graph that is given in advance. *SIAM Journal on Computing*, 34(5):1147–1158, 2005.

[FKR+04] E. Fischer, G. Kindler, D. Ron, S. Safra, and A. Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004. Special Issue on FOCS 2002.

[FM08] E. Fischer and A. Matsliah. Testing graph isomorphism. *SIAM Journal on Computing*, 38(1):207–225, 2008.

[FMS10] E. Fischer, A. Matsliah, and A. Shapira. Approximate hypergraph partitioning and applications. *SIAM Journal on Computing*, 39:3155–3185, 2010.

[FNS04] E. Fischer, I. Newman, and J. Sgall. Functions that have read-twice constant width branching programs are not necessarily testable. *Random Structures and Algorithms*, 24(2):175–193, 2004.

[GGR98] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[GOS+09] P. Gopalan, R. O'Donnell, R. A. Servedio, A. Shpilka, and K. Wimmer. Testing Fourier dimensionality and sparsity. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 500–512, 2009.

[Kis98] A. Kisielewicz. Symmetry groups of boolean functions and constructions of permutation groups. *Journal of Algebra*, 199(2):379 – 403, 1998.

[PRS02] M. Parnas, D. Ron, and A. Samorodnitsky. Testing basic boolean formulae. *SIAM Journal on Discrete Mathematics*, 16(1):20–46, 2002.

[PS10] T. Pitassi and R. Santhanam. Effectively polynomial simulations. In *Proceedings of the first Symposium on Innovations in Theoretical Computer Science (ITCS)*, pages 370–382, 2010.

[Wie64] H. Wielandt. *Finite permutation groups*. Academic Press, New York and London, 1964.

A Notation and standard tools

A.1 Notation

Permutation groups

We need some basic notions from the theory of permutation groups (an exposition can be found in the books by Wielandt [Wie64] and Cameron[Cam99]). Let Ω be a set (which will be assumed finite in this paper). $Sym(\Omega)$ denotes the symmetric group of all permutations of Ω, and $Alt(\Omega)$ is the subgroup of $Sym(\Omega)$ made up of even permutations. When $|\Omega| = n$, we occasionally write $A_n = Alt(\Omega)$ and $S_n = Sym(\Omega)$. For reasons that will become apparent shortly, the product operation we use in $Sym(\Omega)$ is $\pi\sigma \triangleq \sigma \circ \pi$.

A *permutation group* G on Ω is a subgroup of $Sym(\Omega)$, written $G \leq Sym(\Omega)$. The image $\pi(x)$ of $x \in \Omega$ under $\pi \in G$ is often written x^π; under our convention we have $(x^\pi)^\sigma = x^{\sigma\circ\pi} = x^{\pi\sigma}$ for $\pi, \sigma \in G$. The *orbit* of a set $\Delta \subseteq \Omega$ under an arbitrary collection $H \subseteq G$ is the set $\Delta^H = \{\pi(x) \mid \pi \in H, x \in \Delta\}$. When $\Delta = \{x\}$ or $H = \{h\}$ are single-element sets we may simply write x^H or Δ^h.

G is called *transitive* if for every $x, y \in \Omega$ there is $\pi \in G$ with $x^\pi = y$. An intransitive group $G \leq Sym(\Omega)$ partitions Ω into orbits: these are the equivalence classes of the relation $\sim$ given by $x \sim y$ iff there is $\pi \in G$ such that $x^\pi = y$, which occurs iff $x^G = y^G$.

A *group action* of a (general) group G on a set Ω is a homomorphism $\phi : G \to Sym(\Omega)$. (This is what is called a *right* action because of our convention on the composition law in $Sym(\Omega)$). If $\ker\phi = 1_G$, the action is *faithful* and G is isomorphic (via ϕ) to a permutation group on Ω. It is customary to omit the explicit reference to the chosen ϕ and write x^g for $x^{\phi(g)}$ $(g \in G)$. Given an action of G on Ω, we can naturally extend it to define an action on subsets of Ω: $g \in G$ acts on $\mathcal{P}(\Omega)$ by mapping $\Delta \subseteq \Omega$ to Δ^g as defined above.

A *block* of G is a subset Δ of Ω such that for every $\pi \in G$, either $\Delta^\pi = \Delta$ or $\Delta^\pi \cap \Delta = \emptyset$. Obviously, Ω, the empty set $\emptyset$ and each of the singletons $\{i\}_{i\in\Omega}$ are always blocks; we call these the *trivial blocks*. The permutation group G is said to be *primitive* if it is transitive and has no non-trivial blocks. (Only transitive groups are classified as either primitive or imprimitive). If Δ is a block of G, then Ω can be partitioned into a *complete block system*, where every block is of the form Δ^g for some $g \in G$ (so all blocks in a complete block system have the same cardinality).

The *pointwise stabilizer* of $\Delta \subseteq \Omega$ is the set

$$G_\Delta \triangleq \{\pi \in G \mid x^\pi = x \; \forall x \in \Delta\}.$$

Function isomorphism

We consider the right action $\phi : S_n \to Sym(\{0,1\}^n)$ of S_n on $\{0,1\}^n$ defined in the following way: if $\pi \in S_n$, $\phi(\pi) \in Sym(\{0,1\}^n)$ is the permutation mapping each $x = x_1x_2\ldots x_n \in \{0,1\}^n$ to $\phi(\pi)(x) \triangleq x_{\pi(1)}x_{\pi(2)}\ldots x_{\pi(n)}$. As before, we identify π and $\phi(\pi)$ and we write x^π in place of $\phi(\pi)(x)$. Note that $\phi(\pi)$ effectively sends the input at position i into position $\pi^{-1}(x)$, and as a result we have $(x^\sigma)^\pi = x^{\pi\circ\sigma} = x^{\sigma\pi}$. We also write f^π for the function on $\{0,1\}^n$ defined by $f^\pi(x) = f(x^\pi)$; by the observations above, $(f^\pi)^\sigma = f^{\pi\sigma}$.

In this language, the functions f and g are isomorphic (in short, $f \cong g$) if there is $\pi \in S_n$ with $f = g^\pi$. The *distance up to permutations of variables* between f and g is defined by

$$\mathrm{distiso}(f, g) \triangleq \min_{\pi\in S_n} \mathrm{dist}(f^\pi, g).$$

Influence, Juntas, Parities, Cores

Testing f-isomorphism is defined as the problem of testing the property

$$\mathrm{Isom_f} \triangleq \{\mathrm{f}^\pi : \pi \in \mathrm{S_n}\}$$

in the usual property testing terminology. It is thus the task of distinguishing the case $f \cong g$ from the case $\mathrm{distiso}(f, g) \geq \epsilon$.

For a function $g : \{0,1\}^n \to \{0,1\}$ and a set $A \subseteq [n]$, the *influence* of A on g is defined as

$$\mathrm{Inf_g(A)} \triangleq \Pr_{\mathrm{x}\in\{0,1\}^{\mathrm{n}},\ \mathrm{y}\in\{0,1\}^{|\mathrm{A}|}} \left[\mathrm{g(x)} \neq \mathrm{g}(\mathrm{x}_{A\leftarrow y})\right].$$

Thus $\mathrm{Inf_g(A)}$ measures the probability that the value of g changes after a random modification of the bits in A of a random input x. Note that when $|A| = 1$, this value is half that of the most common definition of influence of one variable; for consistency we stick to the previous definition instead in this case as well. For example, every relevant variable of a k-parity ($k \geq 1$) has influence $\frac{1}{2}$.

An index (variable) $i \in [n]$ is *relevant* with respect to g if $\mathrm{Inf_g(\{i\})} \neq 0$. A *$k$-junta* is a function g that has at most k relevant variables; equivalently, there is $S \in \binom{[n]}{k}$ such that $\mathrm{Inf_g([n] \setminus S)} = 0$. Jun_k will denote the class of k-juntas (on n variables), and for $A \subseteq [n]$, Jun_A will denote the class of juntas all of whose relevant variables are contained in A.

By the *core* of a k-junta f we mean the boolean function $\mathsf{core}_k(f) : \{0,1\}^k \to \{0,1\}$ obtained from f by dropping its irrelevant variables (and fixing some arbitrary ordering for the relevant ones).

A.2 Adaptive lower bounds

Let $\mathcal{P}$ be a property (subset) of functions mapping T to $\{0,1\}$. Let

$$\mathcal{R} \subseteq \{f \in \{0,1\}^T \mid \mathrm{dist}(f, \mathcal{P}) \geq \epsilon\}$$

be non-empty. Any tester for $\mathcal{P}$ should, with high probability, accept inputs from $\mathcal{P}$ and reject inputs from $\mathcal{R}$.

We use the following lemma in various lower bound proofs for two-sided adaptive testing. It is proven implicitly in [FNS04], and a detailed proof appears in [Fis01].

Lemma A.1 *Let $\mathcal{P}, \mathcal{R}$ be as in the preceding discussion, and let $\mathcal{D}_{\mathrm{yes}}$ and $\mathcal{D}_{\mathrm{no}}$ be distributions over $\mathcal{P}$ and $\mathcal{R}$, respectively. If q is such that for all $Q \in \binom{T}{q}$ and $a \in \{0,1\}^Q$ we have*

$$(2/3) \Pr_{f\in\mathcal{D}_{\mathrm{yes}}}[f\restriction_Q = a] < \Pr_{f\in\mathcal{D}_{\mathrm{no}}}[f\restriction_Q = a],$$

then any tester for $\mathcal{P}$ with error probability $\leq 1/3$ must make more than q queries.

Remark A.1 *The proof of the lemma is based on a indistinguishability result that a tester needs q queries to tell apart a random $f \sim \mathcal{P}$ from a random $f \sim \mathcal{R}$ (where $\mathcal{P}$ or $\mathcal{R}$ are chosen with probability half). If we drop the condition that $\mathcal{R}$ only contain functions far from $\mathcal{P}$, the implication for property testing lower bounds disappears, but the indistinguishability result still holds.*

B Hypergraph crunching: upper bound

B.1 Hypergraph partition property

Let $H = (V, E)$ be an r-uniform hypergraph and let $\Pi = \{V_1^\Pi, \ldots, V_k^\Pi\}$ be a partition of V. Let us introduce a notation for counting the number of edges from E with a specific placement of their vertices within the partition classes of Π. We denote by Φ the set of all possible mappings $\phi : [r] \to [k]$. We think of every $\phi \in \Phi$ as mapping the vertices of an r-set to the components of Π. We denote by $E_\phi^\Pi \subseteq E$ the following collection of r-sets:

$$E_\phi^\Pi = \{\{v_1, \ldots, v_r\} \in E \mid \ \forall j \in [r],\ v_j \in V_{\phi(j)}^\Pi\}.$$

We now introduce a notion from the work of Fischer, Matsliah and Shapira [FMS10] that generalizes the partition instances that were discussed in the context of graphs by Goldreich, Goldwasser and Ron [GGR98].

Definition B.1 (density tensors) *A* density tensor *of* order k *and arity* r *is a sequence* $\psi = \left\langle \langle \rho_j \rangle_{j\in[k]}, \langle \mu_\phi \rangle_{\phi\in\Phi} \right\rangle$ *of reals (between* 0 *and* 1*) specifying the presumed normalized sizes of* $|V_i^\Pi|$ *and* $|E_\phi^\Pi|$ *of a* k*-wise partition of a hypergraph of arity* r *(whenever* k *and* r *are clear from the context, we call* ψ *simply a* density tensor*).*

In particular, given a partition $\Pi = \{V_1^\Pi, V_2^\Pi, \ldots, V_k^\Pi\}$ *of a hypergraph* H*, we set* ψ^Π *to be the density tensor* $\langle \langle \rho_j^\Pi \rangle_{j\in[k]}, \langle \mu_\phi^\Pi \rangle_{\phi\in\Phi} \rangle$ *with the property that for all* j*,* $\rho_j^\Pi = \frac{1}{n} \cdot |V_j^\Pi|$ *and for all* ϕ*,* $\mu_\phi^\Pi = \frac{1}{n^r} \cdot |E_\phi^\Pi|$*.*

Definition B.2 (partition properties induced by density tensors) *For a fixed hypergraph* H *of arity* r*, a set* Ψ *of density tensors (of order* k *and arity* r*) defines a property of the* k*-wise partitions of* $V(H)$*'s as follows. We say that a partition* Π *of* $V(H)$ *(exactly)* satisfies Ψ *if there exists a density tensor* $\psi = \left\langle \langle \rho_j \rangle_{j\in[k]}, \langle \mu_\phi \rangle_{\phi\in\Phi} \right\rangle \in \Psi$*, such that* ψ *and the density tensor* ψ^Π *of* Π *are equal. Namely,* Π satisfies Ψ *if there is* $\psi = \left\langle \langle \rho_j \rangle_{j\in[k]}, \langle \mu_\phi \rangle_{\phi\in\Phi} \right\rangle \in \Psi$ *such that*

- *for all* $j \in [k]$*,* $\rho_j^\Pi = \rho_j$*;*
- *for all* $\phi \in \Phi$*,* $\mu_\phi^\Pi = \mu_\phi$*.*

We extend this notion of satisfying partitions (and equivalence between density tensors) in two ways as follows.

Definition B.3 (being ϵ-approximately satisfying/equal) *A partition* Π ϵ-approximately *satisfies* Ψ *if there is*
$\psi = \left\langle \langle \rho_j \rangle_{j\in[k]}, \langle \mu_\phi \rangle_{\phi\in\Phi} \right\rangle \in \Psi$ *such that*

- *for all* $j \in [k]$*,* $\rho_j^\Pi = \rho_j$*;*
- *for all* $\phi \in \Phi$*,* $\mu_\phi^\Pi = \mu_\phi \pm \epsilon$*.*

In this case ψ^Π *is* ϵ-approximate *to* ψ*.*

By extension (and with a slight abuse of notation), we say that the hypergraph H itself *satisfies* the property Ψ if there exists a partition Π of H's vertices that satisfies Ψ, and similarly we say that H itself *ϵ-approximately* satisfies the property Ψ if there exists a partition of H's vertices that ϵ-approximately satisfies the property Ψ. In addition, we may consider a specific density tensor ψ as a singleton set $\Psi = \{\psi\}$, and accordingly as a property of partitions.

Let us now describe the hypergraph-partition testing algorithm from [FMS10] (it is stated there for directed hypergraphs, but it also applies to undirected ones). To this end, we define one additional measure of closeness to the property Ψ. The distance of a hypergraph H from the property Ψ is defined as $dist(H, \Psi) = \min_{H'}\{dist(H, H') : H' \text{ satisfies } \Psi\}$. For $\epsilon > 0$ we say that H is *ϵ-far* from satisfying the property Ψ when $dist(H, \Psi) > \epsilon$, and otherwise, H is *ϵ-close* to Ψ. The testing algorithm follows immediately from the following theorem.

Theorem B.4 ([FMS10]) *For every $k, r \in \mathbb{N}$, and set Ψ of density tensors of order k and arity r, there exists a randomized algorithm A_T taking as inputs two parameters $\epsilon, \delta > 0$ and an oracle access to a hypergraph H of arity r, such that*

- *if H satisfies Ψ, then with probability at least $1 - \delta$ the algorithm A_T outputs* ACCEPT*;*
- *if H does not even ϵ-approximately satisfy the property Ψ, then with probability at least $1 - \delta$ the algorithm A_T outputs* REJECT*.*

The query complexity of A_T is bounded by $\log^3(\frac{1}{\delta}) \cdot \text{poly}(k^r, \frac{1}{\epsilon})$, and its running time is bounded by $\log^3(\frac{1}{\delta}) \cdot \exp\left((\frac{r}{\epsilon})^{r \cdot k^r}\right)$.

The algorithm above can be used as a testing algorithm in the traditional sense due to the following trivial observation.

Claim B.1 *Let $\delta < \epsilon/binomkr$. Any hypergraph that δ-approximately satisfies a partition property Ψ is also ϵ-close to satisfying it.* □

B.2 Testing isomorphism to k-simple hypergraphs

Let $H = (V, E)$ be a k-simple r-uniform hypergraph, with the corresponding mapping $\Pi : V \to [k]$ and an hypergraph $\hat{H} = ([k], \hat{E})$ defining the edge patterns of H.

Let $\psi = \left\langle \langle \rho_j \rangle_{j \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \right\rangle$ denote the following density tensor of order k and arity r:

- for all $j \in [k]$, $\rho_j = \frac{\Pi^{-1}(j)}{|V|}$;
- for $\phi \in \Phi$, $\mu_\phi = \frac{E^\Pi_\phi}{|V|^r}$.

Note that a hypergraph H' is isomorphic to H if and only if it satisfies the partition property $\{\psi\}$, hence ϵ-testing isomorphism to H reduces to testing the partition property for $\{\psi\}$ with proximity parameter $\epsilon/\binom{k}{r}$.

C Characterizations of $O(1)$-junto-symmetric families

In this section we show Theorem 1.4. To ease readability we drop the subscripts, i.e. write f and A in place of f_n and A_n. All but one of the implications we need are straightforward:

- $(b) \implies (c)$: $Sym([n] \setminus A) \leq \mathrm{Aut(f)}$ means that f is invariant under permutations of $[n] \setminus A$, i.e. $f(x) = f(y)$ whenever $x\restriction_A = y\restriction_A$ and $|x\restriction_{[n]\setminus A}| = |y\restriction_{[n]\setminus A}|$. These conditions are equivalent to $x\restriction_A = y\restriction_A$ and $|x| = |y|$, so f has the form $f(x) = \hat{f}(|x|, x\restriction_A)$ (where $|A| = O(1)$ by assumption).

- $(c) \implies (d)$: Let $f = \hat{f}(|x|, x\restriction_A)$, $|A| = k = O(1)$. Define $\hat{f}^{(i)}(x) = \hat{f}(i, x\restriction_A)$. Each $\hat{f}^{(i)}$ is a junta on A. The number of A-juntas is only $\ell = 2^{2^k} = O(1)$; let $\jmath_1, \ldots, \jmath_\ell$ be an enumeration of them and let

$$h_i(x) \triangleq \begin{cases} 1 & \text{if } \hat{f}^{(|x|)} = \jmath_i \\ 0 & \text{otherwise} \end{cases}.$$

  Each h_i is a symmetric function, and f can be decomposed into

$$f(x) = \bigvee_{i \in [\ell]} h_i(x) \wedge \jmath_i(x),$$

  which is a boolean combination of ℓ symmetric functions and the j_i, themselves a combination of the k dictatorship functions $\{x_i\}_{i \in A}$.

- $(d) \implies (b)$: Let $f(x) = \hat{f}(s_1(x), \ldots, s_\ell(x), x_{i_1}, \ldots, x_{i_k})$, where $s_1, \ldots, s_\ell$ are symmetric. Set $A = \{i_1, \ldots, i_k\}$ and let $\pi \in Sym([n] - A)$. Each function s_i remains invariant under $Sym([n])$, and each dictatorship x_{i_j} is invariant under $Sym([n] \setminus \{i\}) \supseteq Sym([n] \setminus A)$. Therefore $Sym([n] \setminus A) \leq \mathrm{Aut(f)}$. [2]

- $(c) \implies (a)$: As we saw, if $f \in \mathcal{JS}_k$ and $k = O(1)$, then $\mathrm{Isom(f)} \leq \binom{\mathrm{n}}{\mathrm{k}}\mathrm{k!} = \mathrm{n}^{\mathrm{O(1)}}$.

The only remaining implication is $(a) \implies (b)$. [3] First we need a handy result that provides a lower bound for the index of primitive groups. (Asymptotically better bounds are available [Bab81, Cam81], but this one will suffice).

Theorem C.1 (Bochert's bound [Boc89]; c.f. Thm. 14.2 in [Wie64]) *Let G be a primitive subgroup of S_n, other than S_n and A_n. Then*

$$[S_n/G] \geq \lceil n/2 \rceil !.$$

Lemma C.2 *Let $n \geq 13$, $G \leq S_n$, $G \neq S_n, A_n$. Then*

[2] Note that the fact that $\ell = O(1)$ is immaterial here, and in fact yet another equivalent definition can be given by substituting "any number of symmetric functions" for "$O(1)$-many symmetric functions".

[3] This would be implied by the claim following Theorem 28 in page 586 of [CK91], but unfortunately this claim is in error (as can be seen by taking G_n to be the alternating group A_n). The mistake seems to lie near the end of the proof, after it is shown that $i_n \leq k$ and $|S_n : G_n| \leq n^k$, the claim that $V_n = S_{n-i_n}$ is unjustified (it would seem to assume that in fact $i_n = k$). The lemma does hold for the automorphism groups of boolean functions however as we show, which is the case of interest in both papers.

(a) If G is transitive then

$$[S_n/G] \geq \frac{1}{2}\binom{n}{\lfloor n/2 \rfloor}.$$

(b) Suppose G is intransitive; let Δ be the longest orbit of an element of $[n]$ and $\ell = |\Delta| < n$ its size. Then

$$[S_n/G] \geq \binom{n}{\max(n/2, \ell)}.$$

(c) Under the same conditions as in (b), let

$$H \triangleq G \cap Sym(\Delta) = G \cap S_\ell$$

be the pointwise stabilizer of $[n] \setminus \Delta$. Then

$$[S_\ell/H] \leq \frac{[S_n/G]}{\binom{n}{\ell}}.$$

Proof.

(a) If G is primitive, Bochert's theorem states the bound $[S_n/G] \geq \lceil n/2 \rceil !$, which is stronger for $n \geq 13$. So suppose G is transitive and imprimitive, with a block of imprimitivity of size $2 \leq a \leq n/2, a \mid n$ (and hence $b = n/a$ such blocks because of transitivity). Then

$$|G| \leq (a!)^b b! \leq \lfloor (ab/2) \rfloor ! \lceil (ab/2) \rceil !.$$

To prove the last inequality, observe that for $a = 2$ it reduces to the triviality $b! \geq 2^b$. Hence it suffices to verify that for any $b \geq 2$, the quotient

$$q(a) \triangleq \frac{a!^b}{\lfloor (ab/2) \rfloor ! \lceil (ab/2) \rceil !}$$

is a decreasing function of a. Define the sequences $\{s_i\}, \{t_i\}, i \in [1, ab]$ by $s_i = \lfloor (i + b - 1)/b \rfloor$ and $t_i = \lfloor (i+1)/2 \rfloor$. Then

$$\begin{aligned} q(a) &= \prod_{i=1}^{ab} \frac{s_i}{t_i} = \prod_{i=1}^{(a-1)b} \frac{s_i}{t_i} \cdot \prod_{j=(a-1)b+1}^{ab} \frac{s_j}{t_j} \\ &= q(a-1) \cdot \prod_{j=(a-1)b+1}^{ab} \frac{a}{t_j} \\ &\leq q(a-1), \end{aligned}$$

because $t_{(a-1)b+1} = \lfloor (a-1)b/2 \rfloor + 1 \geq a$ since $b \geq 2$. Therefore

$$[S_n/G] = \frac{n!}{|G|} \geq \frac{1}{2}\binom{n}{\lfloor n/2 \rfloor}.$$

(b) Let $A_1, \ldots, A_m$ be the orbits and $a_i = |A_i|$. Since $G \leq Sym(A_1) \times Sym(A_2) \times \ldots \times Sym(A_m)$, we must have

$$|G| \leq \prod a_i!.$$

Fix ℓ and let us consider the maximum $r(\ell)$ of the expression on the right hand side subject to $0 \leq a_i \leq \ell$, over all choices of $m \geq 2$ (without loss of generality $m = n$). We claim that there are optimal solutions where $a_i = \ell$ for at least one i (in fact precisely $\lfloor n/\ell \rfloor$ of them but we shall not need this). This is due to the inequality $a_i!a_j! \leq 1!(a_i+a_j-1)!$, which tells us that replacing the pair (a_i, a_j) (where $a_i \geq a_j$) by $(a_i + a_j - 1, 1)$ cannot decrease $\prod a_i!$. This replacement preserves $\sum a_i$ and is possible unless either $a_i = \ell$ or $a_j = 0$ for every such pair. The replacing process can not go on forever because if we look at its effect on the sorted sequence $\{a_i\}$, we see that it becomes lexicographically larger at each step. By taking an optimal solution and performing such replacements while possible, we arrive at a solution with some $a_i = \ell$ (as not all pairs can be zero).

Now observe that $\prod a_i! \leq a_i!(n - a_i)!$ for any i. (For example, this can be seen by noting that the left-hand side is the size of the set of permutations $Sym(A_1) \times \ldots \times Sym(A_m)$, and this a subset of $Sym(A_i) \times Sym([n] \setminus A_i)$). So using $a_i = \ell$ for some i we get

$$|G| \leq r(\ell) \leq \ell!(n - \ell)! = \frac{n!}{\binom{n}{\ell}}.$$

This shows that

$$[S_n/G] \geq \binom{n}{\ell}$$

for any ℓ. On the other hand, $r(\ell)$ is by definition an increasing function of ℓ, so the inequality

$$[S_n/G] \geq \binom{n}{\lfloor n/2 \rfloor}$$

holds for any $\ell \leq n/2$.

(c) Because $G \leq H \times Sym([n] - \Delta)$, we can bound

$$|G| \leq |H||S_{n-\ell}| = |H|(n - \ell)!,$$

which yields

$$[S_\ell/H] = \frac{\ell!}{|H|} \leq \frac{\ell!(n - \ell)!}{|G|} = \frac{[S_n/G]}{\binom{n}{\ell}}.$$

□

Lemma C.3 *Let $n \geq 22$, $t \leq n/2$, $[S_n/G] < \frac{1}{4}\binom{n}{t}$, and Δ, ℓ as before. Then $\ell > n - t$ and $Alt(\Delta) \leq G$.*

Proof. If the action of G is transitive on $[n]$ then $[S_n/G] \geq \frac{1}{2}\binom{n}{n/2}$ by Lemma C.2(a), which contradicts our assumptions. So G is not transitive and $\ell < n$. If $\ell \leq n/2$ we have, by Lemma C.2(b), $[S_n/G] \geq \binom{n}{n/2}$, which again is impossible.

We are left with the case $n/2 < \ell < n$. According to Lemma C.2(b),

$$[S_n/G] \geq \binom{n}{\ell} = \binom{n}{n-\ell},$$

so $t > n - \ell$. Let $H = G \cap S_\Delta$. This is actually the pointwise stabilizer of $[n] \setminus \Delta$ in G, and since Δ is an orbit of G it follows that H is normal in G (Proposition 3.1 of [Cam99]). We demonstrate that $A_\Delta \leq H$ by contradiction. We argue by contradiction in order to prove that indeed $A_\Delta \leq H$.

So assume $H \neq S_\Delta, A_\Delta$. Then Lemma C.2 applies to the group H acting on Δ. Let Δ' be the largest orbit of this action and $\ell' = |\Delta'|$. Since G is transitive on Δ and $H \lhd G$, it is not hard to see that the length of any orbit of H on Δ must divide ℓ, i.e. $\ell' \mid \ell$. We distinguish two cases:

- If $\ell' \leq \ell/2$, then
$$[S_\ell/H] \geq \binom{\ell}{\ell/2}$$
by part (b) of the "inner" application of Lemma C.2.
- If $\ell' > \ell/2$, then as we observed that $\ell' \mid \ell$, we must in fact have $\ell' = \ell$, meaning that H is transitive on Δ and
$$[S_\ell/H] \geq \frac{1}{2}\binom{\ell}{\ell/2}$$
by part (a) of the "inner" application of Lemma C.2.

In any case we have

$$[S_\ell/H] \geq \frac{1}{2}\binom{\ell}{\ell/2}.$$

Coupled with part (c) of the "outer" application, i.e.

$$[S_\ell/H] \leq \frac{[S_n/G]}{\binom{n}{\ell}},$$

this yields the contradiction

$$[S_n/G] \geq \frac{1}{2}\binom{\ell}{\ell/2}\binom{n}{\ell} \geq \frac{1}{4}\binom{n}{n/2}.$$

□

Corollary C.4 *Let $n \geq 22$ and $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with $|\mathrm{Isom(f)}| < (1/4)\binom{\mathrm{n}}{\mathrm{t}}$, $t \leq n/2$. Then there is a set Δ of size $\ell > n - t$ such that f is junto-symmetric on Δ. In particular, any poly-symmetric family is junto-symmetric on sets of size $O(1)$.*

Proof. Let $G = \mathrm{Aut(f)}$. Since $|\mathrm{Isom(f)}| = [\mathrm{S_n/G}]$, the previous lemma gives $Alt(\Delta) \leq \mathrm{Aut(f)}$ and $|\Delta| \geq n/2 > 3$. But repeating the argument presented in the introduction shows that for a boolean function this implies $Sym(\Delta) \leq \mathrm{Aut(f)}$. □

This corollary is the last piece we needed to show Theorem 1.4.

D Graphs, algebra number and crunching

Definition D.1 *The* algebra number *of a graph G is the smallest number k for which there exist cliques $C_1, \ldots, C_k$ over subsets of the vertex set of G, such that G can be generated from the edge sets of $C_1, \ldots, C_k$ by taking set unions, intersections and complementations (the latter with respect to the edge set of a complete graph).*

The ϵ-approximate algebra number *is the smallest k such that H is ϵ-close to some graph whose algebra number is k.*

We also define the pairing number *as the smallest k for which there are k* vertex-disjoint *sets $A_1, \ldots A_k \subseteq V$ and a subset $S \subseteq [k] \times [k]$ such that the edge set of G is $E = \cup_{(i,j)\in S} A_i \times A_j$ (note that $i = j$ is allowed). The* ϵ-approximate pairing number *of G is defined similarly.*

Lemma D.2

1. *Any graph with pairing number k is has algebra number $\leq k$.*

2. *Any graph with algebra number k has pairing number $\leq 2^k$.*

3. *Any k-crunchable graph has pairing number k. Conversely, any k-crunchable graph is ϵ-close to having pairing number $\leq k^2/\epsilon$.*

Proof.

1. Immediate.

2. Let $G = (V, E)$ be generated from the edge sets of the cliques $C_1, \ldots, C_k \subseteq V$. For $S \subseteq [k]$, let $A_S = (\cap_{i\in S} C_i) \bigcap (\cap_{i\notin S} \overline{C_i})$. These 2^k sets are disjoint and contain all vertices incident with some edge in G. For all $S, T \subseteq [k]$, if $a_1, a_2 \in A_S$ and $b_1, b_2 \in A_T$, then $(a_1, b_1) \in E$ iff $(a_2, b_2) \in E$ (unless $a_1 = b_1$ or $a_2 = b_2$). This means G has pairing number k since it is possible to write E in the required form.

3. We prove the second statement (the first one is obvious). Suppose G has pairing number k and let $A_1, \ldots, A_k$ be as in Definition D.1. The only reason G may not be k-crunchable is the possible existence of edges between vertices in the same A_i. Divide each A_i into $t = \lceil 1/\epsilon \rceil$ subsets $A_{i1}, \ldots, A_{ik}$ of roughly equal size and remove the edges with both endpoints inside the same A_{ij}. Note that if $|A_{ij}| = \alpha_{ij} n$, then from all $\binom{n}{2}$ edges, the removed ones constitute a fraction bounded by $\sum_{ij} \alpha_{ij}^2/t \leq (\sum \alpha_{ij})^2/t \leq 1/t \leq \epsilon$. Hence this graph is ϵ-close to the original graph, and is also k-crunchable by construction.

□

Corollary D.3 *If the ϵ-approximate algebra number of a graph is more than k, then testing isomorphism to it requires $\Omega(k^{1/4}\sqrt{\epsilon}/r)$ queries.*

E Generalized junta testing

Proof of Lemma 3.3. All known junta testers can be used in a straightforward manner for this generalized property while preserving the exact query complexity. One way to see this is to think of providing the junta tester with a set T of relevant variables for free, and instruct it to seek for relevant blocks outside T just as if the tester had found the variables of T by itself. (Note however that the "partitioning step" must be applied to $[m] \setminus T$).

For a more detailed description we start with an overview of how the tester of [FKR+04] works. Due to the inequality $\mathrm{dist}(f, \mathsf{Jun}_A) \leq \mathrm{Inf_f}([\mathrm{m}] \setminus \mathrm{A})$, the task can be reduced to accepting k-juntas and rejecting functions where the influence outside any set of size k is at least ϵ. Then the tester uses its random coin flips to select a series of disjoint subsets $I_1, \ldots, I_r \subseteq [m]$, and performs a number of repetitions of the basic independence test on each of them. (By an "independence test" on I_i we mean drawing a pair (x, y) from $\mu \times \mu$ conditioned on $x\restriction_{I_i} = y\restriction_{I_i}$, and rejecting when $f(x) \neq f(y)$). These subsets are then shown to satisfy the property

$$\begin{gathered}\text{if } \mathrm{Inf_f}([\mathrm{m}] \setminus \mathrm{A}) \geq \epsilon \text{ for all } A \subseteq [m], |A| = k, \text{ then} \\ \text{at least } k+1 \text{ of the independence tests will be positive.}\end{gathered}$$

(When f is a k-junta, at most k of them will be positive because of the disjointness condition). However, for any $B \subseteq [m]$ the same argument goes through to give a series of disjoint independence tests on $I'_1, \ldots, I'_r \subseteq B$ with the property

$$\begin{gathered}\text{if } \mathrm{Inf_f}(\mathrm{B} \setminus \mathrm{A}) \geq \epsilon \text{ for all } A \subseteq B, |A| = k, \text{ then} \\ \text{at least } k+1 \text{ of the independence tests will be positive.}\end{gathered}$$

(In fact, $I'_1, \ldots, I'_r$ are precisely the queries the junta tester would make for testing k-juntas on $[m] \setminus B$).

To adapt these ideas to our task, note that if $\mathrm{dist}(f, \mathsf{Jun}_k(T)) \geq \epsilon$ then $\mathrm{Inf_f}([\mathrm{m}] \setminus (\mathrm{T} \cup \mathrm{A})) \geq \epsilon$ for any $A \in \binom{[m] \setminus T}{k}$. Let $B = [m] \setminus T$ and $I'_1, \ldots, I'_r \subseteq B$ as before. We simply perform the independence tests of f on $I'_1, \ldots, I'_r$ and reject if at least $k+1$ were positive; both soundness and completeness follow from the preceding comments. Finally, the query complexity remains the same as that of the standard junta tester, and the second part of the last item follows because it is true of the independence tests. □

F Testing isomorphism

F.1 Testing isomorphism to junto-symmetric functions

In an analogous fashion one can reduce the problem of testing isomorphism to g (when g is close enough to $\mathcal{JS}_k$) to testing isomorphism between k-juntas. For this we can use a tolerant tester of isomorphism, except that, in view of Lemma 3.2(e), the set of permutations allowed must be restricted to those fixing V_0 and V_1:

1. Use the algorithm of Theorem 1.5 to accept if $f \in \mathcal{JS}_k$ and reject if $\mathrm{dist}(f, \mathcal{JS}_k) > \epsilon/30$.
2. Perform a suitable test to accept if $d \leq \epsilon/10$ and reject if $d \geq 9\epsilon/10$, where

$$d \triangleq \min_{\pi \in 1_{V_0, V_1} \times Sym(V_2)} \mathrm{dist}(T(f), T(g)^\pi)$$

Ignoring for the moment the implementation details of the second test, we show that the algorithm outlined is an isomorphism tester for $\mathcal{JS}_k$:

Proof of Theorem 1.6. We use the algorithm just described. The claim about the query complexity is clear.

Suppose the test accepts with high probability. Then $\mathrm{dist}(f, \mathcal{JS}_k) \leq \epsilon/30$ and $d \leq 9\epsilon/10$. Since $\mathrm{distiso}(g, \mathcal{JS}_k) \leq 1/k^5$, we have

$$|\mathrm{distiso}(f, g) - d| \leq \epsilon/20 + (1/k^5) + (2k)/\sqrt{2n} \leq \epsilon/20,$$

so $\mathrm{distiso}(f, g) < \epsilon$, as it should.

On the other hand, if $f \cong g$ then $\mathrm{dist}(f, \mathcal{JS}_k) = \mathrm{dist}(g, \mathcal{JS}_k) < 1/k^5$ and $d \leq 2/k^5 + (2k)/\sqrt{2n} < 1/k^4$, meaning that both tests succeed. □

Step 2 can be implemented using standard techniques. To formalize this we use the notion of noisy sampler extractors developed in [CGM11a]. Let $D = V_0 \times V_1$, $f : D \times \{0,1\}^n \to \{0,1\}$ and let $\jmath' \in \mathsf{Jun}_D(A), A \in \binom{[n]}{k}$ be the element of $\mathsf{Jun}_k(D)$ closest to f. Define $\mathsf{core}_{k,D}(\jmath') : D \times \{0,1\}^k \to \{0,1\}$ by

$$\mathsf{core}_{k,D}(\jmath')(x\restriction_D, x\restriction_A) = \jmath'(x).$$

A correct sample for $\mathsf{core}_{k,D}(\jmath')$ (with respect to $\sigma \in 1_D \times S_k$) is a pair (x, a) with $x \in D \times \{0,1\}^k$ and $\mathsf{core}_{k,D}(f)(x^\sigma) = a$. An *$\eta$-noisy sampler* for $\mathsf{core}_{k,D}(\jmath')$ is a procedure to obtain an unlimited sequence of independent samples (x, a) such that each one is correct with probability $1 - \eta$ with respect to some fixed σ, and x follows the distribution $D \times \{0,1\}^k$.

The following two lemmas are all we need.

Lemma F.1 *Suppose* $\mathrm{dist}(f, \mathsf{Jun}_k(D)) < 1/k^5$. *Then there is a* $\mathrm{poly}(k, 1/\epsilon)$*-query nonadaptive algorithm to construct an* $\epsilon/100$*-noisy sampler for* $\mathsf{core}_{k,D}(\jmath')$.

Proof (sketch). We assume familiarity with the proof of Lemma 2 of [CGM11a]. We need two changes. The first is that we substitute the adaptive junta tester of Blais [Bla09] for the junta tester used in the proof. The second one is the observation that we know how the variables in D map to the variables in $\mathsf{core}_k(\jmath')$, so for any $z \in \{0,1\}^n$, we only need to "extract" the setting of the k relevant variables sitting outside A. □

Lemma F.2 *Let* $f, g : D \times \{0,1\}^n \to \{0,1\}$, $g \in \mathsf{Jun}_k(D)$. *Write*

$$d = \min_{\pi \in 1_D \times S_n} \mathrm{dist}(f, g^\pi)$$

Assuming access to an $\epsilon/100$*-noisy sampler for* f, *there is a* $\mathrm{poly}(k/\epsilon)$*-query tester that accepts if* $d \leq \epsilon/10$ *and rejects if* $d \geq 9\epsilon/10$.

Proof (sketch). This is essentially Lemma 1 of [CGM11a]. Construct a sample for $\mathsf{core}_{k,D}(\jmath')$ and take $O(\log k!/\epsilon^2) = O(k \log k/\epsilon^2)$ random samples. These are enough to estimate

$$d' = \min_{\pi \in S_k} \mathrm{dist}(\mathsf{core}_{k,D}(\jmath'), \mathsf{core}_{k,D}(g)^\pi)$$

to within $O(\epsilon)$ additive terms. Finally recall that d' and d are the same up to constant factors (this follows from Lemma 6.1 in [CGM11b]). □

F.2 Tolerant testers

From proposition 2.3 of [Bla09], an $\exp(k/\epsilon)$-query *tolerant* junta tester easily follows. Using this instead of the $\mathrm{poly}(k/\epsilon)$-query junta tester in the proof of the previous theorems, we obtain Theorem 1.8.

G Junto-symmetric functions vs. layered juntas

Now we apply the crunching method to boolean functions. In this setting the procedure is similar to, but not quite the same as, an idea used by Blais and O'Donnell [BO10].

Definition G.1 (Layered juntas) *A function $f : \{0,1\}^n \to \{0,1\}$ is called a* layered k-junta *if there are subsets $J_0, \ldots, J_n \subseteq [n]$, each of size k, and functions $\tilde{f}_0, \ldots \tilde{f}_n : \{0,1\}^k \to \{0,1\}$ so that for all $x \in \{0,1\}^n$,*

$$f(x) = \tilde{f}_{|x|}(x\restriction_{J_{|x|}}).$$

Let $\mathcal{LJ}_k$ denote the class of layered k-juntas, respectively. Note that $\mathcal{JS}_k \subseteq \mathcal{LJ}_k$.

We need a notion of random crunching for functions. The notion for hypergraphs provides a possible definition of function crunching via the equivalence discussed in Section 1.2, but unfortunately this kind of crunching would alter the Hamming weight of inputs, which could be easily detected by the tester for some functions. Here we give a slightly different definition that resolves this issue, but only applies to layered juntas, and also happens to depend on the particular choice of each $\tilde{f}_i$ (remember that they are not uniquely defined).

Definition G.2 (Function crunching) *A* random t-crunching *of the function f defined by $f(x) = \tilde{f}_{|x|}(x\restriction_{J_{|x|}})$ is a function $g \in \mathcal{JS}_t$ obtained as follows:*

1. *pick, uniformly at random, a subset $J \subseteq [n]$ of size t and a mapping $\gamma : [n] \to J$;*
2. *for every $x \in \{0,1\}^n$, let $i_1, \ldots, i_k$ denote the indices in $J_{|x|}$; set $g(x) = \tilde{f}_{|x|}(x_{\gamma(i_1)} \cdots x_{\gamma(i_k)})$ and return g.*

Theorem G.3 *Fix $\epsilon > 0$ and $Q : \mathbb{N} \to \mathbb{N}$, and suppose $f \in \mathcal{LJ}_k$. Then $\Omega(Q(k))$ queries are needed to distinguish a random permutation of f from a random permutation of a random $(k \cdot Q(k))^2$-crunching of f.*

In particular, if f is ϵ-far from $\mathcal{JS}_{(k \cdot Q(k))^2}$, then ϵ-testing isomorphism to f requires $\Omega(Q(k))$ queries.

Proof. Let $\mathcal{D}_{\text{yes}}$ denote the random permutations of f and $\mathcal{D}_{\text{no}}$ the distribution of random permutations of t-crunchings of f. Given $g \in \mathcal{D}_{\text{no}}$ and its corresponding mapping $\gamma : [n] \to J$, call a set of layers $\{\ell_1, \ldots, \ell_m\}$ *collision-free* if γ is injective over $\bigcup_{i \in [m]} J_{\ell_i}$. Let T be a deterministic tester that makes $q = o(Q(k))$ queries. For every $x^1, \ldots, x^q \in \{0,1\}^n$ let $E_{x^1,\ldots,x^q}$ denote the event that the set $\{|x^1|, \ldots, |x^q|\}$ of layers is collision-free with respect to the randomly chosen mapping γ of a function $g \sim \mathcal{D}_{\text{no}}$. Observe that for all $x^1, \ldots, x^q \in \{0,1\}^n$ and $w \in \{0,1\}^q$, conditioned on $E_{x^1,\ldots,x^q}$ we have

$$\Pr_{h \sim \mathcal{D}_{\text{yes}}}[h(x^1), \ldots, h(x^q) = w] = \Pr_{h \sim \mathcal{D}_{\text{no}}}[h(x^1), \ldots, h(x^q) = w].$$

By Lemma A.1 (and the remark following its statement), it is enough to show that $E_{x^1,\ldots,x^q}$ occurs with probability $> 2/3$.

The probability (over $g \in \mathcal{D}_{\text{no}}$) that $\gamma(i) = \gamma(j)$ for a specific pair $i \neq j$ is $(k \cdot Q(k))^{-2}$. The number of different pairs $i, j \in \bigcup_{i \in [q]} J_{|x^q|}$ is bounded by $(kq)^2 = o((k \cdot Q(k))^2)$, hence by the union bound the probability that the set $\{|x^1|, \ldots, |x^q|\}$ of layers is collision-free is $1 - o(1)$. □

Note that this kind of argument admits certain generalizations. For example, by considering functions that have k additional variables outside a known set A (as in Section 3.1) of size $\approx \log n$, one can prove an $\tilde{\Omega}(\sqrt{n})$ lower bound for testing isomorphism to the address function.

H Isomorphism up to linear transformations

Definition H.1 *Two boolean functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^n \to \{0,1\}$ are said to be* linearly isomorphic *if there exists a full-rank linear transformation $A : \{0,1\}^n \to \{0,1\}^n$ such that $f = g \circ A$.*

This is an equivalence relation by virtue of the requirement that A have full rank.

Definition H.2 *Let $f(x) = \sum_{S \in \{0,1\}^n} \hat{f}(S)\chi_S(x)$ be the Fourier expansion of the function $f : \{0,1\}^n \to \mathbb{R}$. Let*

$$A \triangleq \left\{ S \in \{0,1\}^n | \hat{f}(S) \neq 0 \right\}.$$

Then the dimension of the span of A is called the Fourier dimension *of f.*

Lemma H.3 *The function f is linearly isomorphic to some k-junta iff its Fourier dimension is at most k.*

Proof. Suppose f has Fourier dimension k. Then it is a real linear combination of parities whose defining vectors lie on a k-dimensional vector space V. Here we take the parities χ_v to be ± 1-valued. Each parity in V can be written as a product of some parities in a basis for V. It follows that f can be written as a function h (not necessarily linear) of $k' \leq k$ linearly independent parities:

$$f(x) = h(\chi_{v_1}(x), \ldots, \chi_{v_{k'}}(x)) = g(\langle v_1, x\rangle, \langle v_2, x\rangle, \ldots, \langle v_{k'}, x\rangle, \bullet),$$

where g is a junta on the first k' variables, the inner products are taken over $\mathbb{F}_2^n$, and $\bullet$ symbolizes that the remaining $n - k'$ variables are irrelevant. The function g can can easily seen to be boolean-valued on $\{0,1\}^n$ if f is, because all $2^{k'}$ assignments to $\chi_{v_i}(x), i = 1..k'$ are possible. Hence there is a k'-junta $g : \{0,1\}^n \to \{0,1\}$ and a change of basis $A : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that $f = g \circ A$ (take $v_1, \ldots, v_k'$ as the first rows of the matrix associated with A).

Conversely, if $f = g \circ A$ for a k-junta g, then f is a junta on a set P of k parity functions and can be written as a polynomial on those parities. We can replace products of parities in P by a single parity in their span, and this means that f can also be written as a linear combination of the parities in the span of P, so f has Fourier dimension at most k. □

Theorem H.4 *If $f : \{0,1\}^n \to \{0,1\}$ is ϵ-far from having Fourier dimension k then any adaptive ϵ-tester for linear isomorphism to f takes at least $k - 1$ queries.*

Proof. Let $f : \{0,1\}^n \to \{0,1\}$ be ϵ-far from having Fourier dimension k. We will use Theorem A.1 to prove the lower bound. We want to generate two distributions of functions D_Y and D_N for the yes-instances and no-instances respectively:

$\mathcal{D}_{\text{yes}}$: To generate a random function g_Y in $\mathcal{D}_{\text{yes}}$ pick a random linear transformation $L : \{0,1\}^n \to \{0,1\}^n$ of full rank and let $g_Y(x) \triangleq f(Lx)$.

$\mathcal{D}_{\text{no}}$: To generate a random function g_N in $\mathcal{D}_{\text{no}}$ pick a random linear transformation $R : \{0,1\}^n \to \{0,1\}^n$ of rank exactly k and let $g_N(x) \triangleq f(Rx)$.

Note that $\mathcal{D}_{\text{yes}}$ is a distribution supported on the set of those functions which are linearly isomorphic to f. On the other hand, $\mathcal{D}_{\text{no}}$ is supported on those functions that are ϵ-far from linearly isomorphic to f. This is because if $g_N \in \mathcal{D}_{\text{no}}$, then there exists a linear transformation R of rank k such that $g_N(x) = f(Rx)$, so g_N has Fourier dimension k.

Let $Q \subseteq \{0,1\}^n$ and let $q_1, \ldots, q_t$ be a basis of the span of Q. Note that when L is a random linear transformation of full rank then $L(q_1), L(q_2), \ldots, L(q_t)$ are linearly independent. In fact given any t linearly independent vectors $v_1, \ldots, v_t$,

$$\Pr_L[\forall i \ , \ L(q_i) = v_i] = 1/M,$$

where M is the number of distinct sets of t independent vectors.

When R is a random linear transformation of rank k the set of vectors $\{R(q_1), \ldots, R(q_t)\}$ need not be linearly independent in general, but if $t < k$ they are independent with high probability.

Lemma H.5 *If $\{q_1, \ldots, q_t\}$ is a set of linearly independent vectors then when R is a random linear transformation of rank k, then with probability $1 - 1/2^{k-t}$, the set $\{R(q_1), \ldots, R(q_t)\}$ is linearly independent.*

Proof. Let assume that the set $\{R(q_1), \ldots, R(q_t)\}$ is not linearly independent. So there must be a linear combination of the vectors that add up to zero. That is there must be $a_1, \ldots, a_t \in \{0,1\}$ such that $\sum_{i=1}^{t} a_i R(q_i) = 0$. In other words, there exist a vector v in the span of $q_1, \ldots, q_t$ such that $R(v) = 0$.

Because R is a randomly chosen linear transformation of rank k,

$$\forall v \in \{0,1\}^n, \ \Pr_R[R(v) = 0] = \frac{1}{2^k}.$$

So the expected number of vectors in the span of $q_1, \ldots, q_t$ such that $R(v) = 0$ is $1/2^{k-t}$. And thus by Markov's Inequality,

$$\Pr[R(q_1), R(q_2), \ldots, R(q_t) \text{ are linearly independent}] \geq 1 - \frac{1}{2^{k-t}}.$$

□

In fact, conditioned on the event that $\{R(q_1), \ldots, R(q_t)\}$ are linearly independent, any set of linearly independent vectors is equally likely. Thus, for any t linearly independent vectors $v_1, \ldots, v_t$,

$$\Pr_R[\forall i \ R(q_i) = v_i \mid R(q_1), R(q_2), \ldots, R(q_t) \text{ are linearly independent}] = 1/M = \Pr_L[\forall i \ , \ L(q_i) = v_i].$$

And since Q is contained in the span of $q_1, \ldots, q_t$, for all $a \in \{0,1\}^{|Q|}$ we get

$$\begin{aligned}\Pr_{g_N \leftarrow \mathcal{D}_{\text{no}}}\left[g_N\restriction_Q = a\right] &\geq (1-2^{t-k}) \Pr_{g_N \leftarrow \mathcal{D}_{\text{no}}}\left[g_N\restriction_Q = a \mid \{R(q_1), \ldots, R(q_t)\} \text{ is linearly independent}\right] \\ &= (1-2^{t-k}) \Pr_{g_Y \leftarrow \mathcal{D}_{\text{yes}}}\left[g_Y\restriction_Q = a\right].\end{aligned}$$

Therefore, if $t \leq k-2$ we have $\Pr_{g_N \leftarrow \mathcal{D}_{\text{no}}}\left[g_N\restriction_Q = a\right] \geq (3/4)\Pr_{g_Y \leftarrow \mathcal{D}_{\text{yes}}}\left[g_Y\restriction_Q = a\right]$. By Lemma A.1, if f is ϵ-far from having Fourier dimension k then any tester (even an adaptive one) for testing linear isomorphism to f must make at least $k-1$ queries. □

Remark H.1 *Gopalan et al [GOS+09] proved that if $f : \{0,1\}^n \to \{0,1\}$ is $O(2^{-k-1})$-close to having Fourier dimension k then testing linear-isomorphism to f can be done using $O(k2^k)$ queries.*