Tolerant Junta Testing and the Connection to Submodular Optimization and Function Isomorphism

Eric Blais
University of Waterloo

Clément L. Canonne
Stanford University

Talya Eden
Tel Aviv University

Amit Levi
University of Waterloo

Dana Ron
Tel Aviv University

November 2018

Abstract

A function $f\colon \{-1,1\}^n \to \{-1,1\}$ is a k-junta if it depends on at most k of its variables. We consider the problem of *tolerant* testing of k-juntas, where the testing algorithm must accept any function that is ϵ-*close* to some k-junta and reject any function that is ϵ'-far from every k'-junta for some $\epsilon' = O(\epsilon)$ and $k' = O(k)$.

Our first result is an algorithm that solves this problem with query complexity polynomial in k and $1/\epsilon$. This result is obtained via a new polynomial-time approximation algorithm for *submodular function minimization* (SFM) under large cardinality constraints, which holds even when only given an approximate oracle access to the function.

Our second result considers the case where $k' = k$. We show how to obtain a smooth tradeoff between the amount of tolerance and the query complexity in this setting. Specifically, we design an algorithm that given $\rho \in (0,1)$ accepts any function that is $\frac{\epsilon\rho}{16}$-close to some k-junta and rejects any function that is ϵ-far from every k-junta. The query complexity of the algorithm is $O\big(\frac{k\log k}{\epsilon\rho(1-\rho)^k}\big)$.

Finally, we show how to apply the second result to the problem of tolerant isomorphism testing between two unknown Boolean functions f and g. We give an algorithm for this problem whose query complexity only depends on the (unknown) smallest k such that either f or g is *close* to being a k-junta.

1 Introduction

A function $f\colon \{-1,1\}^n \to \{-1,1\}$ is a *k-junta* if it depends on at most k of its variables. Juntas are a central object of study in the analysis of Boolean functions, in particular since they are good approximators for many classes of (more complex) Boolean functions. In the context of learning, the study of juntas was introduced by Blum et al. [11, 12] to model the problem of learning in the presence of irrelevant attributes. Since then, juntas have been extensively studied both in computational learning theory (e.g., [32, 39]) and in applied machine learning (e.g., [25]).

Juntas have also been studied within the framework of property testing. Here the task is to design a randomized algorithm that, given query access to a function f, accepts if f is a k-junta and rejects if f is ϵ-far from every k-junta (i.e., f must be modified in more than an ϵ-fraction of its values in order to be made a k-junta). The algorithm should succeed with high constant probability, and should perform as few queries as possible. The problem of testing k-juntas was first addressed by Fischer et. [21]. They designed an algorithm that queries the function on a number of inputs polynomial in k, and *independent of* n. A series of subsequent works essentially settled the optimal query complexity for this problem, establishing that[1] $\tilde{\Theta}(k/\epsilon)$ queries are both necessary and sufficient [8, 9, 17, 36].

The standard setting of property testing, however, is somewhat brittle, in that a testing algorithm is only guaranteed to accept functions that *exactly* satisfy the property. But what if one wishes to accept functions

[1] We use the notation $\tilde{\Theta}$ to hide polylogarithmic dependencies on the argument, i.e. for expressions of the form $\Theta(f \log^c f)$ (for some absolute constant c).

that are *close* to the desired property? To address this question, Parnas, Ron, and Rubinfeld introduced in [34] a natural generalization of property testing, where the algorithm is required to be *tolerant*. Namely, a *tolerant property testing algorithm* is required to accept any function that is *close* to the property, and, as in the standard model, to reject any function that is far from the property.[2]

As observed in [34], any standard testing algorithm whose queries are uniformly (but not necessarily independently) distributed, is inherently tolerant to some extent. However, for many problems, strengthening the tolerance requires applying different methods and devising new algorithms (see e.g., [24, 34, 22, 1, 27, 31, 19, 14, 7]). Furthermore, there are some properties that have standard testers with sublinear query complexity, but for which any tolerant tester must perform a linear number of queries [20, 38].

The problem of tolerant testing of juntas was previously considered by Diakonikolas et al. [18]. They applied the aforementioned observation from [34] and showed that one of the junta testers from [21] actually accepts functions that are $\mathrm{poly}(\epsilon, 1/k)$-close to k-juntas. Chakraborty et al. [15] observed that the analysis of the (standard) junta tester of Blais [9] implicitly implies an $\exp(k/\epsilon)$-query complexity tolerant tester which accepts functions that are ϵ/C-close to some k-junta (for some constant $C > 1$) and rejects functions that are ϵ-far from every k-junta. Subsequent to the extended abstract version of this work, Levi and Waingarten [30] showed that for *non-adaptive* algorithms, tolerant testing is strictly harder than testing, with a polynomial-factor separation in the query complexity. However, the corresponding question for adaptive tolerant testing of juntas remains open.

1.1 Our results

In this work, we study the question of tolerant testing of juntas from two different angles, and obtain two algorithms with different (and incomparable) guarantees. Further, we show how to leverage one of these algorithms to get a tester for isomorphism between Boolean functions with "instance-adaptive" (defined below) query complexity. The first of our results is a $\mathrm{poly}(k, 1/\epsilon)$-query algorithm, which accepts functions that are close to k-juntas and rejects functions that are far from every $2k$-junta.

Theorem 1.1. *There exists an algorithm that, given query access to a function* $f\colon \{-1,1\}^n \to \{-1,1\}$ *and parameters* $k \geq 1$ *and* $\epsilon \in (0,1)$, *satisfies the following.*

- *If f is $\epsilon/10$-close to some k-junta, then the algorithm accepts with high constant probability.*
- *If f is ϵ-far from every $2k$-junta, then the algorithm rejects with high constant probability.*

The query complexity of the algorithm is $\mathrm{poly}(k, \frac{1}{\epsilon})$.

The algorithm referred to in the theorem can be seen as a relaxed version of a tolerant testing algorithm. Namely, the algorithm rejects functions that are ϵ-far from every $2k$-junta rather than ϵ-far from every k-junta. Similar relaxations have been considered both in the standard testing model (e.g., [33, 26, 28]) and in the tolerant testing model [34]. Although one may hope for a stronger statement (distinguishing functions close to k-junta from those far from k-junta), we note that for most practical purposes the relaxation above is more than sufficient. Among others, this is supported by the following two examples: (i) In the setting we are concerned about, k is to be thought of as very small (constant, or very slowly growing function of n), while n is dauntingly large. Thus, reducing the dimension of the problem from n features to $2k$ is already enough to break the curse of dimensionality (it is interesting to note that this factor-2 relaxation comes, in our case, from reducing an NP-Hard constrained submodular minimization problem to a general tractable submodular minimization one, but on a penalized function). (ii) For *learning* algorithms, and in particular in the testing-before-learning framework, it is common to try and identify the smallest class of functions the unknown target belongs to, in order to apply an efficient learning algorithm tailored to it. In order to do so, one often uses a doubling search (or variant thereof) over the set of parameters (here, k) to identify the "best candidate parameter" while not spending too many resources in this preliminary phase. This implies that, inherently, any such search approach already loses a constant factor in this relevant parameter.

[2]Ideally, a tolerant testing algorithm should work for any given tolerance parameter $\epsilon' < \epsilon$ (that is, accept functions that are ϵ'-close to having the property), and have complexity that depends on $\epsilon - \epsilon'$. However, in some cases the relation between ϵ' and ϵ may be more restricted (e.g., $\epsilon' = \epsilon/c$ for a constant c). A closely related notion considered in [34] is that of *distance approximation* where the goal is to obtain an estimate of the distance that the object has to a property.

However, there are situations where one may already be willing to afford an exponential query complexity (in k) and so may want to obtain the best guarantee achievable within these bounds; or cannot afford this gap between completeness and soundness (e.g., for the sake of a reduction, or because k happens to be moderately large). We thus also study the question of tolerant testing without the above relaxation. That is, when the tester is required to reject functions that are ϵ-far from being a k-junta. We obtain a *smooth tradeoff* between the amount of tolerance and the query complexity. In particular, this tradeoff allows one to recover, as special cases, both the results of Fischer et al. [21] and (an improvement of) Chakraborty et al. [15].

Theorem 1.2. *There exists an algorithm that, given query access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$ and parameters $k \geq 1$, $\epsilon \in (0,1)$ and $\rho \in (0,1)$, satisfies the following.*

- *If f is $\rho\epsilon/16$-close to some k-junta, then the algorithm accepts with high constant probability.*
- *If f is ϵ-far from every k-junta, then the algorithm rejects with high constant probability.*

The query complexity of the algorithm is $O\Big(\frac{k\log k}{\epsilon\rho(1-\rho)^k}\Big)$.

We now discuss some of the implications of this result. Setting $\rho = \Omega(1)$, we obtain a tolerant tester that distinguishes between functions $O(\epsilon)$-close to $\mathcal{J}_k$ and functions ϵ-far from $\mathcal{J}_k$, with query complexity $2^{O(k)}/\epsilon$ – thus matching (and even improving) the simple tester described in Section 3. At the other end of the spectrum, setting $\rho = O(1/k)$ yields a weakly tolerant tester that distinguishes $O(\epsilon/k)$-close to $\mathcal{J}_k$ from ϵ-far from $\mathcal{J}_k$, but with query complexity $\tilde{O}\big(k^2/\epsilon\big)$ – qualitatively matching the guarantees provided by the junta tester of [21]. Moreover, in view of the connection between tolerant testing and distance approximation (see [34]), by adapting the proof of [34, Claim 2] we see that Theorem 1.2 readily implies the following:

Corollary 1.3. *Fix any constant $B > 16$. There exists an algorithm that, given query access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$ and parameters $k \geq 1$, $\epsilon \in (0,1)$, outputs a value $\hat{\epsilon}$ which satisfies the following. With probability at least $2/3$, we have $\operatorname{dist}(f,\mathcal{J}_k) \leq \hat{\epsilon} \leq B\cdot\operatorname{dist}(f,\mathcal{J}_k) + \epsilon$; moreover, the query complexity of the algorithm is $2^{O(k)}/\epsilon \cdot O(\log(1/\epsilon)\log\log(1/\epsilon))$.*

Finally, we show how the above result can be applied to the problem of *isomorphism testing*, which we recall next. Given query access to two unknown Boolean functions $f,g\colon \{-1,1\}^n \to \{-1,1\}$ and a parameter $\epsilon \in (0,1)$, one has to distinguish between (i) f is equal to g up to some relabeling of the input variables; and (ii) $\operatorname{dist}(f, g\circ\pi) > \epsilon$ for every such relabeling π (where $g\circ\pi$ denote the natural function composition). The worst-case complexity of this task is known, with $\tilde{\Theta}\big(2^{\frac{n}{2}}/\sqrt{\epsilon}\big)$ queries being necessary (up to the exact dependence on ϵ) and sufficient [2, 3].

However, is the exponential dependence on n always necessary, or can we obtain better results for "simple" functions? Ideally, we would like our testers to improve on this worst-case behavior, and instead have an *instance-adaptive* query complexity, depending only on some intrinsic parameter of the functions f,g to be tested. This is the direction we pursue here. Let $k^* = k^*(f,g,\gamma)$ be the smallest k such that either f or g is γ-close to being a k-junta. We show that it is possible to achieve a query complexity only depending on this (unknown) parameter, namely of the form $\tilde{O}\big(2^{k^*(f,g,O(\epsilon))/2}/\epsilon\big)$.[3] Moreover, our algorithm offers a much stronger guarantee: it allows *tolerant* isomorphism testing.

Theorem 1.4 (Tolerant isomorphism testing)**.** *There exists an algorithm that, given query access to two functions $f,g\colon \{-1,1\}^n \to \{-1,1\}$ and parameter $\epsilon \in (0,1)$, satisfies the following, for some absolute constant $C \geq 1$.*

- *If f and g are $\frac{\epsilon}{C}$-close to isomorphic, then the algorithm accepts with high constant probability.*
- *If f and g are ϵ-far from isomorphic, then the algorithm rejects with high constant probability.*

The query complexity of the algorithm is $\tilde{O}\big(2^{\frac{k^*}{2}}/\epsilon\big)$ with high-probability (and $\tilde{O}\big(2^{\frac{n}{2}}/\epsilon\ \big)$ in the worst case), where $k^* = k^*(f,g,\frac{\epsilon}{C})$.

[3]It is worth noting that this parameter can be much lower than the actual number of relevant variables for either functions; for instance, there exist functions depending on *all* n variables, yet that are $o(1)$-close to $O(1)$-juntas.

The above statement is rather technical, and requires careful parsing. In particular, the parameter k^* is crucially *not* provided as input to the algorithm: instead, it is discovered adaptively by invoking the tolerant tester of Theorem 1.2. This explains the high-probability bound on the query complexity: with some small probability, the algorithm may fail to retrieve the right value of k^* – in which case it may use instead a larger value, possibly up to n.

Remark 1.5 (On the running time of our algorithms.). We note that, as in previous work on testing juntas, the query complexity depends only on k and $1/\epsilon$ but the running time depends on n (since even querying a single point in $\{-1,1\}^n$ requires specifying n bits).

1.2 Overview and techniques

The proofs of Theorems 1.1 and 1.2 both rely on the notion of the *influence* of a set of variables. Given a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ and a set $S \subseteq [n]$, the influence of the set S (denoted $\mathbf{Inf}_f(S)$) is the probability that $f(x) \neq f(y)$ when x and y are selected uniformly subject to the constraint that for any $i \in \bar{S}$, $x_i = y_i$. The relation between the number of relevant variables and the influence of a set was utilized in previous works. In what follows we let $\mathcal{J}_k$ denote the set of all k-juntas.

Our starting point is similar to the one in [21, 9]. We partition the n variables into $\ell = O(k^2)$ parts, which removes the dependence on n. It is not hard to verify that if f is close to $\mathcal{J}_k$, then there exist k parts for which the following holds. If we denote by $T \subseteq [n]$ the union of variables in these k parts, then the complement set $\bar{T}$ has small influence. On the other hand, Blais [9] showed that if a function is far from $\mathcal{J}_k$, then a random partition into a sufficiently large number of parts ensures the following with high constant probability. For every union T of k parts, the complement set $\bar{T}$ will have large influence. The above gives rise to a $(2^{(1+(o(1))k\log k}/\epsilon)$-query complexity algorithm that distinguishes functions that are $\frac{1}{3}\epsilon$-close to $\mathcal{J}_k$ from functions that are ϵ-far from $\mathcal{J}_k$. The algorithm considers all unions $T \subseteq [n]$ of k parts, estimates the influence of $\bar{T}$, and accepts if there exists a set with sufficiently small estimated influence. In order to obtain an algorithm with better query complexity, we consider two relaxations.

In both relaxations we consider a fixed partition $\mathcal{I} = \{I_1, \ldots, I_\ell\}$ of $[n]$ into $\ell = O(k^2)$ parts and for $S \subseteq [\ell]$ we let $\phi_{\mathcal{I}}(S) \stackrel{\text{def}}{=} \bigcup_{i \in S} I_i$.

Parameterized tolerant testing through submodular minimization. In order to describe the algorithm referred to in Theorem 1.1, it will be useful to introduce the following function. For a Boolean function f and a partition $\mathcal{I}= \{I_1, \ldots, I_\ell\}$, we let $h\colon 2^{[\ell]} \to [0,1]$ be defined as $h(J) \stackrel{\text{def}}{=} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J))$. The starting point of our approach is the observation that the exhaustive search algorithm described previously can be seen as performing a brute-force minimization of h, under a cardinality constraint. Indeed, it effectively goes over all sets $J \subseteq [\ell]$ of size $\ell - k$, estimates $h(J)$, and accepts if there exists a set J for which the estimated value is sufficiently small. With this view, it is natural to ask whether this minimization can be performed more efficiently, by exploiting the fact that by the diminishing marginal property of the influence, h is submodular. That is, for every two sets $J_1 \subseteq J_2$ and variable $i \notin J_2$, it holds that $h(J_1 \cup \{i\}) - h(J_1) \geq h(J_2 \cup \{i\}) - h(J_2)$. While it is possible to find the minimum value of a submodular function in polynomial time given query access to the function, if a cardinality constraint is introduced, then even finding an approximate minimum is hard [37]. In light of the hardness of the problem, we design an algorithm for the following *bi-criteria relaxation*. Given oracle access to a non-negative submodular function $h\colon 2^{[\ell]} \to \mathbb{R}$ and input parameters $\epsilon \in (0,1)$ and $k \in \mathbb{N}$, the algorithm distinguishes between the following two cases:

- There exists a set J such that $|J| \geq \ell - k$ and $h(J) \leq \epsilon$;
- For every set J such that $|J| \geq \ell - 2k$, $h(J) > 2\epsilon$.

Moreover, the algorithm can be adapted to the case where it is only granted access to an approximate oracle for h (for a precise statement, see Theorem 4.3). This is critical in our setting, since $h(J) = \mathbf{Inf}_f(\phi_{\mathcal{I}}(J))$, and we can only estimate the influence of sets of variables.

Subset influence and recycling queries. The key idea behind our second approach is the following. The exhaustive search algorithm estimates the influence of the set of variables $\phi_{\mathcal{I}}(J)$ for every set of indices

$J \subset [\ell]$ such that $|J| = \ell - k$ by performing pairs of queries specifically designed for J. Namely, it queries the value of the function on pairs of points in $\{-1,1\}^n$ that agree on the set $\bar{J}$. If it were possible to use the same queries for estimating the influence of $\phi_{\mathcal{I}}(J)$ for different choices of J, then we could reduce the query complexity. We show that this can be done if we consider the *ρ-biased subset influence* of a set $J \subset [\ell]$, defined next.

Given a partition $\mathcal{I} = \{I_1, \ldots, I_\ell\}$, a parameter $\rho \in (0,1)$, and a set $J \subset [\ell]$, a random ρ-biased subset $S \sim_\rho J$ is a subset of J resulting from taking every index in J to S with probability ρ. The expected influence of a random ρ-biased subset of J, referred to as the ρ-subset influence of J, is $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_{\mathcal{I}}(S))]$. We prove that for every set $J \subseteq [\ell]$, its ρ-subset influence is in $[\frac{\varrho}{3}\mathbf{Inf}_f(\phi_{\mathcal{I}}(J)), \mathbf{Inf}_f(\phi_{\mathcal{I}}(J))]$. A crucial element in our proof is a combinatorial result due to Baranyai [5] on factorization of regular hypergraphs. With this fact in hand, we then present an algorithm that allows to simultaneously estimate the ρ-subset influence of all sets $J \subset [\ell]$ of size $\ell - k$. The query complexity of the algorithm is $O\left(\frac{k \log k}{\epsilon\rho(1-\rho)^k}\right)$.

Application to isomorphism testing: tolerant testing and noisy samplers. The structure of our tolerant isomorphism testing algorithm is quite intuitive, and consists of two phases. In the first phase, we run a linear search on k, repeatedly invoking our tolerant junta tester to discover the smallest value k satisfying $\min(\mathrm{dist}(f, \mathcal{J}_k), \mathrm{dist}(g, \mathcal{J}_k)) \le \epsilon/C$. We note that a similar approach using a tester whose tolerance is only $\mathrm{poly}(\epsilon/k)$ might return a much larger value of k, since as k increases, the allowed tolerance decreases. In the second phase, we use this value of k to tolerantly test isomorphism between f and g. This phase, however, is not as straightforward as it seems: indeed, to achieve the desired query complexity, we would like to test isomorphism – for which we have known algorithms – between f_k and g_k, that is, the k-juntas closest to f and g respectively.

Yet here we face two issues: (i) we do not have query access to f_k and g_k; (ii) even in the completeness case f_k and g_k *need not actually be isomorphic.* Indeed, f and g are only promised to be close to k-juntas, and close to isomorphic. Hence, the corresponding juntas are only guaranteed to be *close* to isomorphic.

Addressing item (ii) relies on adapting the algorithm of [3], along with a careful and technical analysis of the distribution of the points it queries. (This analysis is also the key to providing the tolerance guarantees of our isomorphism tester.) We address item (i) as follows. Our algorithm builds on the ideas of Chakraborty et al. [16], namely on their notion of a "noisy sampler". A noisy sampler is given query access to a function that is promised to be close to some k-junta and provides (almost) uniformly distributed samples labeled (approximately) according to this k-junta. While the [16] noisy sampler works for functions that are $\mathrm{poly}(\epsilon/k)$-close to $\mathcal{J}_k$, we need a noisy sampler that works for functions that are only $\frac{\epsilon}{C}$-close to $\mathcal{J}_k$. To this end, we replace the weakly tolerant testing algorithm of [9] used in the noisy sampler of [16] with our tolerant testing algorithm. The query complexity of the resulting noisy sampler is indeed much higher than that of [16]. However, this does not increase the overall query complexity of our tolerant isomorphism testing algorithm, as stated in Theorem 1.4.

1.3 Future Directions and Open Questions

Our work leaves open several interesting questions, which we discuss below. The first relates to the factor-2 relaxation of Theorem 1.1: can one obtain a *bona fide* $\mathrm{poly}(k)$-query tolerant tester for k-junta, accepting functions $\epsilon/10$-close to some k-junta while rejection functions ϵ-far from every k-junta? On the other hand, Theorem 1.1 only provides a constant-factor distance approximation, even with this relaxation. What is the best possible distance approximation achievable in $\mathrm{poly}(k)$ queries? Switching gears, one may also ask whether tolerant testing (as considered in this work) is *any* harder than testing, which is known to be possible with a near-linear (in k) number of queries. We strongly suspect it to be the case: specifically, we conjecture that tolerant testing requires $\Omega(k^c)$ queries for some $c > 1$.[4] Finally, and more open-ended, is the question of whether the connection between tolerant junta testing (or, more generally, the influence of Boolean functions) and submodular minimization which underlies our main result can lead to further applications and insights.

[4] Interestingly, as mentioned earlier, subsequent work of Levi and Waingarten [30] partially confirmed this conjecture by establishing a polynomial query complexity separation between *non-adaptive* junta testing and tolerant testing.

1.4 Organization of the paper

After introducing the necessary notations and definitions in Section 2, we describe in Section 3 the common starting point of our algorithms – the reduction from n variables to $O(k^2)$ parts. Section 4 then contains the details of the submodular minimization under cardinality constraint underlying Theorem 1.1, which is then implemented in Section 5 with an approximate submodular minimization primitive. We then turn in Section 6 to the proof of Theorem 1.2, before describing in Section 7 how to leverage it to obtain our instance-adaptive tolerant isomorphism testing result.

2 Preliminaries

A *property* $\mathcal{P}$ of Boolean functions is a subset of all these functions, and we say that a function f *has the property* $\mathcal{P}$ if $f \in \mathcal{P}$. The distance between two functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$ is defined as their (normalized) Hamming distance $\operatorname{dist}(f,g) \stackrel{\text{def}}{=} \Pr_x[\, f(x) \neq g(x)\,]$, where x is drawn uniformly at random. Accordingly, for a function f and a property $\mathcal{P}$ we define the distance from f to $\mathcal{P}$ as $\operatorname{dist}(f,\mathcal{P}) \stackrel{\text{def}}{=} \min_{g\in\mathcal{P}} \operatorname{dist}(f,g)$. Given $\epsilon \geq 0$ and a property $\mathcal{P}$, we will say that a function f is *ϵ-far* from $\mathcal{P}$ (resp. *ϵ-close* to $\mathcal{P}$) if $\operatorname{dist}(f,\mathcal{P}) > \epsilon$ (resp. $\operatorname{dist}(f,\mathcal{P}) \leq \epsilon$).

We consider the following definition of tolerant testing of parameterized properties, restated below.

Definition 2.1 (Tolerant Testing of Parameterized Properties)**.** Let $\mathcal{P} = (\mathcal{P}_s)_{s\in\mathbb{N}}$ be a non-decreasing family of properties parameterized by $s \in \mathbb{N}$, i.e. such that $\mathcal{P}_s \subseteq \mathcal{P}_t$ whenever $s \leq t$; and $\sigma\colon \mathbb{N} \to \mathbb{N}$ be a non-decreasing mapping satisfying $\sigma(s) \geq s$ for all s. A *σ-tolerant testing algorithm for $\mathcal{P}$* is a probabilistic algorithm that gets three input parameters $s \in \mathbb{N}$ and $\epsilon_1, \epsilon_2 \in [0,1]$ such that $\epsilon_1 < \epsilon_2$, as well as oracle access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$. The algorithm should output a binary verdict that satisfies the following two conditions.

- If $\operatorname{dist}(f,\mathcal{P}_s) \leq \epsilon_1$ then the algorithm accepts f with probability at least $2/3$.
- If $\operatorname{dist}\big(f,\mathcal{P}_{\sigma(s)}\big) > \epsilon_2$, then the algorithm rejects f with probability at least $2/3$.

In some cases the algorithm is only given one parameter, ϵ_2, setting $\epsilon_1 = r(\epsilon_2)$ for some prespecified function $r\colon (0,1) \to (0,1)$.

The main focus of this work will be the property of being a *junta*, that is, a Boolean function that only depends on a (small) subset of its variables:

Definition 2.2 (Juntas)**.** A Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ is a *k-junta* if there exists a set $T \subseteq [n]$ of size at most k, such that $f(x) = f(y)$ for every two assignments $x, y \in \{-1,1\}^n$ that satisfy $x_i = y_i$ for every $i \in T$. We let $\mathcal{J}_k$ denote the set of all k-juntas (over n variables).

Notations. Hereafter, we denote by log the binary logarithm, by $[n]$ the set of integers $\{1,\dots,n\}$, and by $\mathcal{S}_n$ for the set of permutations of $[n]$. Given two disjoint sets $S, T \subseteq [n]$ and two partial assignments $x \in \{-1,1\}^S$ and $y \in \{-1,1\}^T$, we let $x \sqcup y \in \{-1,1\}^{S\cup T}$ be the partial assignment whose i-th coordinate is x_i if $i \in S$ and y_i if $i \in T$. Given a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ we write $\mathcal{O}_f$ for an oracle providing query access to f. For a set S, we denote by $\binom{S}{r}$ the set of all subsets of S of size r. Given a partition $\mathcal{I} = \{I_1,\dots,I_\ell\}$ of $[n]$ and a set $J \subseteq [\ell]$, we denote by $\phi_{\mathcal{I}}(J)$ the union $\bigcup_{i\in J} I_i$.

A central notion in this work is the *influence* of a set, which generalizes the standard notion of influence of a variable:[5]

Definition 2.3 (Set-influence)**.** For a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, the *set-influence* of a set $S \subseteq [n]$ is defined as

$$\mathbf{Inf}_f(S) = 2\Pr[\, f(x \sqcup u) \neq f(x \sqcup v)\,] \ ,$$

where $x \sim \{-1,1\}^{[n]\setminus S}$, and $u, v \sim \{-1,1\}^S$.

[5]Here, we define the influence of a set with an additional factor 2, so that the (set-)influence of a singleton $\{i\}$ coincides with the standard definition of the influence of the i-th variable (as the latter definition asks that the i-th bit be flipped instead of re-randomized).

3 From n variables to $O(k^2)$ parts

In this section we build on techniques from [21, 9] and describe how to reduce the problem of testing closeness to a k-junta to testing closeness to a k-*part* junta (defined below). The advantage of doing so is that while the former question concerns functions on n variables, the latter *no longer involves* n *as a parameter*: only k and ϵ now have a role to play. We start with a useful definition of k-*part juntas*, and two lemmas regarding their properties with respect to random partitions of the domain.

Definition 3.1 (Partition juntas [10, Definition 5.3], extended)**.** Let $\mathcal{I}$ be a partition of $[n]$ into ℓ parts, and $k \geq 1$. The function $f\colon \{-1,1\}^n \to \{-1,1\}$ is a *k-part junta with respect to* $\mathcal{I}$ if the relevant coordinates in f are all contained in at most k parts of $\mathcal{I}$. Moreover,

(i) f is said to *ϵ-approximate being a k-part junta with respect to* $\mathcal{I}$ if there exists a set $J \in \binom{[\ell]}{\ell-k}$ satisfying $\mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) \leq 2\epsilon$.

(ii) Conversely, f is said to *ϵ-violate being a k-part junta with respect to* $\mathcal{I}$ if for every set $J \in \binom{[\ell]}{\ell-k}$, $\mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) > 2\epsilon$.

Lemma 3.2 ([10, Lemma 5.4])**.** *For $f\colon \{-1,1\}^n \to \{-1,1\}$ and $k \geq 1$, let $\alpha \stackrel{\text{def}}{=} \operatorname{dist}(f, \mathcal{J}_k)$. Also, let $\mathcal{I}$ be a random partition of $[n]$ with $\ell \stackrel{\text{def}}{=} 24k^2$ parts obtained by uniformly and independently assigning each coordinate to a part. With probability at least $5/6$ over the choice of the partition $\mathcal{I}$, the function f $\frac{\alpha}{2}$-violates being a k-part junta with respect to $\mathcal{I}$.*

Lemma 3.3. *For $f\colon \{-1,1\}^n \to \{-1,1\}$ and $k \geq 1$, let $\alpha \stackrel{\text{def}}{=} \operatorname{dist}(f, \mathcal{J}_k)$ and let $\mathcal{I}$ be any partition of $[n]$ into $\ell \geq k$ parts. Then f 2α-approximates being a k-part junta with respect to $\mathcal{I}$.*

Proof: Let $g \in \mathcal{J}_k$ be such that $\operatorname{dist}(f,g) = \operatorname{dist}(f,\mathcal{J}_k) = \alpha$. Let $I_{i_1}, \dots, I_{i_r}$ be the $r \leq k$ parts of $\mathcal{I}$ containing the relevant variables of g. Then, for any set $J \subset [\ell]$ of size $\ell - k$ such that $\{i_1, \dots, i_r\} \subseteq \bar{J}$, we have that when drawing $x \sim \{-1,1\}^{\phi_{\mathcal{I}}(\bar{J})}$, and $u, v \sim \{-1,1\}^{\phi_{\mathcal{I}}(J)}$ the following holds.

$$\begin{aligned}\mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) &= 2\Pr[\, f(x \sqcup u) \neq f(x \sqcup v)\,] \leq 2\Pr[\, f(x \sqcup u) \neq g(x \sqcup u) \text{ or } f(x \sqcup v) \neq g(x \sqcup v)\,] \\ &\leq 2\left(\Pr[\, f(x \sqcup u) \neq g(x \sqcup u)\,] + \Pr[\, f(x \sqcup v) \neq g(x \sqcup v)\,]\right) \leq 2\,(\alpha + \alpha) = 4\alpha\ ,\end{aligned}$$

where the first inequality follows from observing that (as g does not depend on variables in $\phi_{\mathcal{I}}(J)$) one can only have $f(x \sqcup u) \neq f(x \sqcup v)$ if f disagrees with g on at least one of the two points; and the third inequality holds since both $x \sqcup u$ and $x \sqcup v$ are uniformly distributed. ■

The above two lemmas suggest the following approach for distinguishing between functions that are ϵ'-close to some k-junta and functions that are ϵ-far from every k'-junta. Suppose we select a random partition of $[n]$ into $O(k^2)$ parts. Then, with high probability over the choice of the partition, it is sufficient to distinguish between functions that $2\epsilon'$-approximate being a k-junta and functions that $\epsilon/2$-violate being a k'-part junta. Specifically, we get the proposition below, which we apply throughout this work:

Proposition 3.4 (Reduction to part juntas)**.** *Let $\mathcal{T}$ be an algorithm that is given query access to a function $f : \{-1,1\}^n \to \{-1,1\}$, a partition $\mathcal{I} = \{I_1, \dots, I_\ell\}$ of $[n]$ into ℓ parts, and parameters $k \in \mathbb{N}$ and $\epsilon \in (0,1)$. Suppose that $\mathcal{T}$ performs $q(k, \epsilon, \ell)$ queries to f and satisfies the following guarantees, for a pair of functions $r\colon (0,1) \times \mathbb{N} \to (0,1)$ and $r' : \mathbb{N} \to \mathbb{N}$.*

- *If f ϵ'-approximates being a k-part junta with respect to $\mathcal{I}$ and $\epsilon' \leq r(\epsilon, k)$, then $\mathcal{T}$ returns* accept *with probability at least $5/6$;*
- *If f ϵ-violates being a k'-part junta with respect to $\mathcal{I}$ and $k' \geq r'(k)$, then $\mathcal{T}$ returns* reject *with probability at least $5/6$.*

Then there exists an algorithm $\mathcal{T}'$, that given query access to f and parameters $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, satisfies the following.

- *If $\operatorname{dist}(f, \mathcal{J}_k) \leq \frac{\epsilon'}{2}$ and $\epsilon' \leq r(\epsilon, k)$, then $\mathcal{T}'$ outputs* accept *with probability at least $2/3$;*

- *If* $\mathrm{dist}(f, \mathcal{J}_{k'}) > 2\epsilon$ *and* $k' \geq r'(k)$*, then* $\mathcal{T}'$ *outputs* reject *with probability at least* $2/3$*.*

Moreover, the algorithm $\mathcal{T}'$ *has query complexity* $q(k, \epsilon, \ell)$*.*

Proof of Proposition 3.4: The algorithm $\mathcal{T}'$ first obtains a random partition $\mathcal{I}$ of $[n]$ into $\ell \stackrel{\mathrm{def}}{=} 24(k')^2$ parts by uniformly and independently assigning each coordinate to a part. $\mathcal{T}'$ then invokes $\mathcal{T}$ with parameters ϵ, k, ℓ and the partition $\mathcal{I}$. By Lemma 3.2 and the choice of ℓ, with probability at least 5/6 the partition $\mathcal{I}$ is *good* in the following sense. For $\alpha = \mathrm{dist}(f, \mathcal{J}_{k'})$, it holds that f $\frac{\alpha}{2}$-violates being a k'-part junta with respect to $\mathcal{I}$. Conditioned on $\mathcal{I}$ being good, and by Lemma 3.3, we are guaranteed that the following holds.

(i) If $\mathrm{dist}(f, \mathcal{J}_k) \leq \frac{\epsilon'}{2}$, then f ϵ'-approximates being a k-part junta with respect to $\mathcal{I}$;

(ii) If $\mathrm{dist}(f, \mathcal{J}_{k'}) > 2\epsilon$, then f ϵ-violates being a k'-part junta with respect to $\mathcal{I}$.

Therefore, $\mathcal{T}$ will answer as specified by the proposition with probability at least 5/6, making $q(\epsilon, k, \ell)$ queries. Overall, by a union bound, $\mathcal{T}'$ is successful with probability at least 2/3. ■

As an illustration of the above technique, and a warmup towards the (more involved) algorithms of the next sections, we show how to obtain an algorithm $\mathcal{T}'$ as specified in Proposition 3.4 with query complexity $2^{(1+o(1))k \log k}/\epsilon$. Given a partition $\mathcal{I}$ of $[n]$ into ℓ parts, $\mathcal{T}$ considers all $\binom{\ell}{\ell-k}$ sets of variables that result from taking the union of k parts. For each such set T, it obtains an estimate $\widetilde{\mathbf{Inf}}_f(T)$ of the influence of T, by performing $O\left(\frac{\ell \log \ell}{\epsilon}\right)$ queries to f. $\mathcal{T}$ accepts if for at least one of the sets T, $\widetilde{\mathbf{Inf}}_f(T)$ is at most $\frac{3}{2}\epsilon$. Performing $O(\frac{\ell \log \ell}{\epsilon})$ queries to the oracle for each set, ensures that the following holds with high constant probability. For every set T such that $\mathbf{Inf}_f(T) \leq \frac{4}{3}\epsilon$, $\widetilde{\mathbf{Inf}}_f(T) \leq \frac{3}{2}\epsilon$ and for every set T such that $\mathbf{Inf}_f(T) > 2\epsilon$, $\widetilde{\mathbf{Inf}}_f(T) > \frac{3}{2}\epsilon$. Hence, the algorithm $\mathcal{T}$ fulfills the requirements stated in Proposition 3.4 (for $r(\epsilon, k) = \frac{2}{3}\epsilon$ and $r'(k) = k$), and it follows that:

- If f is $\frac{1}{3}\epsilon$-close to some k-junta then $\mathcal{T}'$ accepts with probability at least 2/3.
- If f is ϵ-far from every k-junta then $\mathcal{T}'$ rejects with probability at least 2/3.

Since $\ell = 24k^2$, the query complexity of the algorithm is $\binom{\ell}{\ell-k} \cdot O(\frac{\ell \log \ell}{\epsilon}) = 2^{(1+o(1))k \log k}/\epsilon$.

4 Approximate submodular minimization under a cardinality constraint

In this section we show how a certain bi-criteria approximate version of submodular minimization with a cardinality constraint can be reduced to approximate submodular minimization with no cardinality constraint. This reduction holds even when given *approximate oracle access* to the submodular function, and is meaningful when the cardinality constraint is sufficiently large. Precise details follow.

Definition 4.1 (Approximate oracle)**.** *Let* $h\colon 2^{[\ell]} \to \mathbb{R}$ *be a function. An* approximate oracle *for* h*, denoted* $\mathcal{O}_h^{\pm}$*, is a randomized algorithm that, for any input* $J \subseteq [\ell]$ *and parameters* $\tau, \delta \in (0,1)$*, returns a value* $\tilde{h}(J)$ *such that* $|\tilde{h}(J) - h(J)| \leq \tau$ *with probability at least* $1 - \delta$*.*

Definition 4.2 (Approximate submodular minimization algorithm)**.** *Let* $h\colon 2^{[\ell]} \to \mathbb{R}$ *be a non-negative submodular function and let* $\mathcal{O}_h^{\pm}$ *denote an approximate oracle for* h*. An* approximate submodular function minimization *algorithm (ASFM) is an algorithm that, when given access to* $\mathcal{O}_h^{\pm}$ *and called with input parameters* ξ *and* δ*, returns a value* ν *such that* $|\nu - \min_{J \subseteq [\ell]}\{h(J)\}| \leq \xi$ *with probability at least* $1 - \delta$*.*

In Corollary 5.6 in Section 5 we establish the existence of such an ASFM algorithm. The running time of the algorithm is polynomial in ℓ, logarithmic in the maximal value of the function and linear in the running time of the approximate oracle. We next present an algorithm for approximate submodular minimization under a cardinality constraint.

Algorithm 1 Approximate Submodular Minimization under a Cardinality Constraint($\mathcal{O}_h^{\pm}, \epsilon, \delta, \xi, k$)

1: Let $h'(J) = h(J) - \frac{\epsilon}{k}|J|$ so that for every τ', δ' $\mathcal{O}_{h'}^{\pm}(J, \tau', \delta') = \mathcal{O}_h^{\pm}(J, \tau', \delta') - \frac{\epsilon}{k}|J|$.

2: Let ν be the returned value from invoking an ASFM algorithm with access to $\mathcal{O}_{h'}^{\pm}$ and parameters ξ and δ.

3: Accept if and only if $\nu \le (2 - \frac{\ell}{k}) \cdot \epsilon + \xi$.

Theorem 4.3. *For a non-negative submodular function h, 1 satisfies the following conditions:*

1. *If there exists a set $J \subseteq [\ell]$ such that $|J| \ge \ell - k$ and $h(J) \le \epsilon$, then the algorithm accepts with probability at least $1 - \delta$.*
2. *If for every set $J \subseteq [\ell]$ such that $|J| \ge \ell - 2(1 + \frac{\xi}{\epsilon})k$, we have $h(J) > 2\epsilon + 2\xi$, then the algorithm rejects with probability at least $1 - \delta$.*

Moreover, the second item can be strengthened so that it holds for functions h that satisfy the following: (i) for every set $J \subseteq [\ell]$ such that $|J| \ge \ell - k$, $h(J) > 2\epsilon + 2\xi$ and (ii) for every set $J \subseteq [\ell]$ such that $|J| \ge \ell - 2(1 + \frac{\xi}{\epsilon})k$, $h(J) > \epsilon + 2\xi$.

Proof: First of all, note that the function h' defined in Step 1 is indeed submodular, as the sum of the submodular function h and a modular function (scalar multiple of the cardinality function). By Definition 4.2, with probability at least $1 - \delta$ the value ν defined in Step 2 of the algorithm thus satisfies

$$|\nu - \min_{J \subseteq [\ell]}\{h'(J)\}| \le \xi \; . \tag{1}$$

We start with proving the first item in Theorem 4.3. If there exists a set $J^* \subseteq [\ell]$ such that $|J^*| \ge \ell - k$ and $h(J^*) \le \epsilon$, then

$$\min_{J \subseteq [\ell]}\{h'(J)\} \le h'(J^*) \le \epsilon - \frac{\epsilon}{k}(\ell - k) = \left(2 - \frac{\ell}{k}\right) \cdot \epsilon,$$

and therefore, by Equation (1), with probability at least $1 - \delta$,

$$\nu \le \min_{J \subseteq [\ell]}\{h'(J)\} + \xi \le \left(2 - \frac{\ell}{k}\right) \cdot \epsilon + \xi,$$

and the algorithm accepts.

We divide the analysis of the second item in the theorem into two cases depending on $|J|$.

- For sets J such that $|J| \ge \ell - 2(1 + \frac{\xi}{\epsilon})k$,

  $$h'(J) > 2\epsilon + 2\xi - \frac{\epsilon}{k} \cdot \ell = \left(2 - \frac{\ell}{k}\right)\epsilon + 2\xi,$$

  where in the first inequality we used the fact that for all sets J, $|J| \le \ell$.

- For sets J such that $|J| < \ell - 2(1 + \frac{\xi}{\epsilon})k$, it holds that

  $$h'(J) > 0 - \frac{\epsilon}{k} \cdot \left(\ell - 2\left(1 + \frac{\xi}{\epsilon}\right)k\right) = \left(2 - \frac{\ell}{k}\right) \cdot \epsilon + 2\xi,$$

  where in the first inequality we used the fact that for all sets J, $h(J) \ge 0$.

Hence,

$$\min_{J \subseteq [\ell]}\{h'(J)\} > \left(2 - \frac{\ell}{k}\right) \cdot \epsilon + 2\xi,$$

and by Equation (1), with probability at least $1 - \delta$, it holds that

$$\nu \ge \min_{J \subseteq [\ell]}\{h'(J)\} - \xi > \left(2 - \frac{\ell}{k}\right) \cdot \epsilon + \xi,$$

and the algorithm rejects. ■

5 Approximate submodular function minimization

In this section we use results from [29] to obtain an approximate submodular minimization algorithm, as defined in Definition 4.2.[6] This is done in three steps: (1) We use the known fact that the problem of finding the minimum of a submodular function g can be reduced to finding the minimum of the Lovász extension for that function, denoted $\mathcal{L}_g$. (2) We then extend the results of [29] (and specifically of Theorem 61) and provide a noisy separation oracle for $\mathcal{L}_g$ when only given approximate oracle access to the function g. (3) Finally, we apply Theorem 42 from [29], which provides an algorithm that, when given access to a separation oracle for a function, returns an approximation to that function's minimum value. Note that in this section, we analyze the *time* complexity of our algorithms for submodular function minimization; we will later, in the following sections, switch back to query complexity when applying them to our testing problem, using running time as an upper bound for query complexity.

We start with the following definition of the Lovász extension of a submodular function.

Definition 5.1 (Lovász Extension)**.** *Given a submodular function $g\colon 2^{[\ell]} \to \mathbb{R}$, the* Lovász extension *of g, is a function $\mathcal{L}_g\colon [0,1]^\ell \to \mathbb{R}$, which is defined for all $x \in [0,1]^\ell$ by*

$$\mathcal{L}_g(x) \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{t\sim[0,1]}\left[g(\{\ i:\ x_i \geq t\ \})\right]\ ,$$

where $t \sim [0,1]$ denotes that t is drawn uniformly at random from $[0,1]$.

The following theorem is standard in combinatorial optimization (see e.g. [4] and [23, 35]) and provides useful properties of the Lovász extension.

Theorem 5.2. *The Lovász extension $\mathcal{L}_g$ of a submodular function $g\colon 2^{[\ell]} \to \mathbb{R}$ satisfies the following properties.*

1. *$\mathcal{L}_g$ is convex and $\min_{x\in[0,1]^\ell}\{\mathcal{L}_g(x)\} = \min_{S\subseteq[\ell]}\{g(S)\}$.*
2. *If $x_1 \geq \ldots \geq x_\ell$, then*
$$\mathcal{L}_g(x) = \sum_{i=1}^{\ell}\big(g([i]) - g([i-1])\big)x_i\ .$$

By the first item of Theorem 5.2, in order to approximate the minimum value of a submodular function g, it suffices to approximate the minimum of its Lovász extension. As discussed at the start of the section, this is done by providing a *separation oracle* for $\mathcal{L}_g$.

Definition 5.3 ((Noisy) Separation Oracle [29, Definition 2])**.** *Let h be a convex function over $\mathbb{R}^\ell$ and let Ω be a convex set in $\mathbb{R}^\ell$. A* separation oracle *for h with respect to Ω is an algorithm that for an input $x \in \Omega$ and parameters $\eta, \gamma \geq 0$ satisfies the follows. It either asserts that $h(x) \leq \min_{y\in\Omega}\{h(y)\} + \eta$ or it outputs a halfspace $H \stackrel{\text{def}}{=} \{z : a^T z \leq a^T x + c\}$ such that*

$$\{\ y \in \Omega :\ h(y) \leq h(x)\ \} \subset H\ ,$$

where $a \in [0,1]^\ell$, $a \neq 0$, and $c \leq \gamma\|a\|_2$.

In Theorem 61 in [29] it is shown how to define a separation oracle for a function g when given *exact* query access to g; we adapt the proof to the case where one is only granted access to an approximate oracle for g, and the resulting procedure has small failure probability.

[6] We remark that more efficient algorithms, such as the convex optimization algorithm (given noisy access) of [6] applied to the Lovász extension, might also apply. Such algorithms would yield a better polynomial for the query complexity of our problem; however, they typically require stronger requirements on the noise (e.g., subgaussian distribution for [6]), where our implementation is robust to adversarial noise. This explicit robustness to noise is the reason why we rely on the results of [29], which – to the best of our knowledge – is the first algorithm to provide the robustness guarantees we need.

Algorithm 2 Separation Oracle ($\mathcal{O}_g^{\pm}, \bar{x}, \eta, \gamma, \delta$)

1: Assume without loss of generality that $\bar{x}_1 \geq \bar{x}_2 \geq \ldots \geq \bar{x}_\ell$ (otherwise re-index the coordinates).
2: Let $\tau = \min\{\eta/4\ell, \gamma/2\ell\}$.
3: For each $i \in [\ell]$, let $\tilde{g}([i])$ be the returned value from invoking $\mathcal{O}_g^{\pm}$ on the set $[i]$ with parameters $\frac{\tau^2}{2}$ and $\frac{\delta}{\ell}$.
4: Define $\tilde{a} \in \mathbb{R}^\ell$ by $\tilde{a}_i \stackrel{\text{def}}{=} \tilde{g}([i]) - \tilde{g}([i-1])$ for each $i \in [\ell]$.
5: Let $\tilde{\mathcal{L}}_g(\bar{x}) \stackrel{\text{def}}{=} \tilde{a}^T \bar{x}$.
6: **if** for every $i \in [\ell]$, $|\tilde{a}_i| < \tau$ **then**
7:     **return** $\bar{x}$ (which satisfies "$\mathcal{L}_g(\bar{x}) \leq \min_{y \in [0,1]^\ell}\{\mathcal{L}_g(y)\} + \eta$").
8: **else**
9:     **return** the halfspace $H = \{z : \ \tilde{a}^T z \leq \tilde{\mathcal{L}}_g(\bar{x}) + 2\tau\ell\|\tilde{a}\|_2\}$.
10: **end if**

Lemma 5.1. *Let $g \colon 2^{[\ell]} \to \mathbb{R}$ be a convex function, and let $\Phi_g(\cdot,\cdot)$ denote the running time of the approximate oracle $\mathcal{O}_g^{\pm}$ for g. For every $x \in [0,1]^\ell$, $\eta, \gamma, \delta \in (0,1)$, with probability at least $1-\delta$, 2 satisfies the guarantees of a separation oracle for $\mathcal{L}_g$ (with respect to $[0,1]^\ell$). The algorithm makes ℓ queries to $\mathcal{O}_g^{\pm}$ with parameters $\tau^2/2$ and δ/ℓ, where $\tau = \min\{\eta/4\ell, \gamma/2\ell\}$, and its running time is $\ell \cdot \Big(\Phi_g(\frac{\tau^2}{2}, \delta/\ell) + \log \ell\Big)$.*

In order to prove the above lemma we will use the following theorem from [29].

Theorem 5.4 ([29, Theorem 61], restated)**.** *Let $g \colon 2^{[\ell]} \to \mathbb{R}$ be a submodular function. For every $x \in [0,1]^\ell$,*

$$\sum_{i=1}^{\ell} \big(g([i]) - g([i-1])\big) x_i \leq \mathcal{L}_g(x) \ .$$

Proof of Lemma 5.1: For every $i \in [\ell]$, let $a_i \stackrel{\text{def}}{=} g([i]) - g([i-1])$, and note that by a union bound over all $i \in [\ell]$, we have that $\max_{i \in [\ell]}\{|g([i]) - \tilde{g}([i])|\} \leq \tau^2/2$, with probability at least $1-\delta$. We henceforth condition on this, and observe that this implies that, for any $y \in [0,1]^\ell$,

$$|\tilde{a}^T y - a^T y| \leq 2\ell \cdot \frac{\tau^2}{2} = \ell\tau^2 \ . \tag{2}$$

We next consider two cases. Assume first that there exists an index $i \in [\ell]$ such that $|\tilde{a}_i| \geq \tau$. That is, assume that the condition in Step 6 of the algorithm does not hold. Then we prove that for every $y \in [0,1]^\ell$ such that $\mathcal{L}_g(y) \leq \mathcal{L}_g(\bar{x})$ it holds that $y \in H$, where H is the halfspace defined in Step 9 of the algorithm.

By Theorem 5.4, we have that $\sum_{i=1}^{\ell} a_i \cdot y_i \leq \mathcal{L}_g(y)$ for every $y \in [0,1]^\ell$. Since $\mathcal{L}_g(y) \leq \mathcal{L}_g(\bar{x})$, we get that

$$\tilde{a}^T y \leq a^T y + \ell\tau^2 \leq \mathcal{L}_g(y) + \ell\tau^2 \leq \mathcal{L}_g(\bar{x}) + \ell\tau^2 \ . \tag{3}$$

By Theorem 5.2, together with the assumption that the coordinates of $\bar{x}$ are sorted,

$$\mathcal{L}_g(\bar{x}) = \sum_{i=1}^{\ell} a_i \cdot \bar{x}_i \leq \sum_{i=1}^{\ell} \tilde{a}_i \cdot \bar{x}_i + \ell\tau^2 = \tilde{\mathcal{L}}_g(\bar{x}) + \ell\tau^2. \tag{4}$$

Combining Equation (3) and Equation (4), and since there exists an i such that $|\tilde{a}_i| \geq \tau$,

$$\tilde{a}^T y \leq \tilde{\mathcal{L}}_g(\bar{x}) + 2\ell\tau^2 \leq \tilde{\mathcal{L}}_g(\bar{x}) + 2\ell\tau\|\tilde{a}\|_2 \ .$$

This implies that y is in H and that for $c = 2\tau\ell\|\bar{a}\|_2$ and $\gamma \leq 2\tau\ell$, H fulfills the requirements stated in Definition 5.3.

Now consider the case that $|\tilde{a}_i| \le \tau$ for all $i \in [\ell]$. It follows that for any $y \in [0,1]^\ell$, $-\ell\tau \le \tilde{a}^T y \le \ell\tau$. In particular, we have that $-\ell\tau \le \tilde{\mathcal{L}}_g(\bar{x}) \le \ell\tau$, which implies that for every $y \in [0,1]^\ell$,

$$\tilde{\mathcal{L}}_g(\bar{x}) - 2\ell\tau \;\le\; -\ell\tau \;\le\; \tilde{a}^T y \; .$$

Therefore, for every $y \in [0,1]^\ell$ we get

$$\tilde{\mathcal{L}}_g(\bar{x}) - 3\ell\tau \;\le\; \tilde{a}^T y - \ell\tau \;\le\; a^T y \;\le\; \mathcal{L}_g(y) \; ,$$

where the second inequality follows from Equation (2), and the last inequality follows from Theorem 5.4. Hence, if we let $x^* = \arg\min_x\{\mathcal{L}_g(x)\}$, we have that

$$\tilde{\mathcal{L}}_g(\bar{x}) \;\le\; \mathcal{L}_g(x^*) + 3\ell\tau \; .$$

By Equation (4) we have that $\mathcal{L}_g(\bar{x}) \le \tilde{\mathcal{L}}_g(\bar{x}) + \ell\tau^2$. Hence,

$$\mathcal{L}_g(\bar{x}) \le \mathcal{L}_g(x^*) + 3\ell\tau + \ell\tau^2 \le \mathcal{L}_g(x^*) + 4\ell\tau \; ,$$

and since by the setting of τ in Step 2 of the algorithm, $\tau \le \eta/4\ell$, we get that $\bar{x}$ satisfies

$$\mathcal{L}_g(\bar{x}) \;\le\; \min_{y\in[0,1]^\ell}\{\mathcal{L}_g(y)\} + \eta.$$

Therefore, with probability at least $1-\delta$ the algorithm satisfies the conditions of a separation oracle with parameters η and γ.

The algorithm performs ℓ queries to the approximate oracle for g with parameters $\tau^2/2$ and δ/ℓ, where $\tau = \min\{\eta/4\ell, \gamma/2\ell\}$. Hence, the running time of the algorithm is $\ell \cdot \Phi_g(\frac{\tau^2}{2}, \frac{\delta}{\ell}) + \ell\log\ell$, as it also sorts the coordinates of $\bar{x}$ (in order to re-index the coordinates). ∎

We can now use the separation oracle for $\mathcal{L}_g$ and apply the following theorem to get an approximate minimum of $\mathcal{L}_g$, which is also an approximate minimum of g.

Theorem 5.5 ([29, Theorem 42], restated)**.** *Let h be a convex function on $\mathbb{R}^\ell$ and let Ω be a convex set with constant min-width*[7] *that contains a minimizer of h. Suppose we have a separation oracle for h and that Ω is contained inside $B_\infty(R) \stackrel{\text{def}}{=} \{\, x : \|x\|_\infty \le R \,\}$, where $R > 0$ is a constant. Then there is an algorithm, which for any $0 < \alpha < 1$ and $\eta > 0$ outputs $x \in \mathbb{R}^\ell$ such that*

$$h(x) - \min_{y\in\Omega}\{h(y)\} \le \eta + \alpha \cdot \left(\max_{y\in\Omega}\{h(y)\} - \min_{y\in\Omega}\{h(y)\}\right) \; .$$

In expectation, the algorithm performs $O\left(\ell \cdot \log\left(\frac{\ell}{\alpha}\right)\right)$ calls to 2, and has expected running time of

$$O\left(\ell \cdot \mathrm{SO}(\eta,\gamma)\log\left(\frac{\ell}{\alpha}\right) + \ell^3 \log^{O(1)}\left(\frac{\ell}{\alpha}\right)\right) \; ,$$

where $\gamma = \Theta\left(\frac{\alpha}{\ell^{3/2}}\right)$ and $\mathrm{SO}(\eta,\gamma)$ denotes the running time of the separation oracle when invoked with parameters η and γ.

Corollary 5.6. *Let $g\colon 2^{[\ell]} \to \mathbb{R}$ be a submodular function. There exists an algorithm that, when given access to $\mathcal{O}_g^{\pm}$, and for input parameters $\xi, \delta \in (0,1)$, returns with probability at least $9/10 - 2\delta$ a value $\nu \in \mathbb{R}$ such that $\nu \le \min_{S\subseteq[\ell]}\{g(S)\} + \xi$.*

The algorithm performs $O\left(\ell^2 \log\left(\frac{\ell M}{\xi}\right)\right)$ calls to $\mathcal{O}_g^{\pm}$ with parameters $\frac{\xi^2}{128\ell^5 M^2}$ and $\frac{\delta}{C\ell^2\log\left(\frac{\ell M}{\xi}\right)}$, where $M \stackrel{\text{def}}{=} \max\left\{2\max_{S\subseteq[\ell]}\{|g(S)|\}, \xi/2\right\}$ and $C > 0$ is an absolute constant. The running time of the algorithm is

$$O\left(\ell^2 \cdot \Phi_g\left(\frac{\xi^2}{128\ell^5 M^2}, \frac{\delta}{C\ell^2\log\frac{\ell M}{\xi}}\right)\log\frac{\ell M}{\xi} + \ell^3\log^{O(1)}\frac{\ell M}{\xi}\right) \; ,$$

where Φ_g is the running time of $\mathcal{O}_g^{\pm}$.

[7] For a compact set $K \subseteq \mathbb{R}^\ell$, the min-width is defined as $\min_{a\in\mathbb{R}^\ell:\ \|a\|_2=1}\max_{x,y\in K}\{\,\langle a, x-y\rangle\,\}$. [29, Definition 41]. In particular, it is not hard to see that the set $K = [0,1]^\ell \subseteq B_\infty(1)$ has unit min-width.

Proof: We refer to the algorithm from Theorem 5.5 as *the minimization algorithm* and apply it to $\mathcal{L}_g$, with 2 as a separation oracle. Once the minimization algorithm returns a point $x \in [0,1]^\ell$, we return the value $\nu = \mathcal{O}^{\pm}_{\mathcal{L}_g}(x, \xi/4, \delta)$.

Let $M' \stackrel{\text{def}}{=} 2\max_{S\subseteq[\ell]}\{|g(S)|\}$, and recall that $\mathcal{L}_g(x) = \mathop{\mathbf{E}}_{t\sim[0,1]}[g(\{\ i : x_i \geq t\ \})]$. Hence, $\max_{x\in[0,1]^\ell}\{\mathcal{L}_g(x)\} - \min_{x\in[0,1]^\ell}\{\mathcal{L}_g(x)\} \leq M'$. Setting $\alpha = \xi/(2M)$ and $\eta = \xi/4$ ensures that $0 < \alpha < 1$ and that

$$\eta + \alpha \cdot \left(\max_{x\in[0,1]^\ell}\{\mathcal{L}_g(x)\} - \min_{x\in[0,1]^\ell}\{\mathcal{L}_g(x)\}\right) \leq \eta + \alpha M' \leq 3\xi/4 . \tag{5}$$

The minimization algorithm invokes the separation oracle $C_1 \cdot \ell\log(\ell/\alpha) = C_1 \cdot \ell\log(\ell M/\xi)$ times in expectation, for some constant C_1. We convert this to a worst-case bound as follows. If at some point the number of calls to the separation oracle exceeds $20C_1 \cdot \ell\log(\ell M/\xi)$, then we halt and return fail. Similarly, the algorithm runs in expected time

$$T \stackrel{\text{def}}{=} C_2 \cdot \left(\ell \cdot \mathrm{SO}(\eta,\gamma)\log\left(\frac{\ell}{\alpha}\right) + \ell^3\log^{C_3}\left(\frac{\ell}{\alpha}\right)\right)$$

for some absolute constants $C_2, C_3 > 0$. If at some point the running time exceeds $20T$, then we also halt and return fail. By Markov's inequality, both events each happen with probability at most $1/20$, and therefore by a union bound our algorithm halts and outputs fail with probability at most $1/10$. Hence, every time the minimization algorithm calls the separation oracle with parameters η and γ we invoke 2 with parameters η, γ and $\delta' = \frac{\delta}{10C_1\ell\log\left(\frac{\ell M}{\xi}\right)}$. Therefore, by Lemma 5.1, with probability at least $1 - 1/10 - \delta$ all the calls to 2 satisfy the guarantee of a separation oracle for $\mathcal{L}_g$ with parameters η and γ. By Theorem 5.5 and Equation (5), with probability at least $9/10 - \delta$ the minimization algorithm returns a point x such that

$$\mathcal{L}_g(x) - \min_{y\in[0,1]^\ell}\{\mathcal{L}_g(y)\} \leq \eta + \alpha \cdot \left(\max_{y\in[0,1]^\ell}\{\mathcal{L}_g(y)\} - \min_{y\in[0,1]^\ell}\{\mathcal{L}_g(y)\}\right) \leq \frac{3\xi}{4} ,$$

and with probability at least $9/10 - 2\delta$ the value ν satisfies

$$\nu \leq \min_{y\in[0,1]^\ell}\{\mathcal{L}_g(y)\} + \xi ,$$

as desired.

By the above settings and by Lemma 5.1 we get that $\tau = \frac{\xi}{8\ell^{5/2}M}$ so the running time of each invocation of the separation oracle (recall that each such invocation involves ℓ calls to $\mathcal{O}^{\pm}_g$) is

$$\ell \cdot \Phi_g\left(\frac{\tau^2}{2}, \frac{\delta'}{\ell}\right) + \ell\log\ell = \ell\left(\Phi_g\left(\frac{\xi^2}{128\ell^5M^2}, \frac{\delta}{10C_1\ell^2\log\frac{\ell M}{\xi}}\right) + \log\ell\right) .$$

Since the evaluation of ν in the final step is negligible in the running time of the minimization algorithm, we get that the overall time complexity is

$$O\left(\ell^2 \cdot \Phi_g\left(\frac{\xi^2}{128\ell^5M^2}, \frac{\delta}{10C_1\ell^2\log\frac{\ell M}{\xi}}\right)\log\frac{\ell M}{\xi} + \ell^2\log\ell\cdot\log\frac{\ell M}{\xi} + \ell^3\log^{O(1)}\frac{\ell M}{\xi}\right) ,$$

which gives the stated bound, recalling that $\ell^2\log\ell\cdot\log\frac{\ell M}{\xi} = O(\ell^2\log^2\frac{\ell M}{\xi}) = O(\ell^3\log^{O(1)}\frac{\ell M}{\xi})$.

■

Corollary 5.7. *There exists an algorithm that, when given query access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$ and a partition $\mathcal{I} = \{I_1,\ldots,I_\ell\}$ of $[n]$ into ℓ parts, as well as input parameters $k \in \mathbb{N}, \epsilon, \xi \in (0,1)$, satisfies the following. It has query complexity $\tilde{O}\left(\max\left(\frac{\ell^{12}}{\xi^4}, \frac{\ell^{16}\epsilon^4}{k^4\xi^4}\right)\right)$, and distinguishes with probability at least $5/6$ between the following two cases:*

1. *There exists a set $S \subseteq [\ell]$ such that $|S| \geq \ell - k$ and $h(S) \leq \epsilon$.*

2. *For every set S such that $|S| \geq \ell - 2(1+\frac{\xi}{\epsilon})k$, $h(S) > 2\epsilon + 2\xi$,*

where $h\colon 2^{[\ell]} \to \mathbb{R}$ is defined as $h(S) \stackrel{\text{def}}{=} \mathbf{Inf}_f(\phi_{\mathcal{I}}(S))$.
Moreover, the second item can be strengthened so that it holds for functions f that satisfy the following: (i) for every set S such that $|S| \geq \ell - k$, $h(S) > 2\epsilon + 2\xi$ and (ii) for every set S such that $|S| \geq \ell - 2(1+\frac{\xi}{\epsilon})k$, $h(S) > \epsilon + 2\xi$.

Proof: We apply Corollary 5.6 to $h'\colon 2^{[\ell]} \to \mathbb{R}$, defined as in 1 by $h'(S) \stackrel{\text{def}}{=} h(S) - \frac{\epsilon}{k}|S|$, with ξ, $M \stackrel{\text{def}}{=} \max\left(2\max(2, \frac{\epsilon\ell}{k}), \xi/2\right) = 4\max(1, \frac{\epsilon\ell}{2k})$, and $\delta \stackrel{\text{def}}{=} \frac{1}{30}$. In order to do so, we need to simulate an approximate oracle for h' (as defined in Definition 4.1). Since $h(S) = \mathbf{Inf}_f(\phi_{\mathcal{I}}(S))$, in order to estimate $h'(S)$ within an additive approximation of τ' with probability at least $1-\delta'$, it is sufficient to estimate $\mathbf{Inf}_f(\phi_{\mathcal{I}}(S)) \in [0,2]$ within an additive approximation of τ' with probability at least $1-\delta'$ (indeed, the additional term $\frac{\epsilon}{k}|S|$ can be computed exactly). By Chernoff bounds, this can be done with $\Phi_h(\tau', \delta') = O(\frac{1}{\tau'^2}\log\frac{1}{\delta'})$ queries to f.

This yields an approximate oracle $\mathcal{O}_h^{\pm}$, and therefore $\mathcal{O}_{h'}^{\pm}$ which can be provided to the algorithm of Theorem 4.3 (resulting in a success probability at least $9/10 - 2\delta = 5/6$). The resulting query complexity is

$$O\left(\ell^2 \cdot \Phi_h\left(\frac{\xi^2}{\ell^5 M^2}, \frac{1}{10C_1\ell^2\log\frac{\ell M}{\xi}}\right)\log\frac{\ell M}{\xi} + \ell^2\log\ell + \ell^3\log^{O(1)}\frac{\ell M}{\xi}\right)$$

which, given the above expression for Φ_h, can be bounded as follows.

- If $\epsilon < \frac{2k}{\ell}$, so that $M = 4$, this simplifies as

$$O\left(\frac{\ell^{12}}{\xi^4}\log^2\frac{\ell}{\xi}\right).$$

- If $\epsilon \geq \frac{2k}{\ell}$, which implies that $M = \frac{2\epsilon\ell}{k}$, this becomes

$$O\left(\frac{\ell^{16}\epsilon^4}{k^4\xi^4}\log^2\frac{\ell}{\xi}\right).$$

Observing that the function h is indeed a non-negative submodular function (and that h' is also submodular since it is the sum of a submodular function and a modular function) allows us to conclude by Theorem 4.3. ■

In particular, setting $\xi = \epsilon/(4k)$ we get the following:

Corollary 5.8. *There exists an algorithm that, given query access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$, a fixed partition $\mathcal{I}$ of $[n]$ into $\ell = O(k^2)$ parts, and parameters $k \geq 1$ and $\epsilon \in (0,1)$, satisfies the following. The query complexity of the algorithm is $\tilde{O}\left(\frac{k^{28}}{\epsilon^4} + k^{32}\right) = \mathrm{poly}(k, 1/\epsilon)$, and:*

1. *if f $\frac{\epsilon}{2}$-approximates being a k-part junta with respect to $\mathcal{I}$, then the algorithm accepts with probability at least $\frac{5}{6}$;*

2. *if f $\frac{5}{4}\epsilon$-violates being a $2k$-part junta with respect to $\mathcal{I}$, then the algorithm rejects with probability at least $\frac{5}{6}$.*

Moreover, the second item can be strengthened to "simultaneously $\left(1+\frac{1}{4k}\right)\epsilon$-violates being a k-part junta and $(1+\frac{1}{2k})\frac{\epsilon}{2}$-violates being a $2k$-part junta."

Proof: By applying Corollary 5.7 with $\xi = \epsilon/(4k)$, we get an algorithm that distinguishes between (1) there exists a set $S \subseteq [\ell]$ such that $|S| \geq \ell - k$ and $h(S) \leq \epsilon$; and (2) either (i) for every set S such that $|S| \geq \ell - k$, $h(S) > 2\left(1+\frac{1}{4k}\right)\epsilon$ or (ii) for every set S such that $|S| \geq \ell - (2k+\frac{1}{2})$, $h(S) > \epsilon(1+\frac{1}{2k})$. Since $|S|$ is always an integer, the condition $|S| \geq \ell - (2k+\frac{1}{2})$ is equivalent to $|S| \geq \ell - 2k$ in (2)(ii). This implies the guarantees of the corollary, by the correspondence with partition juntas (Definition 3.1) and using for simplicity that $1 + \frac{1}{4k} \leq \frac{5}{4}$ as $k \geq 1$. (The claimed query complexity is immediate from Corollary 5.7.) ■

The tolerant junta testing theorem (Theorem 1.1) follows immediately from the above, together with Proposition 3.4. With probability at least 5/6, a random partition of the variables into $\ell \stackrel{\text{def}}{=} 192k^2$ parts will have the right guarantees, reducing the problem to distinguishing between $\frac{\epsilon}{2}$-approximating being a k-part junta vs. $\frac{5}{4}\epsilon$-violating being a $2k$-part junta (with regard to this random partition). Overall, by a union bound, the result is therefore correct with probability at least 2/3.

6 A tradeoff between tolerance and query complexity

In this section, we show how to obtain a smooth tradeoff between the amount of tolerance and the query complexity. Formally, we prove Theorem 1.2, restated below.

Theorem 1.2. *There exists an algorithm that, given query access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$ and parameters $k \geq 1$, $\epsilon \in (0,1)$ and $\rho \in (0,1)$, satisfies the following.*

- *If f is $\rho\epsilon/16$-close to some k-junta, then the algorithm accepts with high constant probability.*
- *If f is ϵ-far from every k-junta, then the algorithm rejects with high constant probability.*

The query complexity of the algorithm is $O\Big(\frac{k \log k}{\epsilon\rho(1-\rho)^k}\Big)$.

As discussed in the introduction, this in particular implies the two following results. Setting $\rho = \Omega(1)$, we obtain a tolerant tester that distinguishes between functions $O(\epsilon)$-close to $\mathcal{J}_k$ and functions ϵ-far from $\mathcal{J}_k$, with query complexity $2^{O(k)}/\epsilon$, an improvement over the naive tester from Section 3. On the other hand, choosing $\rho = O(1/k)$ yields a weakly tolerant tester that distinguishes functions $O(\epsilon/k)$-close to $\mathcal{J}_k$ from those ϵ-far from $\mathcal{J}_k$, with query complexity $\tilde{O}\big(k^2/\epsilon\big)$ – thus matching the guarantees provided in [21].

6.1 Useful bounds on the expected influence of a random ρ-subset of a set

In this subsection we formally define the ρ-subset influence of a set and prove that for every set $J \subseteq [\ell]$, its ρ-subset influence is at least $\frac{\rho}{3} \cdot \mathbf{Inf}_f(\phi_{\mathcal{I}}(J))$ and at most $\mathbf{Inf}_f(\phi_{\mathcal{I}}(J))$. Then in the next subsection we provide an algorithm that simultaneously estimates the ρ-subset influence of all subsets J of $[\ell]$ of size $\ell - k$. The query complexity of the algorithm is $O\Big(\frac{k \log k}{\epsilon\rho(1-\rho)^k}\Big)$. We start with a few definitions and notations.

Definition 6.1. *For any $\rho \in (0,1)$ and any set R, we denote by $S \sim_\rho R$ the random ρ-biased subset of R, resulting from including independently each $i \in R$ in S with probability ρ. We refer to such a set S as a* random ρ-subset of R.

Definition 6.2. *For a partition $\mathcal{I} = \{I_1, \dots, I_\ell\}$ and a set $J \subseteq [\ell]$ we refer to the expected value of the influence of a random ρ-biased subset of J, $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_{\mathcal{I}}(S))]$, as the* ρ-subset influence *of J (with respect to $\mathcal{I}$).*

The next lemma describes the connection between the influence of a set J and its ρ-subset influence.

Lemma 6.1. *Let $\mathcal{I} = \{I_1, \dots, I_\ell\}$ be a partition of $[n]$. Then, for any $J \subseteq [\ell]$,*

$$\frac{\rho}{3}\mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) \leq \mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_{\mathcal{I}}(S))] \leq \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)).$$

Proof: The upper bound is immediate by monotonicity of the influence, as $\mathbf{Inf}_f(\phi_{\mathcal{I}}(S)) \leq \mathbf{Inf}_f(\phi_{\mathcal{I}}(J))$ for all $S \subseteq J$. As for the lower bound, let $j = |J|$ and observe that

$$\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_{\mathcal{I}}(S))] = \sum_{s=1}^{j} \sum_{S\subseteq J: |S|=s} \rho^s(1-\rho)^{j-s} \cdot \mathbf{Inf}_f(\phi_{\mathcal{I}}(S)). \tag{6}$$

We will lower bound the sum $\sum_{S\subseteq J:|S|=s} \mathbf{Inf}_f(\phi_{\mathcal{I}}(S))$ for each s separately. In order to do so we define a *legal collection of s-covers for a set J*:

Definition 6.3. *Let J be a set of j elements, and for any $s \in [j]$ consider the family $\binom{J}{s}$ of all $\binom{j}{s}$ subsets of J of size s. We shall say that $\mathcal{C} \subseteq \binom{J}{s}$ is a s-*cover *of J if $\bigcup_{Y \in \mathcal{C}} Y = J$. We shall say that a collection of s-covers $\mathfrak{C}_J = \{\mathcal{C}_1, \dots, \mathcal{C}_r\}$ is a* legal collection of s-covers for J *if each $\mathcal{C}_t \in \mathfrak{C}_J$ is a cover of J and these s-covers are disjoint.*

Thus, we are interested in showing that there exists a legal collection of s-covers for J whose size m is "as big as possible." This is what the next claim guarantees, establishing that there exists such a cover achieving the optimal size:

Claim 6.2. *For any set J of j elements, there exists a legal collection of s-covers $\mathfrak{C}_J$ for J of size at least*

$$|\mathfrak{C}_J| \geq \left\lfloor \frac{\binom{j}{s}}{\lceil \frac{j}{s} \rceil} \right\rfloor .$$

(Moreover, this bound is tight.)

Claim 6.2 follows from a result due to Baranyai [5] on factorization of regular hypergraphs: for completeness, we state this result, and describe how to derive the claim from it, in Appendix B. Observe that if s divides j then $\left\lfloor \frac{\binom{j}{s}}{\lceil \frac{j}{s} \rceil} \right\rfloor = \binom{j-1}{s-1}$; and otherwise

$$\left\lfloor \frac{\binom{j}{s}}{\lceil \frac{j}{s} \rceil} \right\rfloor = \left\lfloor \frac{\frac{j}{s}}{\lceil \frac{j}{s} \rceil} \binom{j-1}{s-1} \right\rfloor \geq \left\lfloor \frac{\frac{j}{s}}{\frac{j}{s}+1} \binom{j-1}{s-1} \right\rfloor \geq \left\lfloor \frac{1}{2} \binom{j-1}{s-1} \right\rfloor \geq \frac{1}{3} \binom{j-1}{s-1} .$$

Therefore,

$$\begin{aligned} \sum_{S \subseteq J : |S| = s} \mathbf{Inf}_f(\phi_{\mathcal{I}}(S)) = \sum_{S \in \binom{J}{s}} \mathbf{Inf}_f(\phi_{\mathcal{I}}(S)) &\geq \sum_{\mathcal{C} \in \mathfrak{C}_J} \sum_{S \in \mathcal{C}} \mathbf{Inf}_f(\phi_{\mathcal{I}}(S)) \\ &\geq |\mathfrak{C}_J| \cdot \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) \geq \frac{1}{3} \binom{j-1}{s-1} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)). \end{aligned}$$

Plugging the above into Equation (6), we obtain that

$$\begin{aligned} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) &\geq \sum_{s=1}^{j} \rho^s (1-\rho)^{j-s} \cdot \left(\frac{1}{3} \binom{j-1}{s-1} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) \right) \\ &= \frac{\rho}{3} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) \sum_{s=1}^{j} \binom{j-1}{s-1} \rho^{s-1} (1-\rho)^{j-s} \\ &= \frac{\rho}{3} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)) (\rho + (1-\rho))^{j-1} = \frac{\rho}{3} \mathbf{Inf}_f(\phi_{\mathcal{I}}(J)), \end{aligned}$$

which concludes the proof. ∎

6.2 Approximation of the ρ-subset influences

We now describe and analyze an algorithm that given a partition $\mathcal{I} = \{I_1, \dots, I_\ell\}$, allows to simultaneously get good estimates of the ρ-subset influences of all subset $J \in \binom{[\ell]}{\ell-k}$. This algorithm is the main building block of the tolerant junta tester of Theorem 1.2.

Algorithm 3 Simultaneously Approximate ρ-subset Influence $(\mathcal{O}_f, \rho, \epsilon, \gamma, k, \ell, \mathcal{I})$

```
1: Set m = C·k log ℓ / (γ²ερ(1−ρ)^k), where C ≥ 1 is an absolute constant.     ▷ C ≥ 256 ln 2 is sufficient.
2: for i = 1 to m do
3:     Let S^i ∼_ρ [ℓ].
4:     Pick x^i ∈ {−1,1}^n uniformly at random, and let z^i be obtained by flipping independently each
       coordinate of x^i_{φ_I(S^i)} with probability 1/2
5:     Set y ← x^i_{φ_I(S̄^i)} ⊔ z^i.
6:     Set ϑ_{S^i} ← 𝟙_{f(x^i)≠f(y^i)} .                        ▷ ϑ_{S^i} is 1 if f(x^i) ≠ f(y^i), 0 otherwise.
7: end for
8: Let 𝒮 be the multiset of subsets S^1, ..., S^m.
9: for every J ∈ ([ℓ] choose ℓ−k) do
10:    Let 𝒮_J ⊆ 𝒮 denote the subset of sets S∈ 𝒮 such that S ⊆ J .
11:    Let ν^ρ_J ← (1/|𝒮_J|) Σ_{S∈𝒮_J} ϑ_S be the estimate of the ρ-subset influence of J.
12: end for
```

Lemma 6.3. *Let $\mathcal{I} = \{I_1, \dots, I_\ell\}$ be a partition of $[n]$. For every $\epsilon \in (0,1)$ and $\rho \in (0,1)$, 3 satisfies that, with probability at least $1 - o(1)$, the following holds simultaneously for all sets $J \in \binom{[\ell]}{\ell-k}$ such that $|J| = \ell - k$:*

1. *if $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_\mathcal{I}(S))] > \frac{\rho\epsilon}{3}$, then the estimate ν^ρ_J is within a multiplicative factor of $(1 \pm \gamma)$ of the ρ-subset influence of J.*
2. *if $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_\mathcal{I}(S))] \le \frac{\rho\epsilon}{4}$, then the estimate ν^ρ_J does not exceed $(1+\gamma)\frac{\rho\epsilon}{4}$.*

Proof: Let $m' \stackrel{\text{def}}{=} \frac{1}{2}(1-\rho)^k \cdot m = \frac{Ck\log\ell}{2\gamma^2\epsilon\rho}$. We first claim that for any fixed set $J \subseteq [\ell]$ of size $\ell - k$, with probability at least $1 - o(\ell^{-2k})$, $|\mathcal{S}_J| \ge m'$. To see why this is true, fix some $J \subset [\ell]$ of size $\ell - k$. For every $i \in [m]$, let $\mathbb{1}_{\{S_i \subseteq J\}}$ be an indicator variable which is equal to 1 if and only if $S_i \subseteq J$. Then, for every $i \in [m]$, $\Pr\left[\mathbb{1}_{\{S_i\subseteq J\}} = 1\right] = (1-\rho)^k$. By a Chernoff bound,

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{\{S_i\subseteq J\}} < \frac{1}{2}\cdot(1-\rho)^k\right] \le e^{-\frac{m}{8}(1-\rho)^k} = e^{-\frac{C\cdot k\log\ell}{8\epsilon\rho\gamma^2}} < 2^{-4k\log\ell} ,$$

for a suitable choice of $C \ge 1$. Therefore, by a union bound over all $\binom{\ell}{\ell-k} = \binom{\ell}{k} = 2^{(1+o(1))k\log\ell}$ sets $J \in \binom{[\ell]}{\ell-k}$, it holds that with probability $1 - o(1)$, for every such J, $|\mathcal{S}_J| \ge m'$. We hereafter condition on this.

We now turn to prove the two items of the lemma. Let $X = \{x^1, \dots, x^m\}$ and $Z = \{z^1, \dots, z^m\}$. For a set S^i, $\mathbf{E}_{x^i,z^i}[\vartheta_S] = \mathbf{Inf}_f(\phi_\mathcal{I}(S))$. Hence, by the definition of ν^ρ_J in Step 11 of the algorithm,

$$\begin{aligned}\mathbf{E}[\nu^\rho_J] &= \mathbf{E}_{\mathcal{S},X,Z}\left[\frac{1}{|\mathcal{S}_J|}\sum_{S\in\mathcal{S}_J}\vartheta_S\right] = \mathbf{E}_\mathcal{S}\left[\frac{1}{|\mathcal{S}_J|}\sum_{S^i\in\mathcal{S}_J}\mathbf{E}_{x^i,y^i}[\vartheta_{S^i}]\right] \\ &= \sum_{S\subseteq J}\Pr[S\in\mathcal{S}]\cdot\mathbf{Inf}_f(\phi_\mathcal{I}(S)) = \mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right]. \qquad (7)\end{aligned}$$

Consider a set J with $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_\mathcal{I}(S))] > \frac{\rho\epsilon}{3}$. By Equation (7), $\mathbf{E}[\nu^\rho_J] > \frac{\rho\epsilon}{3}$. Therefore, by a Chernoff bound, and since for every J, $|\mathcal{S}_J| \ge m' = \frac{Ck\log\ell}{2\gamma^2\epsilon\rho}$,

$$\begin{aligned}\Pr\left[\left|\nu^\rho_J - \mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right]\right| > \gamma\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right]\right] &\le 2e^{-\frac{|\mathcal{S}_J|\gamma^2\cdot\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right]}{3}} \\ &\le 2e^{-\frac{m'\gamma^2\epsilon\rho}{9}} < 2^{-4k\log\ell},\end{aligned}$$

again for a suitable choice of the constant $C \geq 1$. By taking a union bound over all subsets $J \in \binom{[\ell]}{\ell-k}$, we get that, with probability at least $1 - o(1)$, for every J such that $\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right] > \frac{\rho\epsilon}{3}$, it holds that $\nu_J^\rho \in (1 \pm \gamma) \cdot \mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right]$.

Now consider a set $J \subseteq [\ell]$ such that $\left|\bar{J}\right| > \ell - k$ and $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_\mathcal{I}(S))] \leq \frac{\rho\epsilon}{4}$. By a Chernoff bound:

$$\Pr\left[\,\nu_J^\rho > (1+\gamma)\frac{\rho\epsilon}{4}\,\right] \leq e^{-\frac{\gamma^2}{3}\frac{\rho\epsilon}{4}|\mathcal{S}_J|} \leq e^{-\frac{\gamma^2\rho\epsilon}{12}m'} < 2^{-4k\log\ell}.$$

The claim follows by taking a union bound over all subsets $J \in \binom{[\ell]}{\ell-k}$ for which $\mathbf{E}_{S\sim_\rho J}[\mathbf{Inf}_f(\phi_\mathcal{I}(S))] \leq \rho\epsilon/4$. Overall, the conclusions above hold with probability at least $1 - o(1)$, as stated. ■

6.3 Tradeoff between tolerance and query complexity

We now describe how the algorithm from the previous section lets us easily derive the tolerant tester of Theorem 1.2.

Algorithm 4 ρ-Tolerant Junta Tester $(\mathcal{O}_f, \epsilon, \rho, k)$

Create a random partition $\mathcal{I}$ of $\ell = 24k^2$ parts by uniformly and independently assigning each coordinate to a part.
Run 3 with the partition $\mathcal{I}$, $\ell = 24k^2$ and $\gamma = 1/8$.
if there is a set $J \subset [\ell]$ of size $\ell - k$ such that $\nu_J^\rho \leq \frac{9\rho\epsilon}{32}$ **then**
    return accept.
end if
return reject.

Proof of Theorem 1.2: Given Proposition 3.4 it is sufficient to consider a partition $\mathcal{I}$ of size $\ell = 24k^2$ and show that 4 distinguishes with probability at least 5/6 between the following two cases.

1. f $\frac{\rho\epsilon}{8}$-approximates being a k-part junta with respect to $\mathcal{I}$;
2. f $\frac{\epsilon}{2}$-violates being a k-part junta with respect to $\mathcal{I}$

Suppose first that f $\frac{\rho\epsilon}{8}$-approximates being a k-part junta with respect to $\mathcal{I}$. Then by Definition 3.1, there exists a set $J \in \binom{[\ell]}{\ell-k}$ such that $\mathbf{Inf}_f(\phi(J)) \leq \frac{\rho\epsilon}{4}$. By Lemma 6.1, $\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right] \leq \frac{\rho\epsilon}{4}$, and by Lemma 6.3, we have that with probability at least $1 - o(1)$, the estimate ν_J^ρ is at most $(1 + 1/8)\frac{\epsilon\rho}{4} \leq \frac{9\epsilon\rho}{32}$. Therefore, 4 will return accept when considering J.

Consider now the case where f $\frac{\epsilon}{2}$-violates being a k-part junta with respect to $\mathcal{I}$. Hence, by Definition 3.1, every set $J \in \binom{[\ell]}{\ell-k}$ is such that $\mathbf{Inf}_f(\phi(J)) > \epsilon$, and by Lemma 6.1, we have that $\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right] \geq \frac{\rho}{3}\mathbf{Inf}_f(\phi(J)) > \frac{\rho\epsilon}{3}$. Therefore, by Lemma 6.3, with probability at least $1 - o(1)$, for every $J \in \binom{[\ell]}{\ell-k}$

$$\nu_J^\rho \geq \frac{7}{8}\mathbf{E}_{S\sim_\rho J}\left[\mathbf{Inf}_f(\phi_\mathcal{I}(S))\right] > \frac{9\rho\epsilon}{32}\;.$$

Thus, with probability at least $1 - o(1)$, 4 will reject f. ■

7 "Instance-adaptive" tolerant isomorphism testing

In this section, we show how the machinery developed in Section 6, and more precisely the algorithm from Theorem 1.2, can be leveraged to obtain *instance-adaptive tolerant isomorphism testing* between two unknown Boolean functions f and g, as defined below.

We begin with some notation: for $f, g\colon \{-1,1\}^n \to \{-1,1\}$, we denote by $\operatorname{distiso}(f,g)$ the distance between f and the closest isomorphism of g, that is $\operatorname{distiso}(f,g) \overset{\text{def}}{=} \min_{\pi\in\mathcal{S}_n} \operatorname{dist}(f, g\circ\pi)$. Given query

access to two unknown Boolean functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$ and a parameter $\epsilon \in (0,1)$, isomorphism testing then amounts to distinguishing between (i) $\operatorname{distiso}(f,g) = 0$; and (ii) $\operatorname{distiso}(f,g) > \epsilon$.[8]
Our result will be parameterized in terms of the *junta degree* of the unknown functions f and g, formally defined below:

Definition 7.1 (Junta degree)**.** Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and $\gamma \in [0,1]$ a parameter. We define the *γ-junta degree of f* as the smallest integer k such that f is γ-close to being a k-junta, that is

$$k^*(f,\gamma) \stackrel{\text{def}}{=} \min\{\, k \in [n] : \ \operatorname{dist}(f, \mathcal{J}_k) \le \gamma \,\}.$$

Finally, we extend this definition to two functions f, g by setting $k^*(f,g,\gamma) = \min(k^*(f,\gamma), k^*(g,\gamma))$.

With this terminology in hand, we can restate Theorem 1.4:

Theorem 7.2 (Theorem 1.4, rephrased)**.** *There exist absolute constants $c \in (0,1)$, $\epsilon_0 \in (0,1)$ and a tolerant testing algorithm for isomorphism of two unknown functions f and g with the following guarantees. On inputs $\epsilon \in (0,\epsilon_0]$, $\delta \in (0,1]$, and query access to functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$:*

- *if* $\operatorname{distiso}(f,g) \le c\epsilon$*, then it outputs* accept *with probability at least* $1-\delta$*;*
- *if* $\operatorname{distiso}(f,g) > \epsilon$*, then it outputs* reject *with probability at least* $1-\delta$*.*

The query complexity of the algorithm satisfies the following, where $k^* = k^*(f,g,\frac{\rho c\epsilon}{16})$ is the $\frac{\rho c\epsilon}{16}$-junta degree of f and g:

- *it is $\tilde{O}\big(2^{\frac{k^*}{2}} \frac{1}{\epsilon} \log\frac{1}{\delta}\big)$ with probability at least $1-\delta$;*
- *it is always at most $\tilde{O}\big(2^{\frac{n}{2}} \frac{1}{\epsilon} \log\frac{1}{\delta}\big)$.*

Moreover, one can take $c = \frac{1}{1750}$, and $\epsilon_0 \stackrel{\text{def}}{=} \frac{16}{15}(5-2\sqrt{6}) \simeq 0.108$.

7.1 Proof of Theorem 7.2

As described in Section 1.2, our algorithm first performs a linear search on k, invoking at each step the tolerant tester of Section 6 with parameter ϵ', to obtain (with high probability) a value k^* such that $k^*(f,g,\epsilon') \le k^* \le k^*(f,g,\frac{\rho\epsilon'}{16})$. In the second stage, it calls a "noisy sampler" to obtain uniformly random labeled samples from the "cores" of the k^*-juntas closest to f and g (both notions are defined formally in Section 7.1.2), and robustly tests isomorphism between them. We accordingly divide this section in two, proving respectively these two statements:

Lemma 7.3. *There exists an algorithm (5) with the following guarantees. On inputs $\epsilon', \delta \in (0,1)$ and query access to $f, g\colon \{-1,1\}^n \to \{-1,1\}$, it returns a value $0 \le k \le n$, such that:*

- *with probability at least $1-\delta$, we have that:*
  (i) $k^*(f,g,\epsilon') \le k \le k^*(f,g,\frac{\rho\epsilon'}{16})$*;*
  (ii) the algorithm performs $O\Big(2^{\frac{k}{2}+o(k)} \cdot \frac{1}{\epsilon}\log\frac{1}{\delta}\Big)$ *queries;*
- *the algorithm performs at most* $O\big(2^{\frac{n}{2}+o(n)} \cdot \frac{1}{\epsilon}\log\frac{1}{\delta}\big)$ *queries.*

Proposition 7.4. *There exists an algorithm (6) with query complexity $\tilde{O}\Big(\frac{2^{k/2}}{\epsilon}\Big)$ for testing of isomorphism of two unknown functions f and g, under the premise that f is close to $\mathcal{J}_k$. More precisely, there exist absolute constants $c > 0$ and $\epsilon_0 \in (0,1]$ such that, on inputs $k \in \mathbb{N}$, $\epsilon \in (0,\epsilon_0]$ and query access to functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$, the algorithm has the following guarantees. Conditioned on $\operatorname{dist}(f,\mathcal{J}_k) \le c\epsilon$, it holds that:*

[8] Phrased differently, this is testing the property $\mathcal{P} = \Big\{\, (f, f\circ\pi) : \ f \in 2^{2^n}, \pi \in \mathcal{S}_n \,\Big\} \subseteq 2^{2^n} \times 2^{2^n}$.

- *if* $\mathrm{distiso}(f,g) \leq c\epsilon$*, then it outputs* accept *with probability at least* $8/15$*;*
- *if* $\mathrm{distiso}(f,g) > \epsilon$*, then it outputs* reject *with probability at least* $8/15$*.*

Moreover, one can take $c = \frac{1}{1750}$*, and* $\epsilon_0 \stackrel{\text{def}}{=} \frac{16}{15}(5-2\sqrt{6}) \simeq 0.108$.

Theorem 7.2 follows by the combination of Lemma 7.3 and Proposition 7.4.

Proof of Theorem 7.2: Let $\rho \stackrel{\text{def}}{=} 1-\frac{1}{\sqrt{2}}$, and $\epsilon' = c\epsilon$. The algorithm proceeds as follows: it first invokes 5 with inputs $f, g, \epsilon', \delta/2$, and gets by Lemma 7.3, a value $1 \leq k^* \leq n$ such that $k^*(f,g,\epsilon') \leq k^* \leq k^*(f,g,\frac{\rho\epsilon'}{16})$ with probability at least $1-\frac{\delta}{2}$. In particular, conditioning on this we are guaranteed that either f or g is ϵ'-close to some k^*-junta (i.e., by our choice of c, one of the functions is $c\epsilon$-close to $\mathcal{J}_{k^*}$). It then calls 6 with inputs f, g, k^*, ϵ independently $O(\log\frac{1}{\delta})$ times (for probability amplification from $8/15$ to $1-\frac{\delta}{2}$), and accepts if and only if the majority of these executions returned accept. The correctness of the algorithm follows from Proposition 7.4 and the bound on the query complexity follows from the bounds in Lemma 7.3 and Proposition 7.4. ■

7.1.1 Linear search: finding k^*.

Let $\mathcal{T}$ denote the algorithm of Theorem 1.2, with probability of success amplified by standard techniques to $1-\delta$ for any $\delta \in (0,1]$ (at the price of a factor $O\big(\log\frac{1}{\delta}\big)$ in its query complexity); and write $q_{\mathcal{T}}(k,\epsilon,\rho,\delta) = O\Big(\frac{k\log k}{\epsilon\rho(1-\rho)^k}\log\frac{1}{\delta}\Big)$ for its query complexity. 5, given next, performs the linear search for k^*: we then analyze its correctness and query complexity.

Algorithm 5 Junta Degree Finder$(\mathcal{O}_f, \mathcal{O}_g, \epsilon', \rho, \delta)$

Set $\rho \leftarrow 1-\frac{1}{\sqrt{2}}$ and let $\mathcal{T}$ be the algorithm of Theorem 1.2.
for $k = 0$ to n **do**
  Call $\mathcal{T}$ on f with parameters k, ϵ', ρ, and $3\delta/(2\pi^2(k+1)^2)$.
  Call $\mathcal{T}$ on g with parameters k, ϵ', ρ, and $3\delta/(2\pi^2(k+1)^2)$.
  if either call to $\mathcal{T}$ returned accept **then return** k.
  end if
end for
return n

Proof of Lemma 7.3: By a union bound, all executions of $\mathcal{T}$ will be correct with probability at least $1-2\sum_{j=1}^{\infty}\frac{3\delta}{2\pi^2 j^2} = 1-\frac{\delta}{2}$. Conditioning on this, the tester will accept for some k between $k^*(f,g,\epsilon')$ and $k^*(f,g,\rho\epsilon'/16)$. This is true since as long as we invoke $\mathcal{T}$ with values k such that f and g are ϵ'-far from $\mathcal{J}_k$, both invocations of $\mathcal{T}$ will reject. Therefore, once we accept, we have that either f or g is at least ϵ'-close to $\mathcal{J}_k$. Hence, $k \geq k^*(f,g,\epsilon')$. Also, $\mathcal{T}$ is guaranteed to accept on some k' whenever invoked on a function that is $\rho\epsilon'/16$-close to $\mathcal{J}_{k'}$. By definition, $k^*(f,g,\rho\epsilon'/16)$ is such a k' for either f or g; hence, $k \leq k^*(f,g,\rho\epsilon'/16)$.

In the case that all the executions of $\mathcal{T}$ returned correctly, the query complexity is

$$q(\epsilon,f,g) = \sum_{k=0}^{k^*(f,g,\frac{\rho\epsilon'}{16})} 2q_{\mathcal{T}}\bigg(k,\epsilon',\rho,\frac{3\delta}{2\pi^2(k+1)^2}\bigg).$$

By the expression of $q_{\mathcal{T}}$, we get that $q(\epsilon,f,g)$ is upper bounded by

$$q(\epsilon,f,g) \leq \frac{O(1)}{\epsilon\rho}\sum_{k=1}^{k^\star}\frac{k\log k\log\frac{k}{\delta}}{(1-\rho)^k} \leq \frac{O(1)}{\epsilon}(k^\star\log k^\star)^2 2^{k^\star\log\frac{1}{1-\rho}}\log\frac{1}{\delta}$$

where $k^\star \stackrel{\text{def}}{=} k^*(f,g,\frac{\rho\epsilon'}{16})$. In particular, from the choice of ρ, we get $q(\epsilon,f,g) \leq O\Big(2^{\frac{k^\star}{2}+o(k^\star)}\frac{1}{\epsilon}\log\frac{1}{\delta}\Big)$.

(If not all the executions of the tester are successful, in the worst case the algorithm considers all possible values of k, before finally returning n. In this case, the query complexity is similarly bounded by $O\big(2^{\frac{n}{2}+o(n)}\frac{1}{\epsilon}\log\frac{1}{\delta}\big)$.)

■

7.1.2 Noisy samplers and core juntas.

For a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ we denote by $f_k : \{-1,1\}^n \to \{-1,1\}$ the k-junta closest to f. That is, the function $h \in \mathcal{J}_k$ such that $\operatorname{dist}(f,h) = \operatorname{dist}(f,\mathcal{J}_k)$ (if this function is not unique, then we define f_k to be the first according to lexicographic order). Moreover, following Chakraborty et al. [16], for a k-junta $h \in \mathcal{J}_k$ (where we assume without loss of generality that h depends on exactly k variables) we define the *core* of h, as follows. The core of h, denoted $\mathsf{core}_h\colon \{-1,1\}^k \to \{-1,1\}$, is the restriction of h to its relevant variables (where these variables are numbered according to the natural order); so that for some $i_1 \le \dots \le i_k \in [n]$ we have

$$h(x) = \mathsf{core}_h(x_{i_1}, \dots, x_{i_k})$$

for every $x \in \{-1,1\}^n$.

Definition 7.5 ([16, Definition 1])**.** Let $g\colon \{-1,1\}^k \to \{-1,1\}$ be a function and let $\eta,\mu \in [0,1)$. An (η,μ)*-noisy sampler for* g is a probabilistic algorithm $\tilde{g}$ that on each execution outputs a pair $(x,a) \in \{-1,1\}^k \times \{-1,1\}$ such that

(i) For all $y \in \{-1,1\}^k$, $\Pr[\,x = y\,] \in \big[\frac{1-\mu}{2^k}, \frac{1+\mu}{2^k}\big]$;

(ii) $\Pr[\,a = g(x)\,] \ge 1-\eta$;

(iii) the pairs output on different executions are mutually independent.

An η*-noisy sampler* is an $(\eta,0)$-noisy sampler, i.e., one that on each execution selects a uniformly random $x \in \{-1,1\}^k$.

Chakraborty et al. [16] show how to build an efficient $O(\epsilon)$-noisy sampler for core_{f_k}, which is guaranteed to apply as long as $\operatorname{dist}(f,\mathcal{J}_k) = O\big(\epsilon^6/k^{10}\big)$. In more detail, they first run a modified version of the junta tester from [9], which, whenever it accepts, also returns some preprocessing information that enables one to build such a noisy sampler. Moreover, they show that this tester will indeed accept any function that is $O\big(\epsilon^6/k^{10}\big)$-close to $\mathcal{J}_k$ (in addition to rejecting those ϵ-far from it), giving the above guarantee. Using instead (a small modification of) our tolerant tester from Section 6, we are able to extend their techniques to obtain the following – less efficient, but more robust – noisy sampler.

Proposition 7.6 (Noisy sampler for close-to-junta functions)**.** *There are algorithms $\mathcal{A}_P, \mathcal{A}_S$ (respectively preprocessor and sampler), which both require oracle access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$, and satisfy the following properties.*

- *The preprocessor $\mathcal{A}_P$ takes $\epsilon' \in (0,1]$, $\rho \in (0,1)$, $k \in \mathbb{N}$ as inputs, makes $O\Big(\frac{k\log\frac{k}{\epsilon'}}{\epsilon'\rho(1-\rho)^k}\Big)$ queries to f, and either returns fail or a state $\sigma \in \{0,1\}^{\operatorname{poly}(n)}$. The sampler $\mathcal{A}_S$ takes as input such a state $\sigma \in \{0,1\}^{\operatorname{poly}(n)}$, makes a single query to f, and outputs a pair $(x,a) \in \{-1,1\}^k \times \{-1,1\}$. We say that a state σ is γ-*good *if for some permutation $\pi \in \mathcal{S}_k$, $\mathcal{A}_S(\sigma)$ is a γ-noisy sampler for $\mathsf{core}_{f_k} \circ \pi$.*
- *$\mathcal{A}_P(\epsilon',\rho,k)$ fulfills the following conditions:*
  - *(i) If $\operatorname{dist}(f,\mathcal{J}_k) \le \frac{\rho}{16}\epsilon'$, then with probability at least $4/5$, $\mathcal{A}_P$ returns a state σ that is $3\epsilon'$-good.*
  - *(ii) If $\operatorname{dist}(f,\mathcal{J}_k) > \epsilon'$, then with probability at least $4/5$, $\mathcal{A}_P$ returns fail.*
  - *(iii) If $\operatorname{dist}(f,\mathcal{J}_k) \le \epsilon'$, then with probability at least $4/5$, $\mathcal{A}_p$ either returns fail or returns a state σ that is $3\epsilon'$-good.*

The proof of Proposition 7.6 is deferred to Appendix A; indeed, it is almost identical to the proof of Proposition 4.16 in [16], with small adaptations required to comply with the use of the tolerant tester from Section 6 instead of the tester from [9].

We note that the main difference between the guarantees of our noisy sampler and those of the noisy sampler in [16, Lemma 2] lies in the set of functions for which the noisy sampler is required to return a good state. In our case, this set consists of functions that are *somewhat* close to k-juntas. In comparison, the construction from [16] is more query-efficient (only $\tilde{O}(k/\epsilon)$ queries to f in the preprocessing stage), but only guarantees the output of a noisy sampler for functions f that are $O\big(\epsilon^6/k^{10}\big)$-close to $\mathcal{J}_k$.

With these primitives in hand, we are almost ready to prove the main proposition of this subsection, Proposition 7.4. To state the algorithm (6) and proceed with its analysis, we will require the following definition:

Definition 7.7 (Number of violating pairs V_π)**.** Given two sets $Q_1, Q_2 \subseteq \{-1,1\}^k \times \{-1,1\}$ and a permutation $\pi \in \mathcal{S}_k$ we say that pairs $(x, a_1) \in Q_1$ and $(y, a_2) \in Q_2$ are *violating with respect to* π, if $y = \pi(x)$ and $a_1 \neq a_2$. We denote the number of violating pairs with respect to π by V_π.

Algorithm 6 Tolerant isomorphism testing to an unknown f such that $\operatorname{dist}(f, \mathcal{J}_k) \leq c\epsilon$ $(\mathcal{O}_f, \mathcal{O}_g, \epsilon, k)$

1: Let $\mathcal{A}_P, \mathcal{A}_S$ be as in Proposition 7.6, $\rho \leftarrow 1 - \frac{1}{\sqrt{2}}$, $\epsilon' \leftarrow \frac{\epsilon}{16}$, $\alpha \leftarrow 4c\epsilon$.
2: $s \leftarrow C\frac{2^{k/2}}{\epsilon}\sqrt{k \ln k}$, $t \leftarrow (3\alpha + 9\epsilon')\frac{s^2}{2^k}$. ▷ $C > 1$ is an absolute constant.
3: Run the preprocessor $\mathcal{A}_P$ on f and g with parameters ϵ', ρ, k.
4: **if** either invocation of $\mathcal{A}_P$ returned fail **then**
5:     **return** reject.
6: **end if**
7: Using the $3\epsilon'$-noisy sampler $\mathcal{A}_S$ (called with the states returned on Step 3), construct "core" sets $Q_f, Q_g \subseteq \{-1,1\}^k \times \{-1,1\}$ each of size $s \leftarrow C\frac{2^{k/2}}{\epsilon}\sqrt{k \ln k}$.
8: **if** there exist $\pi \in \mathcal{S}_k$ such that $V_\pi \leq t$ **then**
9:     **return** accept.
10: **end if**
11: **return** reject.

Proof of Proposition 7.4: The query complexity is the sum of the query complexities from Steps 3 and 7, i.e.,

$$O\left(\frac{k \log \frac{k}{\epsilon}}{\epsilon \rho (1-\rho)^k}\right) + 2s \cdot 1 = O\left(\frac{2^{k/2}}{\epsilon} k \log \frac{k}{\epsilon} + \frac{2^{k/2}}{\epsilon}\sqrt{k \ln k}\right) = O\left(\frac{2^{k/2}}{\epsilon} k \log \frac{k}{\epsilon}\right).$$

Completeness. Assume that g is $c\epsilon$-close to isomorphic to f, which itself is $c\epsilon$-close to being a k-junta. Therefore, by the triangle inequality and by our choice of $c \leq \frac{\rho}{512}$, $\operatorname{dist}(g, \mathcal{J}_k) \leq 2c\epsilon \leq \rho\epsilon'/16$ as well, so that with probability at least $3/5$ the algorithm does not output reject on Step 5 (we thereafter analyze this case). Moreover, by the triangle inequality there exists a permutation $\pi \in \mathcal{S}_n$ such that $\operatorname{dist}(f_k, g_k \circ \pi) \leq 2c\epsilon + 2c\epsilon = 4c\epsilon \stackrel{\text{def}}{=} \alpha$. In particular, this implies that there exists a permutation $\pi^* \in \mathcal{S}_k$ such that $\operatorname{dist}(\mathsf{core}_{f_k}, \mathsf{core}_{g_k} \circ \pi^*) \leq \alpha$. Let $T^* \subseteq \{-1,1\}^k$ be the disagreement set between core_{f_k} and $\mathsf{core}_{g_k} \circ \pi^*$: by the above $|T^*| \leq \alpha 2^k$.

Let $Q_f^s, Q_g^s \subseteq \{-1,1\}^k$ denote the sets resulting from taking the first element in each pair in Q_f and Q_g respectively. The size of the intersection $Z \stackrel{\text{def}}{=} \left|Q_f^s \cap T^*\right|$ is distributed as a Binomial random variable, namely $Z \sim \operatorname{Bin}\Big(s, \frac{|T^*|}{2^k}\Big)$, and conditioned on Z we have $Z^* \stackrel{\text{def}}{=} \left|Q_f^s \cap Q_g^s \cap T^*\right| \sim \operatorname{Bin}\big(s, \frac{Z}{2^k}\big)$. In particular, we get

$$\mathbf{E}[Z] = \frac{s\,|T^*|}{2^k}, \quad \mathbf{E}[\,Z^* \mid Z\,] = \mathbf{E}\big[\,\left|Q_f^s \cap Q_g^s \cap T^*\right| \mid \left|Q_f^s \cap T^*\right|\,\big] = \frac{sZ}{2^k}\ .$$

Let $\mathcal{A}_S^f$ denote the noisy sampler algorithm when invoked for f, and for every $x \in Q_f^s$ let $\mathcal{A}_S^f(x)$ denote the label given to x by $\mathcal{A}_S^f$. Since $\mathcal{A}_S^f$ is a $3\epsilon'$-noisy sampler for core_{f_k}, $\Pr[\mathcal{A}_S^f(x) \neq \mathsf{core}_{f_k}(x)] \leq 3\epsilon'$. An analogous statement holds for g. We let $N \stackrel{\text{def}}{=} |\{x \in Q_f^s \cap Q_g^s : \mathcal{A}_S^f(x) \neq \mathsf{core}_{f_k}(x) \text{ or } \mathcal{A}_S^g(x) \neq \mathsf{core}_{g_k}(x)\}|$ be the number of common samples incorrectly labelled by either noisy sampler, and observe that N is dominated by a Binomial random variable $\tilde{N} \sim \text{Bin}\left(\left|Q_f^s \cap Q_g^s\right|, 6\epsilon'\right)$.

With this in hand, we can bound $\Pr[V_{\pi^*} > t]$ as follows (recall that $t = 3\alpha + 9\epsilon'$):

$$\begin{aligned}\Pr\left[V_{\pi^*} > (3\alpha + 9\epsilon')\frac{s^2}{2^k}\right] &\leq \Pr\left[|Q_f^s \cap Q_g^s \cap T^*| > 3\alpha\frac{s^2}{2^k}\right] + \Pr\left[N > 9\epsilon'\frac{s^2}{2^k}\right] \\ &\leq \Pr\left[|Q_f^s \cap Q_g^s \cap T^*| > 3\alpha\frac{s^2}{2^k}\right] + \Pr\left[\tilde{N} > 9\epsilon'\frac{s^2}{2^k}\right].\end{aligned}$$

Recall that $Z^* = \left|Q_f^s \cap Q_g^s \cap T^*\right|$. Since $\Pr\left[\left|Q_f^s \cap Q_g^s \cap T^*\right| > 3\alpha\frac{s^2}{2^k}\right]$ is maximized when $|T^*|$ is maximal, we assume without loss of generality that $|T^*| = \alpha 2^k$. We will handle each term separately.

$$\begin{aligned}\Pr\left[Z^* > \frac{3}{2}\cdot\alpha\frac{s^2}{2^k}\right] &= \Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}}\right] \\ &= \Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}} \;\middle|\; Z > \frac{5}{4}\frac{s\,|T^*|}{2^k}\right]\cdot\Pr\left[Z > \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] \\ &\quad + \Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}} \;\middle|\; Z \leq \frac{5}{4}\frac{s\,|T^*|}{2^k}\right]\cdot\Pr\left[Z \leq \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] \\ &\leq \Pr\left[Z > \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] + \Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}} \;\middle|\; Z \leq \frac{5}{4}\frac{s\,|T^*|}{2^k}\right].\end{aligned}$$

We again bound the two terms separately. By the assumption that $|T^*| = \alpha 2^k$ and by the choice of s,

$$\Pr\left[Z > \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] = \Pr\left[Z > \frac{5}{4}\mathbf{E}[Z]\right] < \exp\left(-\frac{1}{3}\cdot\left(\frac{1}{4}\right)^2\cdot\frac{s\,|T^*|}{2^k}\right) < \frac{1}{30}.$$

As for the second term, since $\mathbf{E}[Z^*] = \frac{sZ}{2^k}$ and by the assumption on T^* and the setting of s,

$$\begin{aligned}\Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}} \;\middle|\; Z < \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] &\leq \Pr\left[Z^* > \frac{3}{2}\frac{s^2\,|T^*|}{2^{2k}} \;\middle|\; Z = \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] \\ &= \Pr\left[Z^* > \frac{6}{5}\mathbf{E}[Z^*] \;\middle|\; Z = \frac{5}{4}\frac{s\,|T^*|}{2^k}\right] \\ &< \exp\left(-\frac{1}{3}\cdot\left(\frac{1}{5}\right)^2\cdot\frac{s}{2^k}\frac{s\,|T^*|}{2^k}\right) < \frac{1}{30}\end{aligned}$$

for a sufficiently large constant C in the definition of s.

As for the last term of the initial expression, since $\mathbf{E}\left[\tilde{N}\right] = 6\epsilon'\left|Q_f^s \cap Q_g^s\right|$ we have,

$$\begin{aligned}\Pr\left[\tilde{N} > 9\epsilon'\frac{s^2}{2^k}\right] &\leq \Pr\left[\tilde{N} > 9\epsilon'\frac{s^2}{2^k} \;\middle|\; |Q_f^s \cap Q_g^s| \leq \frac{5}{4}\frac{s^2}{2^k}\right]\cdot\Pr\left[|Q_f^s \cap Q_g^s| \leq \frac{5}{4}\frac{s^2}{2^k}\right] \\ &\quad + \Pr\left[|Q_f^s \cap Q_g^s| > \frac{5}{4}\frac{s^2}{2^k}\right] \\ &\leq \Pr\left[\tilde{N} > 9\epsilon'\frac{s^2}{2^k} \;\middle|\; |Q_f^s \cap Q_g^s| = \frac{5}{4}\frac{s^2}{2^k}\right] + \Pr\left[|Q_f^s \cap Q_g^s| > \frac{5}{4}\frac{s^2}{2^k}\right] \\ &\leq \Pr\left[\tilde{N} > \frac{6}{5}\mathbf{E}\left[\tilde{N}\right] \;\middle|\; |Q_f^s \cap Q_g^s| = \frac{5}{4}\frac{s^2}{2^k}\right] + \Pr\left[|Q_f^s \cap Q_g^s| > \frac{5}{4}\frac{s^2}{2^k}\right] \\ &< \exp\left(-\frac{1}{3}\cdot\left(\frac{1}{6}\right)^2\cdot\frac{16\epsilon'\cdot 5s^2}{4\cdot 2^k}\right) + \exp\left(-\frac{1}{3}\cdot\left(\frac{1}{4}\right)^2\cdot\frac{s^2}{2^k}\right) \leq \frac{1}{15}. \qquad \text{(Actually } o(1).\text{)}\end{aligned}$$

The algorithm will therefore reject with probability at most $\frac{2}{5}+\frac{1}{15}+\frac{1}{15}=\frac{7}{15}$.

Soundness. Assume that $\operatorname{dist}(f,\mathcal{J}_k)\leq c\epsilon$, and that g is ϵ-far from being isomorphic to f. Then one of the following must hold:

1. $\operatorname{dist}(g,\mathcal{J}_k)>\epsilon'$.
2. for all $\pi\in\mathcal{S}_k$, $\operatorname{dist}(\mathsf{core}_{f_k},\mathsf{core}_{g_k}\circ\pi)>\epsilon-(\epsilon'+c\epsilon)>\epsilon-2\epsilon'$.

If the first case holds, then the function will be rejected in Step 3 with probability at least $\frac{4}{5}$, and so the algorithm will reject as desired. We can therefore focus on the second case.

If the second case holds, either the tester rejects in Step 5 (and we are done) or it outputs a state which will be used to get the $3\epsilon'$-noisy sampler. Fix any $\pi\in\mathcal{S}_k$. Since $\operatorname{dist}(\mathsf{core}_{f_k},\mathsf{core}_{g_k}\circ\pi)>(\epsilon-2\epsilon')$, there are $m\stackrel{\text{def}}{=}m(\pi)\geq(\epsilon-2\epsilon')2^k$ inputs $x\in\{-1,1\}^k$ such that $\mathsf{core}_{f_k}(x)\neq\mathsf{core}_{g_k}\circ\pi(x)$. Let $T=T(\pi)\subseteq\{-1,1\}^k$ denote the set of all such inputs (so that $|T|=m$).

We can make a similar argument as for the completeness case: we have that $\left|Q_f^s\cap T\right|$ is a random variable with Binomial distribution (of parameters s, and $\frac{|T|}{2^k}$). Conditioned on $\left|Q_f^s\cap T\right|$, we also have $\left|Q_f^s\cap Q_g^s\cap T\right|\sim\operatorname{Bin}\left(s,\frac{\left|Q_f^s\cap T\right|}{2^k}\right)$, so that

$$\begin{aligned}\mathbf{E}\big[\left|Q_f^s\cap Q_g^s\cap T\right|\big]&=\mathbf{E}\big[\mathbf{E}\big[\left|Q_f^s\cap Q_g^s\cap T\right|\;\big|\;\left|Q_f^s\cap T\right|\big]\big]\\&=\mathbf{E}\left[\frac{s\left|Q_f^s\cap T\right|}{2^k}\right]=\frac{s^2\,|T|}{2^{2k}}\geq(\epsilon-2\epsilon')\frac{s^2}{2^k}=14\epsilon'\frac{s^2}{2^k}.\end{aligned}$$

(Recall that our threshold was set to $t=(3\alpha+9\epsilon')\frac{s^2}{2^k}\leq 12\epsilon'\frac{s^2}{2^k}$.) Moreover, each element $x\in Q_f^s\cap Q_g^s\cap T$ will contribute to V_π with probability at least $(1-3\epsilon')^2>\frac{24}{25}$ (since this is a lower bound on the probability that both $\mathcal{A}_S^f(x)=\mathsf{core}_{f_k}(x)$ and $\mathcal{A}_S^g(x)=\mathsf{core}_{g_k}(x)$, and as $\epsilon'\leq\frac{\epsilon_0}{16}$). As before, we can therefore write, letting $Z\stackrel{\text{def}}{=}\left|Q_f^s\cap Q_g^s\cap T\right|$, and taking $|T|$ to be minimal so that $|T|=(\epsilon-2\epsilon')2^k$,

$$\begin{aligned}\Pr[\,V_\pi>t\,]&\geq\Pr\left[V_\pi>t\;\middle|\;Z\geq\frac{13}{12}t\right]\Pr\left[Z\geq\frac{13}{12}t\right]\\&\geq(1-e^{-\frac{1}{3}\left(\frac{1}{26}\right)^2\cdot\frac{24}{25}\cdot\frac{13t}{12}})\Pr\left[Z\geq\frac{13}{12}t\right]=\left(1-e^{-\frac{t}{2000}}\right)\Pr\left[Z\geq\frac{13}{12}t\right]\qquad\text{(Chernoff bound)}\end{aligned}$$

so that it is sufficient to lower bound $\Pr\big[\,Z\geq\frac{13}{12}t\,\big]$. To do so, we will bound the probability of the two following events:

E1: $Y\stackrel{\text{def}}{=}\left|Q_f^s\cap T\right|<\frac{99}{100}\frac{s|T|}{2^k}$

E2: $Z=\left|Q_f^s\cap Q_g^s\cap T\right|<\frac{99}{100}\frac{s}{2^k}\left|Q_f^s\cap T\right|$, conditioning on $\left|Q_f^s\cap T\right|\geq\frac{99}{100}\frac{s|T|}{2^k}$.

This will be sufficient for us to conclude, as by our choice of $t=(3\alpha+9\epsilon')\frac{s^2}{2^k}$,the setting $|T|=(\epsilon-2\epsilon')2^k=\frac{7}{8}\epsilon 2^k$, and since $\alpha\leq\epsilon'$, we have

$$\frac{13}{12}\cdot t\leq\frac{13}{12}\cdot 12\epsilon'\cdot\frac{s^2}{2^k}=\frac{13}{16}\cdot\epsilon\cdot\frac{s^2}{2^k}\leq\left(\frac{99}{100}\right)^2\frac{s^2|T|}{2^{2k}}\;.$$

Therefore, by a Chernoff bound

$$\begin{aligned}
\Pr\left[\, Z < \frac{13}{12}t \,\right] &\leq \Pr\left[\, Z < \left(\frac{99}{100}\right)^2 \frac{s^2\,|T|}{2^{2k}} \,\right] \\
&\leq \Pr\left[\, Y < \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] + \Pr\left[\, Z < \frac{99}{100}\frac{s}{2^k}Y \;\middle|\; Y \geq \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] \Pr\left[\, Y \geq \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] \\
&\leq \Pr\left[\, Y < \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] + \Pr\left[\, Z < \frac{99}{100}\frac{s}{2^k}Y \;\middle|\; Y \geq \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] \\
&< \exp\left(-\frac{1}{2}\cdot\left(\frac{1}{100}\right)^2 \cdot \frac{s\,|T|}{2^k}\right) + \Pr\left[\, Z < \frac{99}{100}\frac{sY}{2^k} \;\middle|\; Y = \frac{99}{100}\frac{s\,|T|}{2^k} \,\right] \\
&\leq \exp\left(-\frac{1}{2}\cdot\left(\frac{1}{100}\right)^2 \cdot \frac{s(\epsilon-2\epsilon')\cdot 2^k}{2^k}\right) + \exp\left(-\frac{1}{2}\cdot\left(\frac{2}{100}\right)^2 \cdot \frac{s^2(\epsilon-2\epsilon')\cdot 2^k}{2^{2k}}\right) \\
&< \exp(-\tau C^2 k \ln k),
\end{aligned}$$

by the choice $s = C\frac{2^{k/2}}{\epsilon}\sqrt{k\ln k}$, and for some constant $\tau \in (0,1)$. Hence setting C to a sufficiently large constant, the foregoing analysis implies that $\Pr[\, V_\pi \leq t \,] \leq e^{-\frac{t}{2000}} + e^{-\tau C^2 k \ln k} = e^{-\frac{12c+3/4}{2000}\epsilon s^2/2^k} + e^{-\tau C^2 k\ln k} \leq \frac{7}{15k^k}$. A union bound over all $k! < k^k$ permutations $\pi \in \mathcal{S}_k$ finally yields $\Pr[\, \exists \pi,\, V_\pi \leq t \,] \leq \frac{7}{15}$ as claimed. ∎

References

[1] Nir Ailon, Bernard Chazelle, Seshadhri Comandur, and Ding Liu. Estimating the distance to a monotone function. *Random Structures and Algorithms*, 31(3):371–383, 2007.

[2] Noga Alon and Eric Blais. Testing Boolean function isomorphism. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *APPROX–RANDOM 2010*, volume 6302 of *Lecture Notes in Computer Science*, pages 394–405. Springer, 2010.

[3] Noga Alon, Eric Blais, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Nearly tight bounds for testing function isomorphism. *SIAM Journal on Computing*, 42(2):459–493, 2013.

[4] Francis R. Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends in Machine Learning*, 6(2-3):145–373, 2013.

[5] Zsolt Baranyai. On the factorization of the complete uniform hypergraph. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday)*, volume 1, pages 91–108, 1975.

[6] Alexandre Belloni, Tengyuan Liang, Hariharan Narayanan, and Alexander Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 240–265, Paris, France, 03–06 Jul 2015. PMLR.

[7] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Tolerant testers of image properties. In *Proceedings of ICALP*, volume 55 of *LIPIcs*, pages 90:1–90:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[8] Eric Blais. Improved bounds for testing juntas. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 317–330. Springer, 2008.

[9] Eric Blais. Testing juntas nearly optimally. In *Proceedings of STOC*, pages 151–158. ACM, 2009.

[10] Eric Blais. *Testing properties of Boolean functions.* PhD thesis, CMU, 2012.

[11] Avrim L. Blum. Relevant examples and relevant features: Thoughts from computational learning theory. In *AAAI Fall Symposium on 'Relevance*, volume 5, 1994.

[12] Avrim L. Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1):245–271, 1997.

[13] Madeline V. Brandt. Intersecting hypergraphs and decompositions of complete uniform hypergraphs. B.A. Thesis, Reed College, 2015.

[14] Andrea Campagna, Alan Guo, and Ronitt Rubinfeld. Local reconstructors and tolerant testers for connectivity and diameter. In *APPROX-RANDOM*, volume 8096 of *Lecture Notes in Computer Science*, pages 411–424. Springer, 2013.

[15] Sourav Chakraborty, Eldar Fischer, David García-Soriano, and Arie Matsliah. Junto-symmetric functions, hypergraph isomorphism and crunching. In *Conference on Computational Complexity*, pages 148–158. IEEE, 2012.

[16] Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Efficient sample extractors for juntas with applications. volume 6755 of *Lecture Notes in Computer Science*, pages 545–556. Springer, 2011.

[17] Hana Chockler and Dan Gutfreund. A Lower Bound for Testing Juntas. *Information Processing Letters*, 90(6):301–305, 2004.

[18] Ilias Diakonikolas, Homin K Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A Servedio, and Andrew Wan. Testing for concise representations. In *Proceedings of FOCS*, pages 549–558. IEEE, 2007.

[19] Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Transactions on Algorithms (TALG)*, 6(3):52, 2010.

[20] Eldar Fischer and Lance Fortnow. Tolerant Versus Intolerant Testing for Boolean Properties. *Theory of Computing*, 2(9):173–183, 2006.

[21] Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004.

[22] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM Journal on Computing*, 37(2):482–501, 2007.

[23] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[24] Venkatesan Guruswami and Atri Rudra. Tolerant locally testable codes. In *APPROX-RANDOM*, volume 3624 of *Lecture Notes in Computer Science*, pages 306–317. Springer, 2005.

[25] Z. Q. John Lu. The elements of statistical learning: data mining, inference, and prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3):693–694, 2010.

[26] Michael Kearns and Dana Ron. Testing problems with sub-learning sample complexity. In *Proceedings of COLT*, pages 268–279. ACM, 1998.

[27] Swastik Kopparty and Shubhangi Saraf. Tolerant linearity testing and locally testable codes. In *APPROX-RANDOM*, volume 5687 of *Lecture Notes in Computer Science*, pages 601–614. Springer, 2009.

[28] Pravesh Kothari, Amir Nayyeri, Ryan O'Donnell, and Chenggang Wu. Testing surface area. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1204–1214. Society for Industrial and Applied Mathematics, 2014.

[29] Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Proceedings of FOCS*, pages 1049–1065. IEEE, 2015.

[30] Amit Levi and Erik Waingarten. Lower bounds for tolerant junta and unateness testing via rejection sampling of graphs. *CoRR*, abs/1805.01074, 2018. To appear in ITCS'19.

[31] Sharon Marko and Dana Ron. Approximating the distance to properties in bounded-degree and general sparse graphs. *ACM Trans. Algorithms*, 5(2), 2009.

[32] Elchanan Mossel, Ryan O'Donnell, and Rocco P Servedio. Learning juntas. In *Proceedings of STOC*, pages 206–212. ACM, 2003.

[33] Michal Parnas and Dana Ron. Testing the diameter of graphs. *Random Structures & Algorithms*, 20(2):165–183, 2002.

[34] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006.

[35] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2002.

[36] Rocco A. Servedio, Li-Yang Tan, and John Wright. Adaptivity helps for testing juntas. In *Conference on Computational Complexity*, volume 33 of *LIPIcs*, pages 264–279. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

[37] Zoya Svitkina and Lisa Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. *SIAM Journal on Computing*, 40(6):1715–1737, 2011.

[38] Roei Tell. A note on tolerant testing with one-sided error. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 32, 2016.

[39] Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and the closest pair problem. *Journal of the ACM*, 62(2):13, 2015.

A Deferred proof: construction of a noisy sampler

We provide in this appendix the proof of Proposition 7.6, restated below:

Proposition 7.6 (Noisy sampler for close-to-junta functions)**.** *There are algorithms $\mathcal{A}_P, \mathcal{A}_S$ (respectively preprocessor and sampler), which both require oracle access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$, and satisfy the following properties.*

- *The preprocessor $\mathcal{A}_P$ takes $\epsilon' \in (0,1]$, $\rho \in (0,1)$, $k \in \mathbb{N}$ as inputs, makes $O\Big(\frac{k \log \frac{k}{\epsilon'}}{\epsilon' \rho (1-\rho)^k}\Big)$ queries to f, and either returns* fail *or a state $\sigma \in \{0,1\}^{\mathrm{poly}(n)}$. The sampler $\mathcal{A}_S$ takes as input such a state $\sigma \in \{0,1\}^{\mathrm{poly}(n)}$, makes a single query to f, and outputs a pair $(x,a) \in \{-1,1\}^k \times \{-1,1\}$. We say that a state σ is* γ-good *if for some permutation $\pi \in \mathcal{S}_k$, $\mathcal{A}_S(\sigma)$ is a γ-noisy sampler for $\mathsf{core}_{f_k} \circ \pi$.*
- *$\mathcal{A}_P(\epsilon', \rho, k)$ fulfills the following conditions:*
  - *(i) If $\mathrm{dist}(f, \mathcal{J}_k) \leq \frac{\rho}{16}\epsilon'$, then with probability at least $4/5$, $\mathcal{A}_P$ returns a state σ that is $3\epsilon'$-good.*
  - *(ii) If $\mathrm{dist}(f, \mathcal{J}_k) > \epsilon'$, then with probability at least $4/5$, $\mathcal{A}_P$ returns* fail*.*
  - *(iii) If $\mathrm{dist}(f, \mathcal{J}_k) \leq \epsilon'$, then with probability at least $4/5$, $\mathcal{A}_p$ either returns* fail *or returns a state σ that is $3\epsilon'$-good.*

We will very closely follow the argument from the full version of [16] (Proposition 4.16),[9] adapting the corresponding parts in order to obtain our result. For completeness, we tried to make this appendix below self-contained, reproducing almost verbatim several parts of the proof from [16].[10]

Proof of Proposition 7.6: In order to use our result from Section 6 *in lieu* of the junta tester from [9], we first need to make a small modification to our algorithm. Specifically, in its first step our tester will now pick a random partition $\mathcal{I}$ of $[n]$ in $\ell \stackrel{\text{def}}{=} \frac{Ck^2}{\epsilon}$ parts instead of $24k^2$ (for some (small) absolute constant $C > 1$). It is easy to check that both Lemma 3.2 and Lemma 3.3 still hold (e.g., from the proof of [10, Lemma 5.4]), now with probability at least 19/20. Moreover, our modified tolerant tester offers the same soundness and completeness guarantees as Theorem 1.2, at the price of a query complexity $O\Big(\frac{k\log(k/\epsilon)}{\epsilon\rho(1-\rho)^k}\Big)$ (instead of $O\Big(\frac{k \log k}{\epsilon \rho (1-\rho)^k}\Big)$). Moreover, in Step 4 of 4, i.e. when the algorithm found a suitable set $J \subseteq [\ell]$ (of size $\ell - k$) as a witness for accepting, we make the algorithm return $\mathcal{I}$ and the set $\mathcal{J} \stackrel{\text{def}}{=} \{I_j\}_{j \in \bar{J}}$ along with the verdict accept.

We will also require the definitions of the distribution induced by a partition $\mathcal{I}$ and a subset $\mathcal{J} \subseteq \mathcal{I}$, and of such a couple $(\mathcal{I}, \mathcal{J})$ being *good* for a function:

Definition A.1 ([16, Definition 4.6])**.** For any partition $\mathcal{I} = \{I_1, \dots, I_\ell\}$ of $[n]$, and subset of parts $\mathcal{J} \subseteq \mathcal{I}$, we define a pair of distributions:

The distribution $D_{\mathcal{I}}$ on $\{-1,1\}^n$. An element $y \sim D_{\mathcal{I}}$ is sampled by

1. picking $z \in \{-1,1\}^\ell$ uniformly at random among all $\binom{\ell}{\ell/2}$ strings of weight $\frac{\ell}{2}$;
2. setting $y_i = z_j$ for all $j \in [\ell]$ and $i \in I_j$.

The distribution $D_{\mathcal{J}}$ on $\{-1,1\}^{|\mathcal{J}|}$. An element $x \sim D_{\mathcal{J}}$ is sampled by

1. picking $y \sim D_{\mathcal{I}}$;
2. outputting $\mathsf{extract}_{(\mathcal{I},\mathcal{J})}(y)$, where $x = \mathsf{extract}_{(\mathcal{I},\mathcal{J})}(y)$ is defined as follows. For all $j \in [\ell]$ such that $I_j \in \mathcal{J}$:
   - if $I_j \neq \emptyset$, set $x_j = y_i$ (where $i \in I_j$);

[9] The full version can be found at `http://www.cs.technion.ac.il/~ariem/eseja.pdf`.

[10] The reader may notice that Chakraborty et al. rely on a definition of set-influence that differs from ours by a factor 2; we propagated the changes through the argument.

- if $I_j = \emptyset$, set x_j to be a uniformly random bit.

Lemma A.2 ([16, Lemma 4.7]). *$D_{\mathcal{I}}$ and $D_{\mathcal{J}}$ as above satisfy the following.*

- *For all $a \in \{-1,1\}^n$, $\Pr_{\mathcal{I}, y \sim D_{\mathcal{I}}}[\, y = a\,] = \frac{1}{2^n}$.*
- *Assume $\ell > 4\,|\mathcal{J}|^2$. For every $\mathcal{I}$ and $\mathcal{J} \subseteq \mathcal{I}$, the total variation distance between $D_{\mathcal{J}}$ and the uniform distribution on $\{-1,1\}^{|\mathcal{J}|}$ is bounded by $2\,|\mathcal{J}|^2/\ell$. Moreover, the ℓ_∞ distance between the two distributions is at most $4\,|\mathcal{J}|^2/(\ell 2^{|\mathcal{J}|})$.*

Definition A.3 ([16, Definition 4.8]). Given $(\mathcal{I}, \mathcal{J})$ as above and oracle access to $f\colon \{-1,1\}^n \to \{-1,1\}$, we define a probabilistic algorithm $\mathsf{sampler}_{(\mathcal{I},\mathcal{J})}(f)$ that on each execution produces a pair $\langle x, a\rangle \in \{-1,1\}^{|\mathcal{J}|} \times \{-1,1\}$ as follows: first it picks a random $y \sim D_{\mathcal{I}}$, then it queries f on y, computes $x = \mathsf{extract}_{(\mathcal{I},\mathcal{J})}(y)$ and outputs the pair $\langle x, f(y)\rangle$.

Definition A.4 ([16, Definition 4.9]). Given $\alpha > 0$, a function $f\colon \{-1,1\}^n \to \{-1,1\}$, a partition $\mathcal{I} = \{I_1, \dots, I_\ell\}$ of $[n]$ and a subset $\mathcal{J} \subseteq \mathcal{I}$ of k parts, we call the pair $(\mathcal{I}, \mathcal{J})$ *α-good* (with respect to f) if there exists a k-junta $h \in \mathcal{J}_k$ such that the following conditions are satisfied:

1. Conditions on h:
   (a) Every relevant variable of h is also a relevant variable of f_k;
   (b) $\operatorname{dist}(h, f_k) \le \alpha$.
2. Conditions on $\mathcal{I}$:
   (a) For all $j \in [\ell]$, I_j contains at most one variable of core_{f_k};
   (b) $\Pr_{y \sim D_{\mathcal{I}}}[\, f(y) \neq f_k(y)\,] \le 10 \cdot \operatorname{dist}(f, f_k)$.
3. Condition on $\mathcal{J}$: the set $S \stackrel{\text{def}}{=} \bigcup_{I \in \mathcal{J}} I$ contains all relevant variables of h.

Lemma A.5 ([16, Lemma 4.10]). *Let $\alpha, f, \mathcal{I}, \mathcal{J}$ be as in the preceding definition. If the pair $(\mathcal{I}, \mathcal{J})$ is α-good (with respect to f), then $\mathsf{sampler}_{(\mathcal{I},\mathcal{J})}(f)$ (as per Definition A.3) is an (η, μ)-noisy sampler for some permutation of core_{f_k}, with $\eta \le 2\alpha + \frac{4k^2}{\ell} + 10 \cdot \operatorname{dist}(f, f_k)$ and $\mu \le \frac{4k^2}{\ell}$.*

The last piece we shall need is the ability to convert an (η, μ)-noisy sampler to a $(\eta', 0)$-noisy sampler – that is, one whose samples are exactly uniformly distributed.

Lemma A.6 ([16, Lemma 4.4]). *Let $\tilde{g}$ be an (η, μ)-noisy sampler for $g\colon \{-1,1\}^k \to \{-1,1\}$, that on each execution picks x according to some fixed (and fully known) distribution D. Then $\tilde{g}$ can be turned into an $(\eta + \mu)$-noisy sampler $\tilde{g}_{\text{unif}}$ for g.*

With this in hand, we are ready to prove the main lemma:

Lemma A.7 (Analogue of [16, Proposition 4.16]). *The tester from Theorem 1.2, modified as above, has the following guarantees. It has query complexity $O\Big(\frac{k \log(k/\epsilon)}{\epsilon\rho(1-\rho)^k}\Big)$ and outputs, in case of acceptance, a partition $\mathcal{I}$ of $[n]$ in $\ell \stackrel{\text{def}}{=} O\big(k^2/\epsilon\big)$ parts along with a subset $\mathcal{J} \subseteq \mathcal{I}$ of k parts such that for any f the following conditions hold:*

- *if $\operatorname{dist}(f, \mathcal{J}_k) \le \frac{\rho}{16}\epsilon$, the algorithm accepts with probability at least $9/10$;*
- *if $\operatorname{dist}(f, \mathcal{J}_k) > \epsilon$, the algorithm rejects with probability at least $9/10$;*
- *for any f, with probability at least $4/5$ either the algorithm rejects, or it outputs $\mathcal{J}$ such that the pair $(\mathcal{I}, \mathcal{J})$ is $\frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon$-good (as per Definition A.4).*

In particular, if $\operatorname{dist}(f, \mathcal{J}_k) \le \frac{\rho}{16}\epsilon$, then with probability at least $4/5$ the algorithm outputs a set $\mathcal{J}$ such that $(\mathcal{I}, \mathcal{J})$ is $\frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon$-good.

Proof of Lemma A.7: The first two items follow from the analysis of the tester (Theorem 1.2) and the foregoing discussion; we thus turn to establishing the third item.

Called with parameters k, ρ, ϵ, our algorithm, with probability at least 19/20, either rejects or outputs a partition $\mathcal{I}$ of $[n]$ into $\ell = O(k^2)$ parts and set $\mathcal{J} \subseteq \mathcal{I}$ satisfying $\mathbf{Inf}_f(\overline{\phi(J)}) \leq \epsilon$. Let $R \subseteq [n]$ (with $|R| \leq k$) denote the set of relevant variables of f_k, and $V \supseteq R$ (with $|V| = k$) the set of relevant variables of core_{f_k}. Assume that $\operatorname{dist}(f, \mathcal{J}_k) \leq \frac{\rho\epsilon}{16}$.[11] We then have:

- by the above, with probability at least 19/20 the algorithm outputs a set $\mathcal{J} \subseteq \mathcal{I}$ which satisfies
$$\mathbf{Inf}_f(\overline{\phi(J)}) \leq \epsilon;$$
- since $\ell \gg k^2$, with probability at least 19/20 all elements of V fall in different parts of the partition $\mathcal{I}$;
- by Lemma A.2 and by Markov's inequality, with probability at least 9/10 the partition $\mathcal{I}$ satisfies $\Pr_{y\sim\mathcal{D}_\mathcal{I}}[\, f(y) \neq f_k(y)\,] \leq 10 \cdot \operatorname{dist}(f, f_k)$.

So by a union bound, with probability at least 4/5 all three of these events occur. Now we show that conditioned on them, the pair $(\mathcal{I}, \mathcal{J})$ is $(1 + \frac{3}{2}\rho)\epsilon$-good. Let $U \stackrel{\text{def}}{=} R \cap \left(\bigcup_{I\in\mathcal{J}} I\right)$ (informally, U is the subset of the relevant variables of f_k that were successfully "discovered" by the tester). Since $\operatorname{dist}(f, \mathcal{J}_k) \leq \frac{\rho\epsilon}{16}$, we have $\mathbf{Inf}_f(\bar{V}) \leq 4 \operatorname{dist}(f, \mathcal{J}_k) \leq \frac{\rho\epsilon}{4}$. By the subadditivity and monotonicity of influence we get

$$\mathbf{Inf}_f(\bar{U}) \leq \mathbf{Inf}_f(\bar{V}) + \mathbf{Inf}_f(V \setminus U) \leq \mathbf{Inf}_f(\bar{V}) + \mathbf{Inf}_f(\overline{\phi(J)})) \leq \frac{\rho\epsilon}{4} + \epsilon.$$

where the second inequality follows from $V \setminus U \subseteq \overline{\phi(J)}$. This means (see e.g. [10, Lemma 2.21]) that there is a k-junta h on U such that $\operatorname{dist}(f, h) \leq \frac{1}{2}(\frac{\rho\epsilon}{4} + \epsilon)$, and by the triangle inequality $\operatorname{dist}(f_k, h) \leq \frac{1}{2}(\frac{\rho\epsilon}{4} + \epsilon) + \frac{\rho\epsilon}{16} = \frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon$. Based on this h, we can verify that the pair $(\mathcal{I}, \mathcal{J})$ is $\frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon$-good by going over the conditions in Definition A.4. ■

Concluding the proof of Proposition 7.6. We conclude as in Section 4.6 of [16], and start by describing how $\mathcal{A}_P$ and $\mathcal{A}_S$ operate. The preprocessor $\mathcal{A}_P$ starts by calling the tester $\mathcal{T}$ of Lemma A.7. Then, in case $\mathcal{T}$ accepted, $\mathcal{A}_P$ encodes in the state σ the partition $\mathcal{I}$ and the subset $\mathcal{J} \subseteq \mathcal{I}$ output by $\mathcal{T}$ (see Lemma A.7), along with the values of k and ϵ. The sampler $\mathcal{A}_S$, given σ, obtains a pair $\langle x, a\rangle \in \{-1, 1\}^k \times \{-1, 1\}$ by executing $\mathsf{sampler}_{(\mathcal{I},\mathcal{J})}(f)$ (from Definition A.3) once. Now we show how Proposition 7.6 follows from Lemma A.7. The first two items are immediate. As for the third item, notice that we only have to analyze the case where $\operatorname{dist}(f, f_k) \leq \frac{\rho\epsilon}{16}$ and $\mathcal{T}$ accepted; all other cases are taken care of by the first two items. By the third item in Lemma A.7, with probability at least 4/5 the pair $(\mathcal{I}, \mathcal{J})$ is $\frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon$-good. If so, by Lemma A.5, $\mathsf{sampler}_{(\mathcal{I},\mathcal{J})}(f)$ is an (η, μ)-noisy sampler for some permutation of core_{f_k}, where

$$\eta \leq 2 \cdot \frac{1}{2}(1 + \frac{3}{8}\rho)\epsilon + \frac{4k^2}{\ell} + 10 \cdot \operatorname{dist}(f, \mathcal{J}_k) \leq (1 + \frac{3}{8}\rho)\epsilon + \frac{10\rho\epsilon}{16} + \frac{4k^2}{\ell} = (1 + \rho)\epsilon + \frac{4k^2}{\ell}$$

and $\mu \leq \frac{4k^2}{\ell}$. This in turn implies by Lemma A.6 an η'-noisy sampler, for

$$\eta' = \eta + \mu \leq (1 + \rho)\epsilon + \frac{8k^2}{\ell} \leq (2 + \rho)\epsilon \leq 3\epsilon$$

as claimed. (Where we used that $\frac{8k^2}{\ell} \leq \epsilon$ by our choice of ℓ.) ■

B Deferred proof: legal collection of covers

We provide in this appendix the proof of Claim 6.2, restated below:

Definition B.1. Let X be a set of j elements, and for any $s \in [j]$ consider the family $\binom{X}{s}$ of all subsets of X that have size s. We shall say that $\mathcal{C} \subseteq \binom{X}{s}$ is a *s-cover* of X, if $\bigcup_{Y\in\mathcal{C}} Y = X$. We shall say that $\mathcal{C}_1, \dots, \mathcal{C}_m$ is a *legal collection of s-covers for* X, if each $\mathcal{C}_t$ is a cover of X, and these covers are disjoint.

[11] For other f's, the third item follows from the second item.

Claim B.2. *For any set X of j elements, there exists a legal collection of s-covers for X of size at least*

$$m \geq \left\lfloor \frac{\binom{j}{s}}{\left\lceil \frac{j}{s} \right\rceil} \right\rfloor .$$

(Moreover, this bound is tight.)

Proof: This claim follows from a result due to Baranyai [5] on factorization of regular hypergraphs. We state this result, and describe how to derive the claim from it, below (recall that K_n^h denotes the h-regular hypergraph K_n^h on n vertices):[12]

Theorem B.3 (Baranyai's Theorem [5, Theorem 1])**.** *Let n, h be integers satisfying $1 \leq h \leq n$, and $a_1, \dots, a_\ell$ integers such that $\sum_{i=1}^{\ell} a_i = \binom{n}{h}$. Then the edges of K_n^h can be partitioned into hypergraphs $\mathcal{H}_1 \dots, \mathcal{H}_\ell$ such that*

(i) $|\mathcal{H}_i| = a_i$ for all $i \in [\ell]$;

(ii) each $\mathcal{H}_i$ is almost regular*: the number of hyperedges any two vertices $u, v \in \mathcal{H}_i$ participate in differs by at most one (and here, specifically, is either $\left\lceil \frac{a_i h}{n} \right\rceil$ or $\left\lfloor \frac{a_i h}{n} \right\rfloor$).*

We apply Theorem B.3 as follows: setting $m \stackrel{\text{def}}{=} \left\lfloor \frac{\binom{j}{s}}{\left\lceil \frac{j}{s} \right\rceil} \right\rfloor \leq \binom{j-1}{s-1}$ and $\ell \stackrel{\text{def}}{=} m+1$, we let $a_i \stackrel{\text{def}}{=} \left\lceil \frac{j}{s} \right\rceil$ for all $1 \leq i \leq m$, and $a_\ell \stackrel{\text{def}}{=} \binom{j}{s} - \sum_{i=1}^{m} a_i \geq 0$. By the theorem, we obtain a partition of K_j^s into $\ell = m+1$ hypergraphs $\mathcal{H}_1 \dots, \mathcal{H}_\ell$ such that the first m satisfy:

(i) $|\mathcal{H}_i| = \left\lceil \frac{j}{s} \right\rceil$ for all $i \in [m]$;

(ii) for any $i \in [m]$, any vertex $u \in \mathcal{H}_i$ participates in either 1 or 2 hyperedges;

(and we cannot say much about the "remainder" hypergraph $\mathcal{H}_\ell$). Condition (ii) ensures that each of the first m hypergraphs obtained indeed defines a cover of the set of j elements by s-element subsets, while by definition of the partition of the hypergraph we are promised that these m s-covers are disjoint. This proves the lemma, as $\mathcal{H}_1 \dots, \mathcal{H}_m$ then induce a legal cover of X.

As for the optimality of the bound, it follows readily from observing that one must have $m \leq \left\lfloor \frac{\binom{j}{s}}{\lceil j/s \rceil} \right\rfloor$ since for every cover $\mathcal{C}$ we must have $|\mathcal{C}| \geq \lceil j/s \rceil$, and $\left| \binom{X}{s} \right| = \binom{j}{s}$. ■

[12] An exposition of this result and the original proof as given by Baranyai can also be found in [13, Theorem 4.1.1].