Query-efficient computation in property testing and learning theory

García Soriano, D.

Chapter 4

Junto-symmetric functions and hypergraph isomorphism

Now we touch upon the question of when it is possible to test isomorphism with constantly many queries. We prove a characterization of the class of hypergraphs of constant arity (rank) to which isomorphism can be efficiently tested, and make a step towards obtaining a similar characterization for general hypergraphs and boolean functions.

The content of this chapter is based on the paper


4.1 The size of invariance groups

The automorphism group of a function \( f \), also known as its symmetry group or invariance group, is the group of permutations that leave \( f \) invariant:

\[
\text{Aut}(f) \triangleq \{ \pi \in S_n \mid f^{\pi} = f \}.
\]

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

\[
\text{DifPerm}(f) = \{[\pi_1], \ldots, [\pi_t]\}
\]

be the equivalence classes formed. There is a bijection between \( \text{DifPerm}(f) \) and the set \( S_n : \text{Aut}(f) \) of cosets of \( \text{Aut}(f) \); therefore the number \( |\text{DifPerm}(f)| = |\text{Isom}(f)| \) of distinct permutations of \( f \) is equal to the index of \( \text{Aut}(f) \) in \( S_n \), i.e., \( |\text{DifPerm}(f)| = |S_n : \text{Aut}(f)| = n!/|\text{Aut}(f)| \). The size of \( \text{Aut}(f) \) is a rough measure of the amount of symmetry that \( f \) possesses: the larger \( \text{Aut}(f) \), the more
symmetric \( f \) is. A symmetric function satisfies \( \text{Aut}(f) = S_n \) and \( |\text{DifPerm}(f)| = 1 \), whereas a random function has, with high probability, a trivial automorphism group \( \text{Aut}(f) = \{1\} \) and \( |\text{DifPerm}(f)| = n! \) (for example, see [Cla92] for a simple proof of a stronger statement).

Not every group \( G \leq S_n \) can arise as the automorphism group of a boolean function on \( n \) variables; those which can are called \( 2\text{-representable} \). For example, it is not hard to argue that if the alternating group \( A_n (n \geq 3) \) is contained in \( \text{Aut}(f) \), then \( \text{Aut}(f) \) is indeed the whole of \( S_n \); as a result, \( A_n \) is not \( 2\text{-representable} \). 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 \( \pi \) 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 \( \pi \in A_n \subseteq \text{Aut}(f) \) and so \( f(x) = f(y) \). Hence \( A_n \leq \text{Aut}(f) \) implies \( f(x) = f(y) \) for all \( |x| = |y| \), so \( f \) is actually symmetric.

The groups \( G \leq S_n \) that can be represented as \( \text{Aut}(f) \) for some \( k \)-valued function \( f : \{0,1\}^n \to [k] \) are called \( k\text{-representable} \). The representability of \( k\)-valued functions and some generalizations are studied in [CK91, Kis98, Xia05] (see also Chapter 3 of [CK02]). A neat paper of Babai, Beals and Takács-Nagy [BBTN92] exposes a relationship between the circuit complexity of a function \( f \) and the number of orbits of the action of \( \text{Aut}(f) \) on \( \{0,1\}^n \).

We know that \( f\text{-isomorphism} \) can always be tested with \( O(\log |\text{DifPerm}(f)|) \) queries for constant \( \varepsilon \) (Proposition 2.3.2), 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 that \( \text{DifPerm}(f) \) can have for a non-symmetric function \( f \)? A moment’s thought reveals that there are non-symmetric functions with only \( n \) different permutations, like any dictatorship \( f(x_1x_2 \ldots x_n) = x_i \), and indeed this can be shown to be best possible.\(^1\)

4.1.1. PROPOSITION. If \( f : \{0,1\}^n \to \{0,1\} \) is not symmetric and \( n \geq 5 \), then \( |\text{DifPerm}(f)| \geq n \).

Proof. The elements of \( S_n \) act on \( \text{DifPerm}(f) \) by multiplication in a natural way: for each \( \pi \in S_n \) we define a permutation \( \phi(\pi) \) of \( \text{DifPerm}(f) \) by

\[
\phi(\pi)([\sigma]) = [\pi \circ \sigma]
\]

The map \( \phi : S_n \to \text{Sym}(\text{DifPerm}(f)) \) is well-defined since \( [\sigma_1] = [\sigma_2] \) implies \( f^{\sigma_1} = f^{\sigma_2} \) and hence \( f^{\pi \circ \sigma_1} = f^{\pi \circ \sigma_2} \), so \( [\pi \circ \sigma_1] = [\pi \circ \sigma_2] \). Moreover, it is a group homomorphism (where the product operation on both \( S_n \) and \( \text{Sym}(\text{DifPerm}(f)) \) is the usual composition of permutations); this is because \( \phi(1) = 1 \), and \( \phi(\pi_1) \circ \phi(\pi_2) = \phi(\pi_1 \circ \pi_2) \). Therefore its kernel \( \ker \phi \) is a normal subgroup of \( S_n \). The

\(^1\)The claim fails for \( n = 4 \): the function \( f(a,b,c,d) = (a \land b) \lor (c \land d) \) has three different permutations.
only normal subgroups of $S_n$ ($n \geq 5$) are $1, A_n$ and $S_n$ \[\text{Art10} \text{, Theorem 7.4.4}\]. Clearly $\ker \phi \leq \Aut(f)$ and since $\Aut(f)$ does not contain $A_n$ (or else $f$ would be symmetric as argued before), it follows that $\ker \phi = 1$, so $\phi$ is injective. Since the domain of $\phi$ is $S_n$, its image $\Sym(\Isom(f))$ must be at least as large, hence $|\DifPerm(f)| \geq n$.

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 \[\text{PRS02}\], and more generally to $O(1)$-juntas \[\text{FKR}^+04\]. 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 $n$, $\chi_{[n-t]}$. The identity $\chi_{[n-t]}(x) = \chi_{[n]}(x) \oplus \chi_{[n] \setminus [n-t]}(x)$ makes it possible to transform the responses to all queries made for the $t$-junta $\chi_{[n] \setminus [n-t]}$ into the responses to queries 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 $O(t)$ queries. In the same vein, the majority function on the first $n-t$ variables $\text{Maj}_{[n-t]}$ (for $n$ large enough and $t \ll \sqrt{n}$) is very close to the symmetric majority $\text{Maj}_{[n]}$, and it is not hard to see that the standard constant-query test for equality between the tested function and $\text{Maj}_{[n]}$ yields a tester for isomorphism to $\text{Maj}_{[n-t]}$ as well (because its queries are uniformly distributed).

We introduce a notion generalizing all these cases.

4.1.2. DEFINITION. 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

$$f(x) = \tilde{f}(|x|, x|_J)$$

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

The function $\tilde{f}$ is called $k$-junto-symmetric if it is $J$-junto-symmetric on some subset $J$ of size $k$.

The function $\tilde{f}$ above is not completely determined by $f$ on inputs of very small or high weight. For example, let $f$ be 1-junto-symmetric. Then one can define $\tilde{f}(0,1)$ in two different ways that give rise to the same function $f$.

Let $JS_J$ denote the class of $J$-junto-symmetric functions, and $JS_k$ the $k$-junto-symmetric functions. Note that the definition necessitates that 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 $f$.

\footnote{The symbol $\chi$ is usually reserved to a parity taking values in $\pm 1$ so it is a character of $\mathbb{Z}_2^n$, but here we use it for $\{0,1\}$-valued functions.}
Observe that any symmetric function is 0-junto-symmetric, and any $k$-junta is $k$-junto-symmetric. At the other extreme, every function is $(n - 1)$-junto-symmetric. Additional examples of $k$-junto-symmetric functions are $\chi_{[n-k]}$ and $\text{Maj}_{[n-k]}$; in fact, the reader may verify that any $k$-junta whose core function is symmetric must be $\min(k, n - k)$-junto-symmetric.

4.1.3. Definition. Let $F$ denote a sequence $f_1, f_2, \ldots$ of boolean functions with $f_n: \{0, 1\}^n \to \{0, 1\}$ for each $n \in \mathbb{N}^+$. We say that $F$ is an $O(1)$-junto-symmetric family if there exists a constant $k$ such that each $f_i$ is $k$-junto-symmetric.

Interestingly, $O(1)$-junto-symmetric functions were studied by Shannon under the name “partially symmetric functions” [Sha49].

The size of $\text{DifPerm}(f)$ for any $k$-junto-symmetric $f$ is upper-bounded by $\binom{n}{k} k!$, because if $f$ can be written in the form (4.1), then for any $\pi \in S_n$ there is a $k$-subset $T \subseteq [n]$ and a permutation $\sigma \in \text{Sym}(T) \cong S_k$ such that $f^\pi(x) = f(|x|, (x|_T)^\sigma)$. This quantity is $n^{O(1)}$ for constant $k$. Families like this were given a name in [PS10]:

4.1.4. Definition. The family $F$ is poly-symmetric if there exists a constant $c$ such that $|\text{DifPerm}(f_n)| \leq n^c$ for all $n$.

We will occasionally speak of such a family as an $O(1)$-junto-symmetric (or poly-symmetric function when the intended meaning is clear. As it turns out, the two notions just described are the same.

4.2 Characterizing $O(1)$-junto-symmetry

In the following we identify elements of $\text{Sym}(T)$, $T \subseteq [n]$, with elements of $\text{Sym}([n])$ that act as the identity outside $T$.

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

(a) $F$ is a poly-symmetric family;

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

(c) $F$ 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 $n$).
To ease readability, we drop the subscripts in the proof, i.e., write $f$ and $A$ in place of $f_n$ and $A_n$. All but one of the implications we need are straightforward:

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

- $(c) \implies (d)$: Let $f = \check{f}(|x|, x|_A)$, $|A| = k = O(1)$. Define $\hat{f}^{(i)}(x) = \check{f}(i, x|_A)$. Each $\hat{f}^{(i)}$ is a junta on $A$. The number of $A$-juntas is only $\ell = 2^k = O(1)$; let $j_1, \ldots, j_{\ell}$ be an enumeration of them and let

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

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

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

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

- $(d) \implies (b)$: Let $f(x) = \check{f}(s_1(x), \ldots, s_\ell(x), x_{i_1}, \ldots, x_{i_k})$, where $s_1, \ldots, s_\ell$ are symmetric. Set $A = \{i_1, \ldots, i_k\}$ and let $\pi \in Sym([n] \setminus A)$. 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 $Sym([n] \setminus A) \leq Aut(f)$.\footnote{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”.

\footnote{This would be implied by the claim following Theorem 28 on 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. However, the lemma does hold for the automorphism groups of boolean functions however as we show. This is the case of interest in their paper and in this thesis.}}

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

The only remaining implication, which will be shown next\footnote{This is the case of interest in their paper and in this thesis.} is $(a) \implies (b)$. 

\[4.2. \text{Characterizing } O(1)-\text{junto-symmetry} \]
4.2.1 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 \( \Omega \) be a set (which will be assumed finite here, and will often be equal to \([n]\) in our applications). \( \text{Sym}(\Omega) \) denotes the symmetric group of all permutations of \( \Omega \), and \( \text{Alt}(\Omega) \) is the subgroup of \( \text{Sym}(\Omega) \) made up of even permutations. When \(|\Omega| = n\), we occasionally write \( A_n = \text{Alt}(\Omega) \) and \( S_n = \text{Sym}(\Omega) \). The product operation we use in \( \text{Sym}(\Omega) \) is \( \pi \sigma \triangleq \sigma \circ \pi \).

A permutation group \( G \) on \( \Omega \) is a subgroup of \( \text{Sym}(\Omega) \), written \( G \leq \text{Sym}(\Omega) \).

The image \( \pi(x) \) of \( x \in \Omega \) under \( \pi \in G \) is often written \( x^\pi \); under our convention we have \( (x^\pi)^\sigma = x^{\pi \sigma} = x^{\sigma \pi} \) for \( \pi, \sigma \in G \). The orbit of a set \( \Delta \subseteq \Omega \) under an arbitrary collection \( H \subseteq G \) is the set \( \Delta^H = \{ x^{\pi} \mid \pi \in H, x \in \Delta \} \). When \( \Delta = \{ x \} \) or \( H = \{ h \} \) are singletons we may simply write \( x^H \) or \( \Delta^h \).

\( G \) is called transitive if for every \( x, y \in \Omega \) there is \( \pi \in G \) with \( x^\pi = y \). An intransitive group \( G \leq \text{Sym}(\Omega) \) partitions \( \Omega \) into orbits: these can be characterized as 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 \( \Omega \) is a homomorphism \( \phi : G \to \text{Sym}(\Omega) \). (This is what is called a right action because of our convention on the composition law in \( \text{Sym}(\Omega) \).) If \( \ker \phi = 1_G \), the action is faithful and \( G \) is isomorphic (via \( \phi \)) to a permutation group on \( \Omega \). It is customary to omit the explicit reference to the chosen \( \phi \) and write \( x^g \) for \( x^{\phi(g)} \) (\( g \in G \)). Given an action of \( G \) on \( \Omega \), we can naturally extend it to define an action on subsets of \( \Omega \): \( g \in G \) acts on \( \mathcal{P}(\Omega) \) by mapping \( \Delta \subseteq \Omega \) to \( \Delta^g \) as defined above.

A block of \( G \) is a subset \( \Delta \) of \( \Omega \) such that for every \( \pi \in G \), either \( \Delta^\pi = \Delta \) or \( \Delta^\pi \cap \Delta = \emptyset \). Evidently, \( \Omega \), the empty set \( \emptyset \) and each of the singletons \( \{ i \} \in \Omega \) are always blocks; we call these the trivial blocks. The permutation group \( G \) is said to be primitive (group) if it is transitive and has no non-trivial blocks. (Only transitive groups are classified as being primitive or imprimitive.) The intersection of any pair of blocks is itself a block. If \( \Delta \) is a block of \( G \), then \( \Omega \) can be partitioned into a complete block system, where every block is of the form \( \Delta^g \) for some \( g \in G \) (so all blocks in a complete block system have the same cardinality). Any element of \( G \) permutes the blocks in a complete block system among themselves, and also the elements inside each block.

The pointwise stabilizer of \( \Delta \subseteq \Omega \) with regard to \( G \) is the set
\[
G_\Delta \triangleq \{ \pi \in G \mid x^\pi = x \ \forall x \in \Delta \}.
\]

4.2.2 Proof that poly-symmetric \( \equiv O(1) \)-junto-symmetric

First we need a handy result that provides a lower bound for the index of primitive groups. The proof can be found in [Wie64, Theorem 14.2] (asymptotically better bounds are available [Bab81, Cam81], but this one will suffice).
4.2. Characterizing $O(1)$-junto-symmetry

4.2.2. Theorem (Bochert’s bound [Boc89]). Let $G$ be a primitive subgroup of $S_n$, other than $S_n$ and $A_n$. Then

$$[S_n : G] \geq \lfloor n/2 \rfloor !.$$  

4.2.3. Lemma. Let $n \geq 14$, $G \leq S_n$, $G \neq S_n, A_n$. Then

(a) If $G$ is transitive then

$$[S_n : G] \geq \frac{1}{2} \left( \frac{n}{\lfloor n/2 \rfloor} \right).$$  

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

$$[S_n : G] \geq \left( \frac{n}{\max(n/2, \ell)} \right).$$

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

$$H \triangleq G \cap \text{Sym}(\Delta) = G \cap S_\ell$$  
be the pointwise stabilizer of $[n] \setminus \Delta$ (we identify $\text{Sym}(\Delta)$ with $S_\ell$). Then

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

Proof.

(a) If $G$ is primitive, Bochert’s theorem states the bound $[S_n : G] \geq \lfloor n/2 \rfloor !$, which is stronger for $n \geq 14$. So suppose $G$ is transitive and imprimitive, with a block of imprimitivity of size $a$ ($2 \leq a \leq n/2, a \mid n$), and hence $b = n/a \geq 2$ such blocks because of transitivity (see Section 4.2.1). Then

$$|G| \leq (a!)^b b! \leq 2 [(ab/2)]! [(ab/2)]! = 2 [n/2] [n/2],$$

The first inequality holds because there are $b!$ ways of permuting the blocks among themselves, and $a!$ ways of permuting the elements inside a given block. To prove the last inequality, observe that for $a = 2$ it reduces to the triviality $b! \geq 2^{b-1}$. Hence it suffices to verify that for any $b \geq 2$, the quotient

$$q(a) \triangleq \frac{a! b}{[(ab/2)]! [(ab/2)]!}$$

is a decreasing function of $a$. Writing the factors in the numerator and denominator in decreasing order, we have

$$q(a) = \frac{a \cdot a \cdots a (a-1) \cdot (a-1) \cdots (a-1) \cdots}{[ab/2] [ab/2] [ab/2-1] [ab/2-1] \cdots}$$
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

\[
q(a) = \prod_{i=1}^{ab} \frac{s_i}{t_i} = \prod_{i=1}^{(a-1)b} \frac{s_i}{t_i} \cdot \prod_{j=(a-1)b+1}^{ab} \frac{s_j}{t_j}
\]

\[
= q(a - 1) \cdot \prod_{j=(a-1)b+1}^{ab} \frac{a}{t_j}
\]

\[
\leq q(a - 1),
\]

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

\[
[S_n : G] = \frac{n!}{|G|} \geq \frac{1}{2} \left( \frac{n}{[n/2]} \right).
\]

(b) Let \( A_1, \ldots, A_m \) (\( m \geq 2 \)) be the orbits and \( a_i = |A_i| \). Since \( G \) only maps elements of \( A_i \) to elements of the same \( A_i \), we have \( G \leq \text{Sym}(A_1) \times \text{Sym}(A_2) \times \cdots \times \text{Sym}(A_m) \) and therefore

\[
|G| \leq \prod_{i \in [m]} a_i!.
\]

Fix \( n > \ell > 0 \) and let us consider

\[
r_n(\ell) \triangleq \max \{ \prod_{i \in [m]} a_i! \mid m \geq 2, a_i \in \mathbb{N}, 0 \leq a_i \leq \ell, \sum_{i \in [m]} a_i = n \}
\]

Consider the expression inside the maximum in the definition of \( r_n(\ell) \). Without loss of generality, we can take \( m = n \). We claim that it attains its maximum for some solution with \( a_i = \ell \) for at least one \( i \). Take any optimal solution and sort the values in non-increasing order: \( a_1 \geq a_2 \geq \cdots \geq a_t > a_{t+1} = \ldots = a_m = 0 \). If \( a_1 = \ell \) we are done. Otherwise \( a_1 < \ell \) and we must have \( t > 1 \) (with \( a_t > 0 \)) since the total sum is at least \( \ell \). If we replace the pair \((a_1, a_t)\) with \((a_1 + 1, a_t - 1)\) we obtain a feasible solution, and \( \prod a_i! \) increases since \( a_1! a_t! < (a_1 + 1)! (a_t - 1)! \) (as \( a_1 + 1 > a_t \)). This is not possible for an optimal solution, so there is no such pair, meaning that \( a_1 = \ell \).

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

\[
r_n(\ell) \leq \ell!(n-\ell)! = \frac{n!}{\binom{n}{\ell}}
\]
4.2. Characterizing $O(1)$-junto-symmetry

for any $\ell$. When $\ell$ is the size of the largest orbit, we have $|G| \leq r_n(\ell)$, and this shows that

$$[S_n : G] \geq \frac{n!}{r_n(\ell)} \geq \binom{n}{\ell},$$

On the other hand, $r_n(\ell)$ is by definition an increasing function of $\ell$, so the inequality

$$[S_n : G] \geq \frac{n!}{r_n(\ell)} \geq \frac{n!}{r_n(n/2)} \geq \binom{n}{\lfloor n/2 \rfloor}$$

holds when the size of the longest orbit is $\ell \leq n/2$.

(c) Because $G \leq H \times \text{Sym}(\mathbb{N} \setminus \Delta)$, we can bound

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

which yields

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

4.2.4. **Lemma.** Let $n \geq 14$, $t \leq n/2$, $[S_n : G] < \frac{1}{2}\binom{n}{t}$, and $\Delta, \ell$ as before. Then $\ell > n - t$ and $\text{Alt}(\Delta) \leq G$.

**Proof.** If the action of $G$ is transitive on $[n]$ then $[S_n : G] \geq \frac{1}{2}\binom{n}{\lfloor n/2 \rfloor}$ by Lemma 4.2.3(a), which contradicts our assumptions. So $G$ is not transitive and $\ell < n$. If $\ell \leq n/2$ we have, by Lemma 4.2.3(b), $[S_n : G] \geq \binom{n}{\lfloor n/2 \rfloor}$, which again is impossible.

We are left with the case $n/2 < \ell < n$. In accordance with Lemma 4.2.3(b),

$$\binom{n}{t} > \frac{[S_n : G]}{\binom{n}{\ell}} = \frac{n!}{(n-\ell)!},$$

so $t > n - \ell$ (since $n - \ell, t \leq n/2$). Let $H = G \cap S_\Delta$. This is actually the pointwise stabilizer of $[n] \setminus \Delta$ in $G$, and since $\Delta$ is an orbit of $G$ it follows that $H$ is normal in $G$ [Wie64, Proposition 3.1]. We demonstrate that $A_\Delta \leq H$ by contradiction.

So assume $H \neq \text{Sym}(\Delta), \text{Alt}(\Delta)$. Then Lemma 4.2.3 applies to the group $H$ acting on $\Delta$. Let $\Delta'$ be the largest orbit of this action and $\ell' = |\Delta'|$. Since $G$ is transitive on $\Delta$ and $H < G$, it is not hard to see that the length of any orbit of $H$ on $\Delta$ must divide $\ell$, i.e., $\ell' | \ell$. 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 4.2.3.
Chapter 4. Junto-symmetric functions and hypergraph isomorphism

• 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 \( \Delta \) and

\[
[S_\ell : H] \geq \frac{1}{2} \left( \frac{\ell}{\ell/2} \right)
\]

by part (a) of the “inner” application of Lemma 4.2.3.

In any case we have

\[
[S_\ell : H] \geq \frac{1}{2} \left( \frac{\ell}{\ell/2} \right).
\]

Together with part (c) of the “outer” application, i.e.,

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

this yields

\[
[S_n : G] \geq \frac{1}{2} \left( \frac{\ell}{\ell/2} \right) \binom{n}{\ell}.
\]

Now we bound each of these two factors. Using the inequality

\[
\binom{2(m+1)}{m+1} = 2 \binom{2m}{m} \frac{2m+1}{m+1} \leq 4 \binom{2m}{m},
\]

it is possible to show that

\[
\left( \frac{\ell}{\ell/2} \right) \geq \frac{1}{2^{n-\ell}} \binom{n}{n/2}.
\]

Using the fact that \( \ell > n/2 \), we get

\[
\binom{n}{\ell} = \binom{n}{n-\ell} \geq \left( \frac{n}{n-\ell} \right)^{n-\ell} \geq 2^{n-\ell}.
\]

Multiplying these two bounds we obtain the contradiction

\[
[S_n : G] \geq \frac{1}{2} \left( \frac{\ell}{\ell/2} \right) \binom{n}{\ell} \geq \frac{1}{2} \binom{n}{n/2}.
\]

4.2.5. COROLLARY. Let \( n \geq 14 \) and \( f : \{0,1\}^n \to \{0,1\} \) be a boolean function with \(|\text{DiffPerm}(f)| < \frac{1}{2}\binom{n}{t} \), \( t \leq n/2 \). Then there is a set \( \Gamma \) of size \( |\Gamma| < t \) such that \( f \) is junto-symmetric on \( \Gamma \). In particular, any poly-symmetric family is junto-symmetric on sets of size \( O(1) \).
4.3. Testers for junto-symmetric functions

**Proof.** Let $G = \text{Aut}(f)$. Since $|\text{DiffPerm}(f)| = [S_n : G]$, the previous lemma states that if $\Delta$ is the largest orbit, then $|\Delta| \geq n/2 \geq 5$ and $\text{Alt}(\Delta) \leq \text{Aut}(f)$. We show that this means that $f$ is junto-symmetric on $\Gamma \triangleq [n] \setminus \Delta$. Indeed, for any $x \in \{0, 1\}^\Gamma$, we can define a boolean function $g_x : \{0, 1\}^\Delta \to \{0, 1\}$ by $g(z) = f(z \sqcup x)$; then $\text{Alt}(\Delta) \leq \text{Aut}(f) \cap \text{Sym}(\Delta) \leq \text{Aut}(g_x)$, so $g_x$ is a boolean function on more than 4 variables whose automorphism group contains the alternating group. Hence $g_x$ is actually symmetric for all $x$, and $f$ is junto-symmetric on $\Gamma$.

This corollary is the last piece we needed to show Theorem 4.2.1.

4.3 Testers for junto-symmetric functions

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

4.3.1. **Theorem** ([CFGM12, BWY11]). Let $\varepsilon > 0$ and $1/\varepsilon^{1/4} < k < (2n)^{1/12}$. Let $f : \{0, 1\}^n \to \{0, 1\}$ and denote $f^* \in J^k$ the $k$-junto-symmetric function closest to $f$.

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

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

**soundness** If $\text{dist}(f, f^*) \geq \varepsilon$, the algorithm rejects with probability $\geq 2/3$.

See Section 4.3.3 for the proof.

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

4.3.2. **Theorem.** ([CFGM12, BWY11]) Let $k, \varepsilon, f$ as before. There is a poly($k/\varepsilon$)-query $\varepsilon$-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 Section 4.3.4.

4.3.3. **Corollary.** Isomorphism to any poly-symmetric function can be $\varepsilon$-tested with $\text{poly}(1/\varepsilon)$ 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 Section 4.3.5).
4.3.4. Theorem. There is a constant $0 < c < 1$ with the following property. Let $k, \varepsilon, f$ as before.

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

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

In an independent work simultaneous with ours, Blais, Weinstein and Yoshida have also proven the results stated in this section [BWY11]. (Their query complexities are better and the restrictions on the size of $k$ are not present.)

It is possible to define a notion of “symmetric influence” that characterizes closeness to junto-symmetric functions up to a factor of two, just as influence does for closeness to juntas. The resulting definition does not enjoy the subadditivity property, which is crucial for the proofs of the standard junta testers (Section 3.6). Although this approach can be made to work with some technical work [BWY11], here we take a different route.

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.

4.3.1 Preliminary observations

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 $\ell$:

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

For any $B \in B_{\ell,x}$, either $x^B \subseteq x$ or $x \subseteq x^B$ holds, depending on whether $|x| \geq \ell$ or $|x| \leq \ell$. The set $B_{\ell,x}$ 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 $r(\ell,x) \in B_{\ell,x}$ for all $\ell, x$. We need a lemma concerning the probability that $B = r(\ell,x)$ happens to intersect some small set $A$, when $(\ell,r,x)$ 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 $\{0,1\}^n$. 

4.3. Testers for junto-symmetric functions

4.3.5. Lemma. Let \( A \subseteq [n] \). Then

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

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

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

On the other hand, the size of any element \( B \) of \( B_{\ell,x} \) is \( |\ell - |x|| \) by definition. We can write \( \ell = |y| \) for uniformly random \( y \in \{0, 1\}^n \), so \( \mathbb{E} [\|B\|] = \mathbb{E} [||x| - |y||] \).

Recalling that \( \mathbb{E} [\|x\|] = \mathbb{E} [||y||] = n/2 \), \( \mathbb{E} [|x|^2] = \mathbb{E} [|y|^2] = \text{Var} [\|x\|] + \mathbb{E} [|x|^2] = \frac{1}{4} n(n + 1) \) and applying Cauchy-Schwarz,

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

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

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

\[\square\]

Let us define a transformation \( T \) mapping each function \( f : \{0, 1\}^n \rightarrow \{0, 1\} \) to \( T(f) : \mathcal{L} \times \mathcal{R} \times \{0, 1\}^n \rightarrow \{0, 1\}^n \) 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 \( (\ell, x) \), one string \( x^{r(\ell,x)} \) of Hamming weight \( \ell \) with minimum distance to \( x \); the choice is independent of all choices for any other pair when \( r \) ranges uniformly over \( \mathcal{R} \).

We want to argue about \( T(f) \) as a function in its own right, on a larger set of variables. 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 \( f \). The reader who so wishes may think of \( T(f) \) as a function on \( \{0, 1\}^{\lceil \log(n+1) \rceil} \times \{0, 1\}^{\lceil \log |\mathcal{R}| \rceil} \times \{0, 1\}^n \), although this is not strictly necessary; in this case \( V_0, V_1, V_2 \) would indicate disjoint input bit variables with sizes

\[
|V_0| = \log(n + 1), |V_1| = \log |\mathcal{R}|, |V_2| = n
\]

(but note that the input distribution on \( \{0, 1\}^{V_0} \) is not uniform).
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 $\ell$ 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, i.e., the variables in $V_1$ are irrelevant to $\psi(g)$. The intuition is that $T$ maps junto-symmetric functions $f$ on $A$ into functions that are close to juntas on $(V_0 \cup A)$ because $V_1$ and $V_2 \setminus A$ will be nearly irrelevant to $T(f)$; while $\psi$ maps these functions on an extended domain that are juntas on $V_0 \cup A$ into junto-symmetric functions on $A$ defined on $\{0,1\}^n$.

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

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 $\{0,1\}^n \to \{0,1\}$;
- $j, j_1, j_2: \{0,1\}^n \to \{0,1\}$ are junto-symmetric on $A$;
- $j': \mathcal{L} \times \mathcal{R} \times \{0,1\}^n \to \{0,1\}$ is a member of $\text{Jun}_{V_0}(A)$;
- $\pi \in \text{Sym}(V_0) \times \text{Sym}(V_2)$ (we identify $\pi$ with an element of $\text{Sym}(V_2)$ as well).

4.3.6. **Lemma.** The mappings $T$ and $\psi$ satisfy the following properties:

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

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

$$\text{dist}(j', T(\psi(j'))) \leq \frac{|A|}{\sqrt{2n}}.$$

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

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

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

$$\text{distiso}(j_1, j_2) = \text{distiso}(\psi(j_1), \psi(j_2)).$$
4.3. Testers for junto-symmetric functions

(f) The bounds

\[ |\text{distiso}(f, g) - d| \leq \text{dist}(f, JS_k) + \text{dist}(g, JS_k) + \frac{2k}{\sqrt{2n}}. \]

hold for

\[ d \triangleq \min_{\pi \in V_0 \times V_1} \text{dist}(T(f)^\pi, T(g)). \]

Proof.

(a) For any \( \ell \), the distribution of \( x^{r(\ell, x)} \) for random \( x, r \) is uniform over all strings of weight \( \ell \). 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

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

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

\[ \text{dist}(\pi', T(\psi(\pi'))) = \Pr[\pi'(\ell, r, x) \neq \psi(\pi')(x^{r(\ell, x)}) = \pi'(\ell, \bullet, x^{r(\ell, x)})] \]

\[ \leq \Pr[r(\ell, x) \cap A \neq \emptyset] \]

\[ \leq \frac{|A|}{\sqrt{2n}} \]

by Lemma 4.3.5.

(c) This follows from (b) because any \( \pi \in JS(A) \) can be written in the form \( \psi(\pi') \) for some (in fact, many) \( \pi' \in \text{Jun}_k(V_0) \).

(d) Let \( \pi \) be \( k \)-junto-symmetric and \( \pi' \in \text{Jun}_k(V_0) \) with \( \psi(\pi') = \pi \). Then by the triangle inequality and parts (c) and (a),

\[ \text{dist}(T(f), \pi') \leq \text{dist}(T(f), T(\pi)) + \text{dist}(T(\pi), \pi') \leq \text{dist}(f, \pi) + \frac{k}{\sqrt{2n}}, \]

so \( \text{dist}(T(f), \text{Jun}_k(V_0)) \leq \text{dist}(f, JS_k) + k/(2\sqrt{n}) \). Likewise, if \( \pi' \) is a junta on \( V_0 \cup A \) where \( |A| = k \), then

\[ \text{dist}(f, \psi(\pi')) = \text{dist}(T(f), T(\psi(\pi'))) \]

\[ \leq \text{dist}(T(f), \pi') + \text{dist}(\pi', T(\psi(\pi'))) \]

\[ \leq \text{dist}(T(f), \pi') + \frac{k}{\sqrt{2n}}, \]

which proves the inequality \( \text{dist}(f, JS_k) \leq \text{dist}(T(f), \text{Jun}_k(V_0)) + k/(2\sqrt{n}) \).

(e) Clear.

(f) Follows from (d), (e) and the triangle inequality for \text{distiso}. 

\[ \square \]
4.3.2 Generalized junta testing

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

4.3.7. Lemma. For any product distribution $\mu$ and any constant $p < 1$, there is an algorithm

\begin{align*}
\text{GeneralizedJuntaTester}_{\mu, p}(f, k, \varepsilon, T)
\end{align*}

that, with probability at least $p$,

- accepts if $f \in \text{Jun}_k(T)$.
- rejects if $\text{dist}(f, \text{Jun}_k(T)) \geq \varepsilon$;
- makes $\Theta(k^4 \log(k+1)/\varepsilon)$ non-adaptive queries, and the marginal distribution of each query is $\mu$.

Note that standard junta testing corresponds to $T = \emptyset$.

Proof. All known junta testers can be used in a straightforward manner for this generalized property preserving the exact query complexity. One way to see this is to think about 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 $[m] \setminus T$.)

Recall from Section 3.6 that the non-adaptive junta tester produces a number of disjoints subsets $I_1, \ldots, I_r \subseteq [m]$ satisfying the property written on page 56. 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

- if $\text{Inf}_f(B \setminus A) \geq \varepsilon$ for all $A \subseteq B, |A| = k$, then at least $k + 1$ of the independence tests will be positive.

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

To adapt these ideas to our task, note that if $\text{dist}(f, \text{Jun}_k(T)) \geq \varepsilon$ then $\text{Inf}_f([m] \setminus (T \cup A)) \geq \varepsilon$ for any $A \in \binom{[m] \setminus T}{k}$ (Lemma 3.6.1). 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. \qed
4.3.3 Testing junto-symmetry

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

1. Let \( q = \theta(k^4 \log(k + 1)/\varepsilon) \) bound the query complexity of Step 3.

2. Make queries to \( T(f) \) to test that \( \text{Inf}_{T(f)}(V_1) < \frac{1}{18q} \) with confidence > \( 8/9 \) by performing an independence test (Lemma 3.2.3); take \( O(q) \) random pairs \((\ell, r, x), (\ell, r', x)\) and compare \( T(f) \) on them. If it isn’t, reject.

3. Reject iff \text{GENERALIZEDJUNTA TESTER}_{\mu, 8/9}(T(f), k, \varepsilon/5, V_0 \cup V_1) \) rejects.

**Proof of Theorem 4.3.1.** The algorithm is clearly non-adaptive and its query complexity is \( \Theta(q) = \Theta(k^4 \log(k + 1)/\varepsilon) \). We assume that \( n \) is large enough for \( 2k/\sqrt{2n} < 1/(18q) < \varepsilon/5 \) to hold (small constant values for \( n \) can be dealt with separately in the tester).

The probability that an incorrect assessment is given by either the junta tester in step 3 or the influence test in step 2 is less than \( 2/9 < 2/3 \). So if the overall test accepts with probability \( \geq 2/3 \), then \( T(f) \) must be \( \varepsilon/5 \)-close to a junta \( f' \) on \( V_0 \cup V_1 \cup A \), \( |A| \leq k \). In particular \( \text{Inf}_{T(f)}(V_2 \setminus A) \leq \varepsilon/5 \). Moreover, since the influence test succeeded we also have \( \text{Inf}_{T(f)}(V_1) < \varepsilon/5 \). Therefore \( \text{Inf}_{T(f)}(V_1 \cup (V_2 \setminus A)) \leq 2\varepsilon/5 \), which means (by Lemma 3.6.1) that \( T(f) \) is in fact \( 4\varepsilon/5 \)-close to a junta on \( V_0 \cup A \). Consequently, \( f \) is \( 4\varepsilon/5 + k/\sqrt{2n} < \varepsilon \)-close to junto-symmetric on \( A \) (Lemma 4.3.6), proving soundness.

On the other hand, suppose \( f \) is \( 1/(18q) \)-close to a junto-symmetric function \( f' \). Then there is \( f' \in \text{Jun}_k(V_0) \) with \( \text{dist}(T(f), f') \leq 1/(18q) + k/\sqrt{2n} < 1/(9q) \). Recall that every query of the junta tester to \( T(f) \) follows the distribution \( L \times R \times \{0, 1\}^n \) (third item of Lemma 4.3.7), and this translates into uniform queries to \( f' \) (we showed that \( x^{r}(l, x) \) is uniformly distributed during the course of the proof of Lemma 4.3.6(a)). As the tester is non-adaptive, this means that the expected number of queries exposing a difference between \( T(f) \) and \( f' \) is \( 1/9 \), so with probability \( 8/9 \) the tester can’t see the difference between \( T(f) \) and \( f' \). Hence we are effectively testing \( f' \) 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 \( \text{Inf}_{f}(V_1) = 0 \), we also have \( \text{Inf}_{T(f)}(V_1) \leq 2 \cdot \text{dist}(T(f), f') \leq 2k/\sqrt{2n} < 1/(18q) \) and step 2 also accepts with probability \( 8/9 \). This establishes completeness.

4.3.4 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 \( JS_k \)) to testing isomorphism between \( k \)-juntas. For this we can use a tolerant tester of isomorphism, except that, in view of
Chapter 4. Junto-symmetric functions and hypergraph isomorphism

Lemma 4.3.6(e), the set of permutations allowed must be restricted to those fixing \(V_0\) and \(V_1\):

1. Use the algorithm of Theorem 4.3.1 to accept if \(f \in JS_k\) and reject if \(\text{dist}(f, JS_k) > \varepsilon/30\).

2. Perform a suitable test to accept if \(d \leq \varepsilon/10\) and reject if \(d \geq 9\varepsilon/10\), where
   \[
   d \triangleq \min_{\pi \in V_0 \times V_1 \times \text{Sym}(V_2)} \text{dist}(T(f), T(g)^\pi)
   \]

Ignoring for the moment the implementation details of the second test, we show that the algorithm outlined is an isomorphism tester for \(JS_k\):

**Proof of Theorem 4.3.2.** We use the algorithm just described. The claim about the query complexity is clear.

Suppose the test accepts with high probability. Then \(\text{dist}(f, JS_k) \leq \varepsilon/30\) and \(d \leq 9\varepsilon/10\). Since \(\text{distiso}(g, JS_k) \leq 1/k^5\), we have
\[
|\text{distiso}(f, g) - d| \leq \varepsilon/30 + 1/k^5 + 2k/\sqrt{2n} \leq \varepsilon/20,
\]
so \(\text{distiso}(f, g) < \varepsilon\), as it should.

On the other hand, if \(f \simeq g\) then \(\text{dist}(f, JS_k) = \text{dist}(g, JS_k) < 1/k^5\) and \(d \leq 2/k^5 + (2k)/\sqrt{2n} < 1/k^4\), meaning that both tests succeed. If \(\text{distiso}(f, g) < 1/k^5\), then it also accepts with high probability because we can argue as before that since the test makes \(O(k^4)\) queries that are individually uniformly distributed.

Step 2 can be implemented using sample extractors. Let \(D = V_0 \times V_1\), \(f : D \times \{0, 1\}^n \to \{0, 1\}\) and let \(f' \in \text{Jun}_D(A)\), \(A \in \left(\binom{n}{k}\right)\) be the element of \(\text{Jun}_k(D)\) closest to \(f\). Define \(\text{core}_{k, D}(f') : D \times \{0, 1\}^k \to \{0, 1\}\) by
\[
\text{core}_{k, D}(f')(x \upharpoonright D, x \upharpoonright A) = f'(x).
\]

A correct sample for \(\text{core}_{k, D}(f')\) (with respect to \(\sigma \in 1_D \times S_k\)) is a pair \((x, a)\) with \(x \in D \times \{0, 1\}^k\) and \(\text{core}_{k, D}(f)(x^\sigma) = a\). An \(\eta\)-noisy sampler for \(\text{core}_{k, D}(f')\) 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 \(\sigma\), and \(x\) follows the distribution \(D \times \{0, 1\}^k\).

The following two lemmas are all we need.

**4.3.8. Lemma.** Suppose \(\text{dist}(f, \text{Jun}_k(D)) < 1/k^5\). Then there is a \(\text{poly}(k, 1/\varepsilon)\)-query non-adaptive algorithm to construct an \(\varepsilon/100\)-noisy sampler for \(\text{core}_{k, D}(f')\).

**Proof (sketch).** This is essentially Theorem 3.5.2. We need two changes. The first is that we substitute the adaptive junta tester for the junta tester used in
4.4. Hypergraph isomorphism

the proof. The second one is the observation that we know how the variables in \( D \) map to the variables in \( \text{core}_{k,D}(j') \), so for any \( z \in \{0,1\}^n \), we only need to “extract” the setting of the \( k \) relevant variables sitting outside \( A \).

4.3.9. Lemma. Let \( f, g : D \times \{0,1\}^n \rightarrow \{0,1\} \), \( g \in \text{Jun}_k(D) \). Write\n
\[
d = \min_{\pi \in 1_{D} \times S_n} \text{dist}(f, g^\pi)
\]

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

Proof (sketch). This is essentially Lemma \[3.5.4\]. Construct a sample for \( \text{core}_{k,D}(j') \) and take \( O(\log k!/\varepsilon^2) = O(k \log k/\varepsilon^2) \) random samples. These are enough to estimate\n
\[
d' = \min_{\pi \in S_k} \text{dist}(\text{core}_{k,D}(j'), \text{core}_{k,D}(g)^\pi)
\]

to within \( O(\varepsilon) \) additive error. Finally recall that \( d' \) and \( d \) are the same up to constant factors (this follows from Lemma \[3.3.1\]).

4.3.5 Tolerant testers

Using the tolerant tester in the third item of Theorem \[3.6.8\] instead of the poly\((k/\varepsilon)\)-query junta tester in the proof of the previous theorems, we obtain Theorem \[4.3.4\].

4.4 Hypergraph isomorphism

It is possible to establish a link 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.

Now define the distance between two hypergraphs \( H = (V, E) \) and \( H' = (V, E') \) on the same set of vertices by \( \text{dist}(H, H') = |E \oplus E'|/2^n \), 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 \( f : \{0,1\}^n \rightarrow \{0,1\} \) 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

\[
f \cong g \iff f^{-1}(1) \cong g^{-1}(1) \text{ as hypergraphs},
\]

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 \textit{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 [BC08b] studied this question and obtained worst-case query-complexity bounds for the case of uniform hypergraphs. Yet a characterization of the testability of isomorphism between uniform hypergraphs remains 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 \textit{homomorphism} between \( H = (V, E) \) and \( \tilde{H} = (\tilde{V}, \tilde{E}) \) is a mapping \( \Pi: V \to \tilde{V} \) such that for all \( \{v_1, \ldots, v_r\} \in V \), the implication \( \{v_1, \ldots, v_r\} \in E \implies \{\Pi(v_1), \ldots, \Pi(v_r)\} \in \tilde{E} \) holds. The homomorphism \( \Pi \) is called \textit{full} (and \( H \) is said to be \textit{fully homomorphic} to \( \tilde{H} \)) if it holds in both directions, i.e., if

\[
\{v_1, \ldots, v_r\} \in E \iff \{\Pi(v_1), \ldots, \Pi(v_r)\} \in \tilde{E}.
\]

Note that the size of \( \tilde{V} \) may be much smaller than the size of \( V \).

\textbf{4.4.1. Definition.} An \( r \)-uniform hypergraph \( H \) is \textit{k-crunchable} if it is fully homomorphic to an \( r \)-uniform hypergraph with \( \leq k \) vertices.

The \textit{crunching number} of \( H \) is the smallest \( k \) such that \( H \) is \( k \)-crunchable.

The \( \varepsilon \)-approximate crunching number of \( H \), denoted \( \text{CrunchNum}_\varepsilon(H) \), is the smallest \( k \) such that \( H \) is \( \varepsilon \)-close to a \( k \)-crunchable \( r \)-uniform hypergraph.

The \( \varepsilon \)-testing number of \( H \), denoted \( \text{TestNum}_\varepsilon(H) \), is the minimum \( q \) for which there exists an \( \varepsilon \)-tester 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 (see Lemma 4.4.6).

We prove the following.

\textbf{4.4.2. Theorem (Chakraborty et al. [CFGM12]).} For every \( r \in \mathbb{N}, \varepsilon > 0 \) there exists a pair of functions \( L_{\varepsilon,r}(t) \) and \( U_{\varepsilon,r}(t) \), with \( \lim_{t \to \infty} L_{\varepsilon}(t) = \infty \), such that for every \( r \)-uniform hypergraph \( H \) we have

\[
L_{\varepsilon,r} (\text{CrunchNum}_\varepsilon(H)) \leq \text{TestNum}_\varepsilon(H) \leq U_{\varepsilon,r}(\text{CrunchNum}_{\varepsilon/3}(H)).
\]
The original proof of Fischer for (a statement equivalent to) the special case of Theorem 4.4.2 when \( r = 2 \) applied the highly acclaimed Szeméredi regularity lemma [Sze76] 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.

Now we prove Theorem 4.4.2. The functions \( L_\varepsilon \) and \( U_\varepsilon \) can be extracted from the proofs of the lower bound and the upper bound, respectively.

### 4.4.1 Lower bound via crunching

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

\[
E' = \left\{ \{v_1, \ldots, v_k\} \mid \{\Pi(v_1), \ldots, \Pi(v_k)\} \in E \right\}.
\]

A \( k \)-crunch of a hypergraph is a \( \Pi \)-crunch for some \( \Pi \) with an image of size \( \leq k \).

Note that every \( k \)-crunch is a \( k \)-crunchable hypergraph (as witnessed by the same mapping \( \Pi \)). When \( \Pi \) is injective, a \( \Pi \)-crunch 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 crunchs from injective ones (permutations). A **random \( k \)-crunch 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 \( \Pi \)-crunch of \( H \).

Now define the distribution \( D^k_{H} \) by drawing a random permutation of a random \( k \)-crunch of \( H \). Also write \( D_{H} \) for the uniform distribution over all permutations of \( H \).

**Lemma.** Let \( H \) be an \( r \)-uniform hypergraph and define \( D_{H} \) and \( D^k_{H} \) as before. Then it is impossible to distinguish a random \( \tilde{H} \sim D_{H} \) from a random \( \tilde{H} \sim D^k_{H} \) 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 \( \Pi \) is injective on \( Q \), the distribution of replies to queries \( e_1, \ldots, e_q \) is identical for \( D_{H} \) and \( D^k_{H} \). But \( E_Q(\Pi) \) occurs except with probability at most \( |Q|^2/k = o(1) \) as the choice of \( \Pi \) 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 $\hat{H}$ is drawn from $\mathcal{D}_H$ as when it is drawn from $\mathcal{D}_H^k$. We conclude by Lemma 1.5.6 that the tester cannot distinguish $\mathcal{D}_H$ from $\mathcal{D}_H^k$ with $q$ queries and success probability $\geq 2/3$.

4.4.5. COROLLARY. If an $r$-uniform hypergraph is $\varepsilon$-far from being $k$-crunchable, then $\varepsilon$-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, we show that being $O(1)$-chunchable is equivalent to having “algebra number” $O(1)$ as well.

4.4.6. DEFINITION. 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 $\varepsilon$-approximate algebra number is the smallest $k$ such that $H$ is $\varepsilon$-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 $E = \{\{v, w\} | v \in A_i, w \in A_j, (i, j) \in S, v \neq w\}$ (note that $i = j$ is allowed but loops are not). The $\varepsilon$-approximate pairing number of $G$ is defined similarly.

4.4.7. LEMMA.

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

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

3. Any $k$-crunchable graph has pairing number $k$. Conversely, any graph with pairing number $k$ is $\varepsilon$-close to being $k^2/\varepsilon$-crunchable.

Proof.

1. Let $cl(A)$ denote the edge set of the clique with vertex set $A \subseteq V$. It is enough to show that for disjoint $A_1, A_2 \subseteq V$, the set of edges between $A_1$ and $A_2$ is in the algebra generated by $cl(A_1), cl(A_2)$ and $cl(A_1 \cup A_2)$. This is easy to see because the set of edges in question is equal to $cl(A_1) \cup cl(A_2) \cap cl(A_1 \cup A_2)$. 


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 $A_S = (\cap_{i \in S} C_i) \cap (\cap_{i \notin S} C_i)$. 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 4.4.6. 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 $t = \lceil 1/\varepsilon \rceil$ subsets $A_{i1}, \ldots, A_{ik}$ of roughly equal size and remove the edges with both endpoints inside the same $A_{ij}$. If $n$ is divisible by $t$, then from all $\binom{n}{2}$ possible edges, the removed ones constitute a fraction bounded by $t \cdot (1/t^2) = 1/t \leq \varepsilon$; a simple argument shows that the same bound still holds in the general case. Hence this graph is $\varepsilon$-close to the original graph, and is also $k$-crunchable by construction.

4.4.2 Upper bound via partition properties

For the upper bound we need to discuss “partition properties” of hypergraphs, which generalizes those discussed in the context of graphs by Goldreich, Goldwasser and Ron [GGR98]. A graph partition instance $\psi$ is composed of an integer $k$ specifying the number of sets in the required partition $V_1, \ldots, V_k$ of the graph’s vertex set, and intervals specifying the allowed ranges for the number of vertices in every $V_i$ and the number of edges between every $V_i$ and $V_j$ for $i \leq j$. Many problems, such as $k$-corollability and maximum clique, can be easily formulated in this framework. In [GGR98] the authors presented algorithms for testing if a graph satisfies a certain partition property. We use a similar notion for hypergraphs, taken from the work of Fischer, Matsliah and Shapira [FMS10]. They work with directed hypergraphs, but we state their results in terms of undirected hypergraphs.

**Hypergraph partition property**

Let $H = (V, E)$ be a directed $r$-uniform hypergraph. 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 $\Pi$. We denote by $\Phi$ the set of all possible mappings $\phi: [r] \to [k]$. We think of every $\phi \in \Phi$ as mapping the vertices of an $r$-tuple to the components of $\Pi$. We denote by $E_\phi^\Pi \subseteq E$ the following collection of $r$-tuples:

$$E_\phi^\Pi = \left\{ (v_1, \ldots, v_r) \in E \mid \forall j \in [r] : v_j \in V_{\phi(j)}^\Pi \right\}.$$
4.4.8. Definition. A density tensor of order $k$ and arity $r$ is a sequence $\psi = \langle \langle \rho_j \rangle_{j \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle$ of reals between 0 and 1. (The interpretation is that they specify the presumed normalized sizes of $|V_i^j|$ and $|E_\phi^j|$ of a $k$-partition of a hypergraph of arity $r$.) Whenever $k$ and $r$ are clear from context, we call $\psi$ simply a density tensor.

In particular, given a $k$-partition $\Pi = \{V_1^k, V_2^k, \ldots, V_k^k\}$ of a hypergraph $H$, we set $\psi^\Pi$ 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 $j$, $\rho_j^\Pi = \frac{1}{n} \cdot |V_j^k|$ and for all $\phi$, $\mu_\phi^\Pi = \frac{1}{n^r} \cdot |E_\phi^k|$.

4.4.9. Definition. For a fixed hypergraph $H$ of arity $r$, a set $\Psi$ of density tensors (of order $k$ and arity $r$) defines a property of the $k$-partitions of $V(H)$ as follows. We say that a partition $\Pi$ of $V(H)$ (exactly) satisfies $\Psi$ if there exists a density tensor $\psi = \langle \langle \rho_j \rangle_{j \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle \in \Psi$, such that $\psi$ and the density tensor $\psi^\Pi$ of $\Pi$ are equal. Namely, $\Pi$ satisfies $\Psi$ if there is $\psi = \langle \langle \rho_j \rangle_{j \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle \in \Psi$ 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: one with respect to the edge density parameters $\langle \mu_\phi \rangle$, and the other with respect to the usual closeness measures between hypergraphs.

4.4.10. Definition. A $k$-partition $\Pi$ $\varepsilon$-approximately satisfies $\Psi$ if there is $\psi = \langle \langle \rho_j \rangle_{j \in [k]}, \langle \mu_\phi \rangle_{\phi \in \Phi} \rangle \in \Psi$ such that

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

In this case $\psi^\Pi$ is $\varepsilon$-approximate to $\psi$.

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

We define one additional measure of closeness to the property $\Psi$. The distance of a hypergraph $H$ from the property $\Psi$ is defined as $\text{dist}(H, \Psi) = \min_{H': \text{dist}(H, H') \leq \varepsilon} \text{dist}(H', \Psi)$. For $\varepsilon > 0$ we say that $H$ is $\varepsilon$-far from satisfying the property $\Psi$ when $\text{dist}(H, \Psi) \geq \varepsilon$, and otherwise, $H$ is $\varepsilon$-close to $\Psi$. The testing algorithm follows immediately from the following theorem.

4.4.11. Theorem (Fischer, Matsliah & Shapira [FMS10]). For any two $k, r \in \mathbb{N}$, and any set $\Psi$ of density tensors of order $k$ and arity $r$, there exists a randomized algorithm $A_T$ taking as inputs two parameters $\varepsilon, \delta > 0$ and an oracle access to a hypergraph $H$ of arity $r$, such that
4.4. Hypergraph isomorphism

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

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

For us it is enough to consider set $Ψ$ with a single partition property $ψ$.

Hypergraphs with small approximate crunching number

Let $H = (V,E)$ be a directed $k$-crunchable $r$-uniform hypergraph. We can define crunchings of directed hypergraphs in a similar manner, with the corresponding mapping $Π: V \rightarrow [k]$ and an hypergraph $\tilde{H} = ([k], \tilde{E})$ defining the edge patterns of $H$, i.e.,

$$\tilde{E} = \{(v_1, \ldots, v_k) \mid (Π(v_1), \ldots, Π(v_k)) ∈ E\}.$$  

The algorithm above can be used as a testing algorithm in the traditional sense on account of the following observations.

4.4.12. Lemma. Let $ε_0 < ε/k^r$. Any directed hypergraph that $ε_0$-approximately satisfies a partition property $Ψ$ is also $ε$-close to satisfying it.

Proof. Let $Π$ be a partition witnessing the fact that the hypergraph $ε_0$-approximately satisfies $Ψ$. For every $φ ∈ Φ$, we can add or remove $ε_0n^r$ edges to/from $E_Π^H$ so that the resulting graph exactly satisfies $Ψ$. Since $|Φ| = k^r$, this entails changing less than an $ε$-fraction of all possible edges.

4.4.13. Lemma. Let $H_0, H_1$ denote directed hypergraphs on $n$ vertices, where $H_1$ is the closest $k$-crunchable hypergraph to $H_0$. Suppose $H_0$ is $ε/3$-close to $H_1$ and the crunch is defined via the map $Π: V(H_0) → V(H_1)$. We can assume $V(H_0) = V(H_1) = [n]$. Let $ψ = \langle \langle ρ_j \rangle_j \in [k], \langle µ_φ \rangle_φ ∈ Φ \rangle$ denote the following density tensor of order $k$ and arity $r$:

- for all $j ∈ [k]$, $ρ_j = \frac{Π^{-1}(j)}{n}$;
- for $φ ∈ Φ$, $µ_φ = \frac{E_Π^φ(H_0)}{n^r}$.

Let $ε_0 < 2ε/(9k^r)$. Then $H_0$ satisfies $\{ψ\}$, and every hypergraph that $ε_0$-approximately satisfies $\{ψ\}$ is $8ε/9$-close to being isomorphic to $H_0$. 
Proof. Observe that in the density tensor above, the partition sizes $\rho_j$ are defined by $H_1$, but the edge densities $\mu_\phi$ are those of $H_0$. Note that $H_0$ satisfies $\{\psi\}$ by definition. The $k$-crunchable hypergraph $H_1$ does not satisfy the property $\psi$, but it does satisfy a related partition property $\psi_1$ with the same partition sizes $\{\rho_j\}$ but where the edge densities $\{\mu_\phi\}$ are all zero or one.

Take any hypergraph $H_3$ that $\varepsilon_0$-approximately satisfies $\psi$. By Lemma 4.4.12, it is $2\varepsilon/9$-close to some hypergraph $H_2$ that satisfies $\psi$. We show that $H_2$ is $\varepsilon/3$-close to satisfying $\psi_1$. The reason is that for any $\phi \in \Phi$, the number of edges of type $\phi$ that we need to change is $\min(\mu_\phi, 1 - \mu_\phi)n^r$. So by modifying

$$\sum_{\pi \in \Phi} \min(\mu_\phi, 1 - \mu_\phi)n^r$$

edges we can obtain a hypergraph that satisfies $\psi_1$. But this expression is also the number of edges that we need to change from $H_0$ so that it satisfies $\psi_1$, which is the distance between $H_0$ and $H_1$, hence at most $\varepsilon/3$.

Moreover, the only $r$-uniform directed hypergraph that satisfies $\psi_1$ is $H_1$, up to isomorphism. Therefore $H_2$ is $\varepsilon/3$-close to isomorphic to $H_1$, and $2\varepsilon/3$-close to $H_0$. Hence $H_3$ is $8/9$-close to begin isomorphic to $H_0$. \(\Box\)

For undirected hypergraphs, simply replace each edge $\{v_1, \ldots, v_r\}$ with all $r!$ directed edges of the form $(v_{\pi(1)}, \ldots, v_{\pi(r)})$ for a permutation $\pi: [r] \rightarrow [r]$, and test isomorphism to this directed version. The following follows.

4.4.14. Theorem. Let $\varepsilon \in (0, 1)$. Testing isomorphism to an $r$-uniform hypergraph that is $\varepsilon/3$-close to $k$-crunchable can be done with $\text{poly}(k^r/\varepsilon)$ queries.

4.4.3 Proof of the characterization

Proof of Theorem 4.4.2. By definition, any hypergraph $H$ is $\varepsilon$-far from $(\text{CrunchNum}_{\varepsilon}(H) - 1)$-crunchable, so by Lemma 4.4.5, we have

$$\text{TestNum}_{\varepsilon}(H) \geq L_{\varepsilon,r}(\text{CrunchNum}_{\varepsilon}(H))$$

for some $L_{\varepsilon,r}(t) = \Omega\left(\frac{\sqrt{t^r}}{r}\right)$. Clearly $\lim_{n \rightarrow \infty} L_{\varepsilon}(t) = \infty$.

For the upper bound, any hypergraph $H$ is $\varepsilon/3$-close to $\text{CrunchNum}_{\varepsilon/3}(H)$-crunchable. Hence we have by Theorem 4.4.14 that

$$\text{TestNum}_{\varepsilon}(H) \leq U_{\varepsilon/3,r}(\text{CrunchNum}_{\varepsilon/3}(H)),$$

for some appropriate polynomial $U_{\varepsilon,r}(t) = \text{poly}(t^r/\varepsilon)$. \(\Box\)
4.5 Junto-symmetric functions vs. layered juntas

Now we address the question of 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 applies the crunching method to boolean functions. In this setting the procedure resembles an idea used by Blais and O’Donnell [BO10].

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

\[
    f(x) = \tilde{f}_{|x|}(x_{|J_x|}),
\]

Perhaps it should be stressed that layered \( k \)-juntas are not, in general, \( k \)-juntas. Let \( L\mathcal{J}_k \) denote the class of layered \( k \)-juntas, respectively. Note that \( JS_k \subset L\mathcal{J}_k \).

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 4.4, 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 \( \tilde{f}_i \).

4.5.2. Definition. A random \( t \)-crunch of the function \( f \) defined by \( f(x) = \tilde{f}_{|x|}(x_{|J_x|}) \) is a function \( g \in JS_t \) obtained as follows:

1. pick, uniformly at random, a subset \( J \subseteq [n] \) of size \( t \) and a mapping \( \gamma: [n] \rightarrow J \);
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 \( g \).

4.5.3. Theorem (Chakraborty et al. [CFGM12]). Fix \( \varepsilon > 0 \) and \( Q: \mathbb{N} \rightarrow \mathbb{N} \), and suppose \( f \in L\mathcal{J}_k \). Then \( \Omega(Q(k)) \) queries are needed to distinguish a random permutation of \( f \) from a random permutation of a random \((k \cdot Q(k))^2\)-crunch of \( f \).

In particular, if \( f \) is \( \varepsilon \)-far from \( JS_{(k \cdot Q(k))^2} \), then \( \varepsilon \)-testing isomorphism to \( f \) requires \( \Omega(Q(k)) \) queries.

Proof. Let \( D_{\text{yes}} \) denote the random permutations of \( f \) and \( D_{\text{no}} \) the distribution of random permutations of \( t \)-crunchs of \( f \). Given \( g \in D_{\text{no}} \) and its corresponding mapping \( \gamma: [n] \rightarrow J \), call a set of layers \( \{\ell_1, \ldots, \ell_m\} \) collision-free if \( \gamma \) is injective on
Chapter 4. Junto-symmetric functions and hypergraph isomorphism

\[ \bigcup_{i \in [m]} J_{1i} \]. Let there 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 containing inputs queried is collision-free with respect to the randomly chosen mapping \( \gamma \) 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 1.5.6 (and the ensuing remark), it is enough to show that \( E_{x^1, \ldots, x^q} \) occurs with probability \( > \frac{2}{3} \).

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 sort of argument admits certain generalizations. For example, we can consider functions that have few additional variables outside a known set \( A \) (as in Section 4.3.3), as the address function on \( n \) variables (which behaves as a 1-junta for any fixed setting of the \( \approx \log n \) addressing variables). Choosing \( q \) small enough for address function to be far from isomorphic to a \( q^2 \)-crunching of the addressee variables gives a lower bound of \( \Omega(q) \) for testing isomorphism to it.

### 4.6 Linear isomorphism

We turn our attention now to a more general notions of isomorphism, namely 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). We show that functions that are far from having constant Fourier dimension are hard to test for isomorphism.

**4.6.1. Definition.** Two boolean functions \( f, g: \{0, 1\}^n \to \{0, 1\} \) 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.

**4.6.2. Definition.** 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

\[
A \triangleq \{ S \in \{0, 1\}^n \mid \hat{f}(S) \neq 0 \}.
\]

Then the dimension of the span of \( A \) is called the **Fourier dimension** of \( f \).
4.6. Linear isomorphism

4.6.3. Lemma. 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 \( \chi_v \) to be \( \pm 1 \)-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 easily seen to be boolean-valued on \( \{0, 1\}^n \) if \( f \) is, because all \( 2^{k'} \) assignments to \( \chi_v(x), i = 1 \ldots k' \) are possible. Hence there is a \( k' \)-junta \( g: \{0, 1\}^n \rightarrow \{0, 1\} \) and a change of basis \( A: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n \) 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 whose defining vector is in the linear span of the defining vectors of the parities in the product. 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 \( k \).

4.6.4. Theorem. If \( f: \{0, 1\}^n \rightarrow \{0, 1\} \) is \( \varepsilon \)-far from having Fourier dimension \( k \) then any adaptive \( \varepsilon \)-tester for linear isomorphism to \( f \) takes at least \( k - 1 \) queries.

Proof. Let \( f: \{0, 1\}^n \rightarrow \{0, 1\} \) be \( \varepsilon \)-far from having Fourier dimension \( k \). We will use Lemma 1.5.6 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:

\( D_Y \): To generate a random function \( g_Y \) in \( D_Y \) pick a random linear transformation \( L: \{0, 1\}^n \rightarrow \{0, 1\}^n \) of full rank and let \( g_Y(x) \overset{\Delta}{=} f(Lx) \).

\( D_N \): To generate a random function \( g_N \) in \( D_N \) pick a random linear transformation \( R: \{0, 1\}^n \rightarrow \{0, 1\}^n \) of rank exactly \( k \) and let \( g_N(x) \overset{\Delta}{=} f(Rx) \).

Note that \( D_Y \) is a distribution supported on the set of those functions which are linearly isomorphic to \( f \). On the other hand, \( D_N \) is supported on those functions that are \( \varepsilon \)-far from linearly isomorphic to \( f \). This is because if \( g_N \in D_N \), 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 \( v_1, \ldots, v_t \),

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

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.

4.6.5. Lemma. 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 us 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 exists 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 \( E \) 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 \( v_1, \ldots, v_t \),

\[
\Pr_R[\forall i \ R(q_i) = v_i \mid E] = 1/M = \Pr_L[\forall i : L(q_i) = v_i].
\]

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

\[
\Pr_{g_N \leftarrow D_{no}} \left[ g_N \upharpoonright_Q = a \right] \geq (1 - 2^{t-k}) \Pr_{g_N \leftarrow D_{no}} \left[ g_N \upharpoonright_Q = a \mid E \right]
\]

\[
= (1 - 2^{t-k}) \Pr_{g_Y \leftarrow D_{yes}} \left[ g_Y \upharpoonright_Q = a \right].
\]

Therefore, if \( t \leq k - 2 \) we have

\[
\Pr_{g_N \leftarrow D_{no}} \left[ g_N \upharpoonright_Q = a \right] \geq (3/4) \Pr_{g_Y \leftarrow D_{yes}} \left[ g_Y \upharpoonright_Q = a \right].
\]
By Lemma 1.5.6, if $f$ is $\varepsilon$-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.

4.6.6. Remark. Gopalan et al \cite{GOS+09} proved that if $f : \{0, 1\}^n \rightarrow \{0, 1\}$ has Fourier dimension $k$ then testing linear isomorphism to $f$ can be done using $O(k2^k)$ queries.

4.7 Summary

We proved that isomorphism to any boolean function on the $n$-dimensional hypercube with a polynomial number of distinct permutations can be tested with a number of queries that is independent of $n$. To do this we introduced the notion of junto-symmetric functions, proved its equivalence with poly-symmetric functions, and reduced the problem to certain generalizations of junta testing. We also showed some partial results in the converse direction. A complete characterization, however, remains open.

We also considered isomorphism testers against a uniform hypergraph that is given in advance. Our results regarding the latter topic generalize the known classification of the testability of graph isomorphism, and in the process we also provided a simpler proof of his original result which avoids the use of Szemerédi’s regularity lemma.

Some related problems were also discussed, such as testing isomorphism up to linear transformations.