Proceedings of Machine Learning Research vol 195:1–17, 2023 36th Annual Conference on Learning Theory

Complexity of High-Dimensional Identity Testing with Coordinate Conditional Sampling

Antonio Blanca ABLANCA@CSE.PSU.EDU
Department of Computer Science and Engineering, Pennsylvania State University

Zongchen Chen ZONGCHEN@MIT.EDU
Department of Mathematics, Massachusetts Institute of Technology

Daniel Štefankovič STEFANKO@CS.ROCHESTER.EDU
Department of Computer Science, University of Rochester

Eric Vigoda VIGODA@UCSB.EDU
Department of Computer Science, University of California, Santa Barbara

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We study the identity testing problem for high-dimensional distributions. Given as input an explicit distribution μ, an $\varepsilon > 0$, and access to sampling oracle(s) for a hidden distribution π, the goal in identity testing is to distinguish whether the two distributions μ and π are identical or are at least ε-far apart. When there is only access to full samples from the hidden distribution π, it is known that exponentially many samples (in the dimension) may be needed for identity testing, and hence previous works have studied identity testing with additional access to various "conditional" sampling oracles. We consider a significantly weaker conditional sampling oracle, which we call the Coordinate Oracle, and provide a computational and statistical characterization of the identity testing problem in this new model.

We prove that if an analytic property known as approximate tensorization of entropy holds for an n-dimensional visible distribution μ, then there is an efficient identity testing algorithm for any hidden distribution π using $\widetilde{O}(n/\varepsilon)$ queries to the Coordinate Oracle. Approximate tensorization of entropy is a pertinent condition as recent works have established it for a large class of high-dimensional distributions. We also prove a computational phase transition: for a well-studied class of n-dimensional distributions, specifically sparse antiferromagnetic Ising models over $\{+1, -1\}^n$, we show that in the regime where approximate tensorization of entropy fails, there is no efficient identity testing algorithm unless $\mathsf{RP} = \mathsf{NP}$. We complement our results with a matching $\Omega(n/\varepsilon)$ statistical lower bound for the sample complexity of identity testing in the Coordinate Oracle model.

1. Introduction

A fundamental problem in statistics and machine learning is the identity testing problem (also known as the goodness-of-fit problem). Roughly speaking, we are explicitly given a visible distribution μ and oracle access to samples from an unknown/hidden distribution π; the goal is to determine if these distributions are identical using as few samples from π as possible.

The complexity of identity testing for general distributions is now well-understood; this includes conditions on the visible and hidden distributions which enable efficient identity testing; see (Canonne, 2020) for a comprehensive survey. An intriguing line of work considers a different perspective: what additional assumptions *on the sampling oracle* for the hidden distribution

are required to ensure efficient identity testing. We present tight results with more modest oracle assumptions than considered previously.

Let us begin with a formal definition of the classical identity testing framework. Let $\mathcal{X}$ be a finite state space of size $N = |\mathcal{X}|$, and let $d(\cdot,\cdot)$ denote a metric or divergence between distributions over $\mathcal{X}$; the standard choices for $d(\cdot,\cdot)$ are total variation distance (TV distance) or Kullback–Leibler divergence (KL divergence). For a distribution μ over $\mathcal{X}$ and a parameter $\varepsilon > 0$, denote by ID-TEST$(d,\varepsilon;\mu)$ the identity testing problem for μ: given as input the full description of the visible distribution μ, and given access to a sampling oracle for an unknown distribution π, our goal is to distinguish between the cases $\pi = \mu$ vs. $d(\pi,\mu) \geq \varepsilon$ with probability at least $2/3$.

For a distribution μ over $\mathcal{X}$, there are efficient identity testing algorithms with sample complexity $O(\sqrt{N}/\varepsilon^2)$ which matches, asymptotically, the information-theoretic lower bound; see (Valiant and Valiant, 2017; Paninski, 2008) for landmark results and (Chan et al., 2014; Acharya et al., 2015; Valiant and Valiant, 2017; Diakonikolas and Kane, 2016; Goldreich, 2016; Diakonikolas et al., 2021; Canonne and Sun, 2022) for other relevant works. (We recall that the sample or query complexity of an identity testing algorithm is the number of queries it sends to the sampling oracle.)

In practice, data is often high-dimensional, which raises the question of whether identity testing can be solved more effectively for high-dimensional distributions; this will be our focus. To be more precise, let $\mathcal{K} = \{1,\ldots,k\}$ be an alphabet (spin/color) set and let $\mathcal{X} = \mathcal{K}^n$ be a product space of dimension n. We study the identity testing problem ID-TEST$(d,\varepsilon;\mu)$ for n-dimensional distributions μ over $\mathcal{X}$.

Identity testing for high-dimensional distributions has recently attracted some attention, see, e.g., (Daskalakis and Pan, 2017; Daskalakis et al., 2019; Bezáková et al., 2019; Blanca et al., 2020; Canonne et al., 2020; Bhattacharyya et al., 2021, 2022). The focus is on visible distributions μ that have a $\mathrm{poly}(n)$ size description or parametrization; otherwise one could not hope to design efficient testing algorithms. Such distributions include product distributions (including the uniform distribution), Bayesian nets, and undirected graphical models (also known as spin systems) among others.

The goal is to design identity testing algorithms with $\mathrm{poly}(n)$ sample complexity and running times. It is known, however, that identity testing may require a super-polynomial (in n) number of samples (Bezáková et al., 2019; Blanca et al., 2020). (The algorithms for the general identity testing problem have sample complexity $\Omega(k^{n/2}/\varepsilon^2)$ in the high-dimensional setting since $|\mathcal{X}| = k^n$.)

Consequently, in order to design efficient algorithms, there are two types of further conditions that one may attach to the identity testing problem. The first approach is to restrict the unknown distribution π to be in some particular class of distributions; a natural example is to require that π is from the same class as μ. For example, (Bhattacharyya et al., 2021) studies the setting where both μ and π are product distributions, (Canonne et al., 2020; Daskalakis and Pan, 2017) requires μ and π to be Bayesian nets, (Daskalakis et al., 2019) studies the problem when μ and π are Ising models. More recently, (Bhattacharyya et al., 2022) considers the case where μ is a product distribution, and π is a Bayesian net. While such an approach leads to fruitful results for testing high-dimensional distributions, it is not ideal from a practical perspective, where π can be, for example, a "noisy" version of μ and may not necessarily belong to a nice class of distributions.

An alternative approach to overcome the apparent intractability of identity testing in the high-dimensional setting is to assume access to stronger sampling oracles from the hidden distribution π; specifically, access to conditional sampling oracles for π (in addition to the sampling oracle for π). This approach for high-dimensional distributions is the focus of this paper.

There are several types of conditional sampling oracles, and here we mention the most popular choices. The first is the general conditional sampling oracle—see (Canonne et al., 2015; Chakraborty et al., 2016; Falahatgar et al., 2015)—which given any subset $\mathcal{X}'$ of the space $\mathcal{X}$ generates a sample from the projection of π to $\mathcal{X}'$; that is, the oracle returns an element x from $\mathcal{X}'$ with probability $\pi(x)/\pi(\mathcal{X}')$. This oracle is not well-suited for the high-dimensional setting because the query subset $\mathcal{X}'$ could be exponentially large in n, and thus one could not hope to formulate the queries to the oracle efficiently (unless restricted to a special class of subsets $\mathcal{X}'$).

The second is the pairwise conditional sampling oracle (Pairwise Oracle) which takes a pair of configurations and generates a sample from the distribution restricted to these two choices: given $x, y \in \mathcal{X}$ the oracle returns x with probability $\pi(x)/(\pi(x)+\pi(y))$ and y otherwise; see (Canonne et al., 2015). The queries for Pairwise Oracle can be easily formulated for high-dimensional distributions, and identity testing has been studied in this setting. Recently, (Narayanan, 2021) provided an identity testing algorithm for the Pairwise Oracle model with $\widetilde{O}(\sqrt{n}/\varepsilon^2)$ sample complexity and a matching statistical lower bound; the $\widetilde{O}$ notation hides poly-logarithmic factors in n and $1/\varepsilon$.

The other conditional oracle previously studied in the high-dimensional setting is the subcube conditional sampling oracle (Subcube Oracle) introduced by Bhattacharyya and Chakraborty (2018) and also studied in (Canonne et al., 2021; Chen et al., 2021a). A query to the Subcube Oracle consists of a subset $\Lambda \subseteq [n] = \{1, \ldots, n\}$ of variables and a configuration $x \in \mathcal{K}^\Lambda$ on Λ. If $\pi(x) > 0$, the Subcube Oracle returns a sample $x' \in \mathcal{K}^{[n]\setminus\Lambda}$ from the conditional distribution $\pi(\cdot \,|\, x)$. For the Subcube Oracle, an identity testing algorithm using $\widetilde{O}(n^2/\varepsilon^2)$ queries was given in (Bhattacharyya and Chakraborty, 2018); improved algorithms were presented for uniformity testing in (Canonne et al., 2021) and for testing juntas in (Chen et al., 2021a).

In this work, we study identity testing for high-dimensional distributions under a weaker conditional sampling oracle, which we call the Coordinate Oracle. The Coordinate Oracle corresponds to the Subcube Oracle restricted to query sets Λ where $|\Lambda| = n-1$; that is, we fix the configuration at all but one coordinate and look at the conditional distribution at this particular coordinate given a fixed configuration on the remaining coordinates. Hence, access to the Coordinate Oracle is a much weaker assumption than access to the Subcube Oracle. We also note that the Subcube Oracle model can be significantly harder to simulate. For instance, for the classical ferromagnetic Ising model simulating the Coordinate Oracle is trivial, but sampling conditionally on arbitrary configurations, as required by the Subcube Oracle, is computationally hard (Goldberg and Jerrum, 2007).

Access to the Coordinate Oracle is also a weaker assumption than access to the Pairwise Oracle in the following sense. When $k = 2$ and $\mathcal{X} = \{0,1\}^n$, Coordinate Oracle access corresponds to Pairwise Oracle access restricted to pairs of configurations that differ in *exactly one coordinate*. When $k \geq 3$, one can simulate an δ-approximate Coordinate Oracle with Pairwise Oracle access in $\mathrm{poly}(k, \log(1/\delta))$ time (or a perfect one with $\mathrm{poly}(k)$ expected time) using a Markov chain. In addition, as in the case of the Subcube Oracle, simulating the Pairwise Oracle can be computationally more demanding than simulating the Coordinate Oracle. For example, in the context of the ferromagnetic Ising model on an n-vertex bounded degree graphs, a query to the Coordinate Oracle will require $O(1)$ random bits, but queries to the Pairwise Oracle may require $\Omega(n)$ random bits.

We provide here a computational and statistical characterization of the identity testing problem in the Coordinate Oracle model. Our focus is on imposing no conditions on the hidden distribution π, other than access to Coordinate Oracle, and explore which conditions on the visible distribution μ are necessary and sufficient for identity testing. We mention that the Coordinate Oracle oracle

has already been implicitly used in (Canonne et al., 2021) for uniformity testing (i.e., the special case of testing whether π is the uniform distribution).

Algorithmic results. For our algorithmic work we consider the identity testing problem under KL divergence, which we denote by $D_{\mathrm{KL}}(\cdot \,\|\, \cdot)$. From an algorithmic perspective, the choice of KL divergence is a natural one since, by Pinsker's inequality, an indentity testing algorithm for ID-TEST$(D_{\mathrm{KL}}(\cdot \,\|\, \cdot), 2\varepsilon^2; \mu)$ yields one for ID-TEST$(d_{\mathrm{TV}}(\cdot,\cdot), \varepsilon; \mu)$ (i.e., for identity testing under TV distance); the reverse is not true in general.

We start by introducing a key analytic property for the visible distribution, known as *approximate tensorization of entropy* (Caputo et al., 2015), which we will show is a sufficient (and essentially also necessary) condition for efficient identity testing in the high-dimensional setting. Approximate tensorization of entropy roughly states that the entropy of a distribution is bounded by the sum of the average conditional entropies at each coordinate.

Definition 1 (Approximate Tensorization of Entropy) *A distribution μ fully supported on $\mathcal{K}^n$ satisfies* approximate tensorization of entropy *with constant C if for any distribution π over $\mathcal{K}^n$:*

$$D_{\mathrm{KL}}(\pi \,\|\, \mu) \leq C \sum_{i=1}^{n} \mathbb{E}_{x \sim \pi_{n\setminus i}} \Big[D_{\mathrm{KL}}(\pi_i(\cdot \mid x) \,\|\, \mu_i(\cdot \mid x)) \Big], \tag{1}$$

where $\pi_{n\setminus i}(\cdot)$ denotes the marginal distribution of π on $[n] \setminus \{i\}$, and $\pi_i(\cdot \mid x)$ and $\mu_i(\cdot \mid x)$ denote the marginals of π and μ, respectively, on the i-th coordinate conditional on x.

The constant C achieves the minimum $C = 1$ when μ is a product distribution. More details about approximate tensorization and equivalent formulations are provided in the full version of this paper Blanca et al. (2022b).

Approximate tensorization of entropy is known to imply optimal mixing times of single-site update Markov chains, known as the Gibbs sampler or Glauber dynamics (Cesi, 2001; Chen et al., 2021d). It is also used to establish modified log-Sobolev inequalities and the concentration of Lipschitz functions under the distribution (Bobkov and Götze, 1999; Montenegro and Tetali, 2006).

There are a plethora of recent results establishing approximate tensorization in a wide variety of settings. In particular, (Chen et al., 2020) showed that the spectral independence condition introduced by (Anari et al., 2020) implies approximate tensorization of entropy for sparse undirected graphical models (i.e., spin systems on bounded degree graphs). Furthermore, recent works showed that spectral independence (and hence approximate tensorization) is implied by certain forms of correlation decay (Chen et al., 2020, 2021b; Feng et al., 2021), path coupling for local Markov chains (Blanca et al., 2022a; Liu, 2021), and the stability of the partition function (Chen et al., 2021c). As such, approximate tensorization is now known to hold with constant $C = O(1)$ (independent of n) for a variety of high-dimensional distributions; see, e.g., (Chen et al., 2021c; Blanca et al., 2022a; Liu, 2021; Chen et al., 2022; Galanis et al., 2022; Friedrich et al., 2022).

We show that approximate tensorization of the visible distribution μ yields an efficient identity testing algorithm, provided access to the Coordinate Oracle and the General Oracle for the hidden distribution π. Access to the General Oracle oracle (i.e., to independent full samples from π) is a standard assumption for testing under conditional sampling oracles. In particular, the General Oracle corresponds to Subcube Oracle restricted to $\Lambda = \emptyset$, so access to the Subcube Oracle implies access to the General Oracle, and previous work under the Pairwise Oracle assumes access to the General Oracle as well.

For our algorithmic result, we have four additional basic assumptions on the visible distribution μ. Specifically, we require that:

(i) μ has a description (parametrization) of $\mathrm{poly}(n)$ size;

(ii) the Coordinate Oracle can be implemented efficiently for the *visible* distribution μ;

(iii) μ is η-*b*alanced: there is a lower bound η so that the conditional probability of any spin $a \in \mathcal{K}$ at any coordinate i, fixing any configuration on $[n] \setminus \{i\}$, is at least η;

(iv) μ is fully supported on $\mathcal{K}^n$.

We discuss these assumptions in detail below (see Remark 3). Our algorithmic result for the Coordinate Oracle model follows.

Theorem 2 *Given a distribution μ over $\mathcal{X} = \mathcal{K}^n$ satisfying (i)-(iv) and* Approximate Tensorization *with constant C, there is a testing algorithm for* ID-TEST$(D_{\mathrm{KL}}(\cdot \,\|\, \cdot), \varepsilon; \mu)$ *with* Coordinate Oracle *and* General Oracle *access with $\widetilde{O}(n/\varepsilon)$ sample complexity and polynomial running time.*

A more precise theorem statement indicating the explicit dependence on C and η in the sample complexity is provided in the full version Blanca et al. (2022b). Several applications of Theorem 2 for well-studied high-dimensional distributions, including product distributions, sparse undirected graphical models in the so-called tree uniqueness regime, and distributions satisfying a Dobrushin-type uniqueness condition (Marton, 2019) are also given in the full version of this paper Blanca et al. (2022b).

We shall see in what follows that our algorithmic result for the Coordinate Oracle model in Theorem 2 is tight, both statistically and computationally; that is, we establish a matching $\Omega(n/\varepsilon)$ sample complexity lower bound and show that there is a class of high-dimensional distributions where identity testing is computationally hard in exactly the same settings where approximate tensorization of entropy does not hold.

A surprising feature of our algorithm is that it bypasses sampling from visible distribution μ; it does not even require the concentration of any statistics under μ. As in some of the previous algorithms for high-dimensional testing—e.g., those in (Daskalakis et al., 2019; Canonne et al., 2021)—our algorithm starts by "localizing" the testing problem (i.e., reducing it to a one-dimensional setting). For this, we crucially use the Approximate Tensorization of entropy of the visible distribution.

We then consider the problem of testing general (one-dimensional) distributions under KL divergence. It turns out that this problem has been largely overlooked in the literature (the aforementioned known results for identity testing are all under TV distance). This is likely because there are pairs of distributions with infinite KL divergence but arbitrarily small TV distance, and so testing under KL divergence is considered unsolvable in a worst-case sense; see (Daskalakis et al., 2018). However, we can aim for algorithms with sample complexities that depend on the visible distribution; i.e., instance-specific bounds instead of worst-case ones, as done in (Valiant and Valiant, 2017) for testing under TV distance.

We provide here an algorithm for the classical identity testing problem (that is, only access to the General Oracle is assumed) for general distributions under KL divergence; the sample complexity of our algorithm depends on the visible distribution (see Lemma 9). This is a key technical development towards establishing Theorem 2, and one we believe could be of independent interest.

Remark 3 *We pause now to discuss assumptions (i)-(iv) in Theorem 2. As mentioned, condition (i) is necessary as otherwise one can not hope to design testing algorithms with* $\mathrm{poly}(n)$ *running times. Condition (ii) formally states that for any coordinate* i*, and any fixed assignment* σ *for the other* $n-1$ *coordinates, we can compute the conditional distribution at* i *given* σ *in polynomial time. This is equivalent to requiring that a step of the Gibbs Sampler Markov chain for* μ *can be implemented efficiently; we believe (ii) is a mild assumption.*

The notion of η*-balancedness in condition (iii) is s byproduct of working with KL-divergence and is closely related to other coordinate marginal conditions that are required for efficient learning and sampling; specifically, under the assumption that* μ *has full support, it is equivalent to the notions* δ*-biased in (Klivans and Meka, 2017) and of* b*-marginally bounded distributions from (Chen et al., 2021c; Blanca et al., 2022a). Finally, we note that condition (iv) is also a byproduct of working with KL divergence but can be relaxed; we could require instead that the support of* π *is a subset of the support of* μ*. We emphasize that these conditions are all for the visible distribution* μ*, and that we impose no restrictions on the hidden distribution* π *(other than oracle access). For example, the uniform distribution, product distributions, and undirected graphical models (e.g., the Ising and hard-core models) satisfy the conditions in Theorem 2.*

Computational hardness results. We show next that the algorithmic result in Theorem 2 is computationally tight. In particular, for the antiferromagnetic Ising model (defined below), we establish the following computational phase transition for identity testing in the Coordinate Oracle model: (i) when approximate tensorization holds the problem can be solved efficiently, and (ii) when approximate tensorization does not hold, there is no polynomial-time testing algorithm unless $\mathsf{RP} = \mathsf{NP}$.

We do not directly prove that identity testing is hard when approximate tensorization fails. We show instead that the same strong correlations that cause approximate tensorization to fail, combined with the hardness of identifying the ground states of the model in the presence of strong correlations, imply the hardness of identity testing. (The ground states are the most likely configurations in the model, and for the antiferromagnetic Ising model correspond to the maximum cuts of the graph.)

We introduce the Ising model next, which is the simplest and most well-studied example of an undirected graphical model. Given a graph $G = (V, E)$, the set of configurations of the model is denoted by $\Omega = \{+1, -1\}^V$. For a real-valued parameter β, the probability of a configuration $\sigma \in \Omega$ is given by the Gibbs or Boltzmann distribution:

$$\mu_{G,\beta}(\sigma) = \frac{1}{Z_{G,\beta}} \cdot \exp\left(\beta \sum\nolimits_{\{v,w\}\in E} \sigma_v \sigma_w\right), \tag{2}$$

where the normalizing constant $Z_{G,\beta}$ is known as the partition function. When $\beta > 0$ the model is ferromagnetic/attractive and when $\beta < 0$ then the model is antiferromagnetic/repulsive.

The antiferromagnetic Ising model undergoes an intriguing computational phase transition at the threshold $\beta_c(d) = -\frac{1}{2}\ln(\frac{d}{d-2})$ for the parameter β. This threshold corresponds to the so-called uniqueness/non-uniqueness phase transition on the infinite d-regular tree; roughly speaking, in graphs of maximum degree at most d, when $\beta > \beta_c(d)$ long-range correlations die off, whereas when $\beta < \beta_c(d)$ long-range correlations persist.

For constant $d \geq 3$ and all $0 > \beta > \beta_c(d)$, the approximate sampling and counting (i.e., approximating the partition function $Z_{G,\beta}$) problems can be solved efficiently on any graph of maximum degree d (Chen et al., 2021d). Moreover, approximate tensorization holds in this regime, and hence Theorem 2 applies for identity testing in the Coordinate Oracle model. In contrast, it is also known

that when $\beta < \beta_c(d)$ there are no polynomial-time approximate sampling or counting algorithms unless RP $=$ NP (Sly and Sun, 2014; Galanis et al., 2016).

We establish here the computational hardness of identity testing in the Coordinate Oracle model in the same parameter regime $\beta < \beta_c(d)$, which thereby exhibits a similar computational phase transition for identity testing for the class of antiferromagnetic Ising models.

Theorem 4 *For sufficiently large constant $d \geq 3$ and constant $\beta < 0$, consider identity testing for the family of antiferromagnetic Ising models on n-vertex graphs of max degree d with parameter β.*

(i) If $\beta > \beta_c(d)$, then there exists a polynomial-time algorithm for identity testing under KL divergence with Coordinate Oracle *and* General Oracle *access with sample complexity $\widetilde{O}(n/\varepsilon)$;*

(ii) If $\beta < \beta_c(d)$, then there is no polynomial-time algorithm for identity testing under KL divergence with Coordinate Oracle *and* General Oracle *access unless* RP $=$ NP.

There are few analogous computational hardness results for identity testing; most lower-bound results in this setting are information-theoretic. The few examples appeared in (Bezáková et al., 2019; Blanca et al., 2020). These results apply to the identity testing problem with access only to General Oracle and require both the hidden and visible models to be Ising models. In our current setting, the visible model is an Ising model, but the hidden is an arbitrary high-dimensional distribution. This is a significant conceptual difference, and the techniques from (Bezáková et al., 2019; Blanca et al., 2020) do not easily extend (see Remark 10).

At a high level, as in (Bezáková et al., 2019), we prove the hardness result in Theorem 4(ii) using a reduction from the maximum cut problem. That is, given a graph $G = (V, E)$, we construct a testing instance that if solved, would find the maximum cut of G. In this approach, constructing a testing instance of small degree is a key challenge, and the "degree reducing" gadgets from (Bezáková et al., 2019; Blanca et al., 2020) no longer work in our setting.

Instead, we use a gadget introduced in (Sly, 2010) to establish the computational hardness of approximate counting antiferromagnetic spin systems. An interesting technical aspect of our proof is that we are required to design polynomial-time sampling algorithms to simulate the hidden oracles. This is difficult for us because sampling antiferromagnetic Ising models throughout the non-uniqueness regime, i.e., for all $\beta < \beta_c(d)$, is a notoriously hard problem (the problem is NP-hard even for regular graphs). We manage to design efficient sampling algorithms for our testing instances using the recent algorithmic result of Koehler, Lee, and Risteski (2022) that give an approximate sampling algorithm for Ising models when the edge interaction matrix has low rank, in conjunction with the sampling methods from (Jenssen et al., 2020) that use polymer models. A detailed overview of our reduction is given in Section 2.2. We mention that in the reductions in (Bezáková et al., 2019; Blanca et al., 2020), sampling is trivial, since there it is assumed that $\beta \ll \beta_c(d)$ (specifically, $|\beta|d = \Omega(\log n)$) and the instance is bipartite, so the Gibbs distribution concentrates in the configurations that align with the bi-partition; see Remark 10 for a detailed account of the novelties in our reduction to establish Theorem 4(ii).

Finally, we mention that the hardness result in Theorem 4(ii) extends to *any* conditional sampling oracle that could be implemented in polynomial time for the antiferromagnetic Ising model, and thus applies to identity testing in the Pairwise Oracle model, complementing the algorithmic results from (Canonne et al., 2015; Narayanan, 2021). On the other hand, they do not extend to the Subcube Oracle model since we do not know how to simulate this oracle efficiently.

Statistical lower bounds. We present next an information-theoretic lower bound for identity testing in the Coordinate Oracle model that matches the sample complexity of our testing algorithm for this model (Theorem 2). Our lower bound is for the special case of uniformity testing under TV distance when $k = 2$; i.e., the visible distribution is the uniform distribution over $\{0,1\}^n$.

Theorem 5 *Let μ be the uniform distribution over $\{0,1\}^n$. Then, any identity testing algorithm for* ID-TEST$(d_{\mathrm{TV}}(\cdot,\cdot),\varepsilon;\mu)$ *with access to both the* Coordinate Oracle *and the* General Oracle *requires $\Omega(n/\varepsilon^2)$ samples.*

A direct corollary of this result is that solving the identity testing problem under KL divergence requires $\Omega(n/\varepsilon)$ samples in the Coordinate Oracle model, thus showing that the sample complexity of our algorithm in Theorem 2 is asymptotically tight (up to logarithmic in n and $1/\varepsilon$ factors).

Our proof of Theorem 5 follows a well-known strategy. We construct a family of "bad" distributions $\mathcal{B}$, each of which has TV distance at least ε from the uniform distribution μ over $\{0,1\}^n$. The lower bounds follow from the fact that, for this carefully constructed family $\mathcal{B}$, one can not distinguish between sequences of independent samples from μ or from a distribution π chosen uniformly at random from $\mathcal{B}$. However, since our setting is adaptive, i.e., the choice of conditional queries of the testing algorithm may depend on the output to previous ones, we need to consider *query histories*, as in (Canonne et al., 2015; Narayanan, 2021). (Roughly speaking, a query history is a sequence of queries that the testing algorithm sends the oracle along with the outputs from the oracle.) To show that two query histories are indistinguishable (under μ or π), we use ideas from (Canonne et al., 2015) and the so-called *hybrid argument* in cryptography; see (Goldreich, 2004).

New results for the Subcube Oracle **model.** While the main focus of this work is the study of identity testing under weaker oracle assumptions (i.e., the Coordinate Oracle model), we also provide new results for identity testing in the previously studied Subcube Oracle model. Our first results for this model is an improved identity testing algorithm.

Theorem 6 *Let μ be an η-balanced distribution fully supported on $\mathcal{K}^n$ that has a $\mathrm{poly}(n)$ size parameterization. If we can compute the marginal probability at any coordinate conditioned on any partial configuration on any subset of coordinates, then there is an identity testing algorithm for* ID-TEST$(D_{\mathrm{KL}}(\cdot\,\|\,\cdot),\varepsilon;\mu)$ *for the* Subcube Oracle *model with $\widetilde{O}(n/\varepsilon)$ sample complexity and running time that depends on the time it takes to compute the coordinate conditional marginals.*

This algorithm, compared to the best-known algorithm for the Subcube Oracle model in (Bhattacharyya and Chakraborty, 2018), additionally requires that μ is η-balanced, but improves the sample complexity significantly from $\widetilde{O}(n^2/\varepsilon^2)$ to $\widetilde{O}(n/\varepsilon)$. In addition, compared to Theorem 2, this result for the Subcube Oracle does *not* require approximate tensorization of entropy. In fact, we point out several relevant settings where Theorem 6 applies, but approximate tensorization fails (or we do not know if it holds) and hence Theorem 2 does not apply: undirected graphical models (e.g., Ising model) on trees, Bayesian networks, mixtures of product distributions, and high-temperature Ising models and monomer-dimer models (i.e., weighted matchings) on arbitrary graphs. We remark that, similar to Theorem 2, Theorem 6 also holds under the weaker assumption that the support of μ contains the support of π; see Blanca et al. (2022b) for more details.

We also provide a matching lower bound for uniformity testing in the Subcube Oracle.

Theorem 7 *Let μ be the uniform distribution over $\{0,1\}^n$. Then, any identity testing algorithm for* ID-TEST$(D_{\mathrm{KL}}(\cdot\,\|\,\cdot),\varepsilon;\mu)$ *with access to the* Subcube Oracle *requires $\Omega(n/\varepsilon)$ samples.*

Note that when μ is the uniform distribution, then $\eta = \Theta(1)$, so Theorems 6 and 7 provide asymptotically matching sample complexity bounds for identity testing under KL divergence. Interestingly, if one considers uniformity testing under TV distance and Subcube Oracle access, then the recent work (Canonne et al., 2021) shows that $\widetilde{O}(\sqrt{n}/\varepsilon^2)$ oracle queries suffice. As far as we know, it is unclear if the testing algorithm from (Canonne et al., 2021) with sublinear sample complexity can be used for other high-dimensional distributions, e.g., general product distributions.

Our last result concerns *tolerant* identity testing in the Subcube Oracle model. In this problem, the goal is to distinguish between the cases $D_{\mathrm{KL}}(\pi \,\|\, \mu) \le \delta$ and $D_{\mathrm{KL}}(\pi \,\|\, \mu) \ge \delta + \varepsilon$ for $\delta, \varepsilon > 0$; identity testing corresponds to $\delta = 0$. We show that, under the same assumptions as in Theorem 6, one can estimate $D_{\mathrm{KL}}(\pi \,\|\, \mu)$ within additive error ε using $\widetilde{O}(n^4/\varepsilon^4)$ queries to the Subcube Oracle.

Theorem 8 *Let μ be an η-balanced visible distribution fully supported on $\mathcal{K}^n$ that has a $\mathrm{poly}(n)$ size parametrization. Suppose we can compute the marginal probability for μ at any coordinate conditioned on any partial configuration on a subset of coordinates. Given access to the* Subcube Oracle *for a hidden distribution π, there is an algorithm that for any $\varepsilon > 0$ computes $\widehat{S}$ such that, with probability at least $2/3$, we have $|\widehat{S} - D_{\mathrm{KL}}(\pi \,\|\, \mu)| \le \varepsilon$. The sample complexity of the algorithm is $\widetilde{O}\left(n^4/\varepsilon^4\right)$. The running time of the algorithm depends on the time it takes to compute the coordinate conditional marginals.*

2. Overview of Techniques

We present proof overviews for our main results in the Coordinate Oracle model: our testing algorithm (Theorem 2), the computational hardness (Theorem 4(ii)), and the lower bound (Theorem 5).

2.1. Algorithmic result for Coordinate Oracle model: Theorem 2

Suppose μ is the visible distribution and let π be an arbitrary distribution over $\mathcal{K}^n$. If approximate tensorization of entropy holds for μ with constant C, the following holds:

$$D_{\mathrm{KL}}(\pi \,\|\, \mu) \le Cn\, \mathbb{E}_{(i,x)}\left[D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x)\right],$$

where $i \in [n]$ is a uniformly random coordinate, $x \in \mathcal{K}^{n\setminus i}$ is generated from the marginal distribution $\pi_{n\setminus i}$ of π on $[n] \setminus \{i\}$, $p_i^x = \pi_i(\cdot \mid x)$, and $q_i^x = \mu_i(\cdot \mid x)$ (see Definition 1). Therefore, to distinguish between the cases $\pi = \mu$ and $D_{\mathrm{KL}}(\pi \,\|\, \mu) \ge \varepsilon$, it suffices to distinguish between:

$$p_i^x = q_i^x \text{ for all pairs } (i,x) \quad \text{vs.} \quad \mathbb{E}_{(i,x)}\left[D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x)\right] \ge \frac{\varepsilon}{Cn}.$$

This is the first step towards localizing the testing problem to a single coordinate. Now, under the η-balanced assumption for μ, we have that $0 \le D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x) \le \ln(1/\eta)$. Hence, if $\mathbb{E}_{(i,x)}\left[D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x)\right] \ge \frac{\varepsilon}{Cn}$, one can show via a reverse Markov inequality that there exists an integer $\ell \ge 1$, such that $2^\ell = O(n)$ and

$$\mathbb{P}_{(i,x)}\Big(D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x) \ge 2^\ell \cdot \frac{\varepsilon}{2Cn}\Big) \ge \frac{1}{2^\ell D}, \tag{3}$$

where $D = \Theta\big(\log(\frac{n\cdot\log(1/\eta)}{\varepsilon})\big)$.

With (3), it is not difficult to find a pair (i,x) such that $D_{\mathrm{KL}}(p_i^x \,\|\, q_i^x) \ge 2^\ell \cdot \frac{\varepsilon}{2Cn}$. This can done by first generating $O(\mathrm{poly}(n) \cdot D)$ pairs (i_t, x_t), by choosing $i_t \in [n]$ uniformly at random

and then using the General Oracle to sample the partial configuration x_t on $[n] \setminus \{i_t\}$. Then, we can exhaustively check for each ℓ (note that $\ell = O(\log n)$) whether among the generated pairs (i_t, x_t)'s there is one, say (i, x), satisfying that $D_{\mathrm{KL}}\left(p_i^x \,\|\, q_i^x\right) \geq 2^{\ell} \cdot \frac{\varepsilon}{2Cn}$. By (3), this will likely be the case.

In conclusion, we reduce (or localized identity testing for μ to solving identity testing for the one-dimensional distributions $p := p_i^x$ and $q := q_i^x$ on a domain of size k with respect to KL divergence. We assume q can be computed for the visible distribution μ (this is assumption (iii) in Theorem 2), and we have access to a sampling oracle for p_i^x using the Coordinate Oracle for π.

As mentioned earlier, in the distribution testing literature, testing under KL divergence has been overlooked. This is because there are pairs of distributions with infinite KL divergence but arbitrarily small TV distance, which entails that identity testing requires arbitrarily many samples even though the KL divergence is arbitrarily large. For example, this happens when q is the distribution on a single point 0 and p is the Bernoulli distribution with arbitrarily small mean (Daskalakis et al., 2018). As such, identity testing under KL divergence has been considered unsolvable in the sense of worst-case sample complexity for arbitrary p and q.

However, the identity testing problem under KL divergence makes perfect sense for specific visible distributions q if we are interested in the instance-specific sample complexity instead of the worst-case one, as in (Valiant and Valiant, 2017) under TV distance. Namely, for a given distribution q, what is the number of samples required, potentially depending on q, for the identity testing problem for q under KL divergence? We give next a first attempt at solving this problem. The sample complexity of our testing algorithm depends on the minimum probability $\eta = \min_{a \in \mathcal{K}} q(a)$.

Lemma 9 *Let $k \in \mathbb{N}^+$ and let $\varepsilon > 0$, $\eta \in (0, 1/2]$. Given a visible distribution q over domain $\mathcal{K}$ of size k such that $q(a) \geq \eta$ for any $a \in \mathcal{K}$, and given sample access to an unknown distribution p over $\mathcal{K}$, there exists a polynomial-time identity testing algorithm that distinguishes with probability at least $2/3$ between the cases $p = q$ or $D_{\mathrm{KL}}\left(p \,\|\, q\right) \geq \varepsilon$. The sample complexity of the identity testing algorithm is $O\big(\min\big\{\frac{1}{\varepsilon\sqrt{\eta}}, \frac{\sqrt{k}\ln(1/\eta)}{\varepsilon^2}\big\}\big)$ for $k \geq 3$ and $O\big(\frac{\ln(1/\eta)}{\varepsilon}\big)$ for $k = 2$.*

We remark that the dependency on η in the sample complexity is inevitable.

A natural first approach to identity testing under KL divergence to prove Lemma 9 is a reduction to testing under TV distance via the so-called reversed Pinsker's inequality: $D_{\mathrm{KL}}\left(p \,\|\, q\right) \leq (2/\eta) d_{\mathrm{TV}}\left(p, q\right)^2$. The sample complexity of such algorithm is $O(\sqrt{k}/(\varepsilon\eta))$. This is not optimal: for example, if q is the uniform distribution over $\mathcal{K}$ one has $\eta = 1/k$ and so the sample complexity is $O(k^{3/2}/\varepsilon)$, but one would expect the sample complexity to be $O(\sqrt{k}/\varepsilon)$, by analogy to what happens for testing in TV distance. A better reduction is to testing under ℓ_2 distance via the inequality: $D_{\mathrm{KL}}\left(p \,\|\, q\right) \leq (1/\eta)\left\|p - q\right\|_2^2$. We then need to distinguish between $p = q$ and $\left\|p - q\right\|_2 \geq \sqrt{\varepsilon\eta}$, which allows us to apply results from (Diakonikolas and Kane, 2016) and obtain an algorithm with sample complexity of $O(\left\|q\right\|_2/(\varepsilon\eta))$; this time we get the $O(\sqrt{k}/\varepsilon)$ sample complexity bound when q is the uniform distribution.

However, two major challenges ought to be solved for this approach to work. First, while $\left\|q\right\|_2$ can be bounded for certain specific distributions q (e.g., the uniform distribution), in general we do not have a bound for $\left\|q\right\|_2$. This can be solved via the *flattening* method from (Diakonikolas and Kane, 2016) which, roughly speaking, constructs a new testing instance (i.e., distributions p' and q' over $\mathcal{K}'$) that is equivalent to the initial one with the additional property that $\left\|q'\right\|_2$ is small. The idea is to divide "heavy" elements (those $a \in \mathcal{K}$ with large density $q(a)$) into many copies so that $q'(a') \approx 1/\ell$ for all $a' \in \mathcal{K}'$ where $\ell = |\mathcal{K}'|$; i.e., q' is close to uniform.

The second challenge is that, even if when $\|q\|_2$ small, the dependency on η could still be inverse-polynomial. This is particularly problematic when η decays sharply as k grows; e.g., $\eta = 2^{-k}$. To overcome this, we divide the domain $\mathcal{K}$ into two parts, those with small density $q(a) < \zeta$ and those with large density $q(a) \geq \zeta$ for some parameter ζ. We then deal with the two parts separately by running different testing algorithms on each. This can be viewed as a simple application of the *bucketing* technique from (Batu et al., 2001) using only two buckets.

While flattening and bucketing were previously known, the novelty of our approach is to combine them to get a stronger bound for the sample complexity, specifically by selecting the right scale ℓ for flattening and the right threshold ζ for bucketing. Our bound achieves $O(\sqrt{k}/\varepsilon)$ for q with $\eta = \Theta(1/k)$ such as the uniform distribution, and also maintains a $\sqrt{k}$ dependency even for biased q of tiny η, with only a logarithmic dependency on $1/\eta$.

2.2. Computational hardness in the Coordinate Oracle model: Theorem 4(ii)

We establish hardness of the identity testing problem as stated in Theorem 4(ii) for the antiferromagnetic Ising model with Coordinate Oracle and General Oracle access via a reduction from the maximum cut problem. Let $\{G = (V_G, E_G), k\}$ be an instance of the maximum cut problem. That is, we want to check whether $\textsf{max-cut}(G) < k$ or $\textsf{max-cut}(G) \geq k$. In our reduction, we construct an identity testing instance for the antiferromagnetic Ising model, feed it as input to a presumed testing algorithm, and claim that the output of the algorithm solves $\{G = (V_G, E_G), k\}$ with probability at least $2/3$; this is not possible unless $\mathsf{RP} = \mathsf{NP}$.

We start by constructing the multi-graph $F = (V_F, E_F)$ by adding two special vertices s and t to G; i.e., $V_F = V_G \cup \{s, t\}$. These two vertices are connected with $N^2 - k$ edges, where $N = |V_G|$. We also add N edges between s and each vertex of V_G and do the same for t so that:

1. When $\textsf{max-cut}(G) < k$, the cut $(\{s, t\}, V_G)$ of size $2N^2$ is the unique maximum cut of F;
2. When $\textsf{max-cut}(G) \geq k$, there exists another cut in F, other than $(\{s, t\}, V_G)$, of size $\geq 2N^2$.

This is because for $S \subset V_G$, the cut $(S \cup \{s\}, V_G \setminus S \cup \{t\})$ of F will have size: $\textsf{max-cut}(G) + |S|N + |V_G \setminus S|N + N^2 - k = 2N^2 + \textsf{max-cut}(G) - k$, which is $\geq 2N^2$ only when $\textsf{max-cut}(G) \geq k$.

We consider the antiferromagnetic Ising model on F. There is a natural bijection between the cuts of F and the configurations of the Ising model. In particular, each cut $(S, V_F \setminus S)$ of F corresponds to exactly two Ising configurations: vertices in S are assigned $+1$ and those in $V_F \setminus S$ are assigned -1 (and vice versa). From the definition of the model (see (2)) we also see that the "ground states" of the antiferromagnetic Ising model on F, that is the configurations of maximum probability in the Gibbs distribution, correspond precisely to the maximum cuts of F.

Let Ω be the set of all cuts of F and let Ω_0 be the set of all cuts $(S, V_F \setminus S)$ of F *except* those where $s \in S$, $t \in V_F \setminus S$, and the corresponding cut for G, i.e., $(S \setminus \{s\}, V_F \setminus \{S, t, s\})$, has size $\geq k$. This way, if $\textsf{max-cut}(G) < k$, then $\Omega_0 = \Omega$, and if $\textsf{max-cut}(G) \geq k$, then $\Omega \setminus \Omega_0$ contains the cuts of F corresponding to cuts of G of size $\geq k$.

We set the visible distribution of our testing instance to be the Gibbs distribution $\mu_{F,\beta}$ of the antiferromagnetic Ising model on F with $\beta < \beta_c(d) < 0$ in the tree non-uniqueness region. The hidden distribution will be $\mu_{F,\beta}(\cdot \mid \Omega_0)$; that is, $\mu_{F,\beta}$ conditioned on configurations that correspond to cuts in Ω_0. Our construction ensures that if $\textsf{max-cut}(G) < k$, then $\Omega = \Omega_0$ and so $\mu_{F,\beta}(\cdot \mid \Omega_0) = \mu_{F,\beta}$. Moreover, when $\textsf{max-cut}(G) \geq k$, we have $\Omega \neq \Omega_0$ and $\mu_{F,\beta}(\cdot \mid \Omega_0) \neq \mu_{F,\beta}$. In fact, it can

be shown that the TV distance between $\mu_{F,\beta}(\cdot \mid \Omega_0)$ and $\mu_{F,\beta}$ is $1 - o(1)$; intuitively, this is because $\Omega \setminus \Omega_0$ contains large cuts of F that account for a non-trivial portion of the probability mass of $\mu_{F,\beta}$.

Our reduction is then completed by generating samples from $\mu_{F,\beta}(\cdot \mid \Omega_0)$ and giving these samples and $\mu_{F,\beta}$ to the identity testing algorithm as input. The testing algorithm is guaranteed to succeed with probability at $2/3$. If the algorithm detects that the samples did not come from $\mu_{F,\beta}$, it means that $\mathsf{max\text{-}cut}(G) \geq k$; otherwise, it means that $\mathsf{max\text{-}cut}(G) < k$. Hence, we have a polynomial running time algorithm that solves the maximum cut problem with probability at least $2/3$, which is not possible unless $\mathsf{RP} = \mathsf{NP}$.

There are two important complications in this approach. First, F is a multi-graph of unbounded degree, and our goal is to establish hardness for the class of antiferromagnetic Ising models graphs of maximum degree $d = O(1)$ when $\beta < \beta_c(d)$. Second, we do not know how to generate samples from $\mu_{F,\beta}(\cdot \mid \Omega_0)$ efficiently in polynomial time.

Let us address first how we solve the issue of F being a multi-graph with large maximum degree. For this, we use a "degree reducing" gadget; the one we use was introduced in (Sly, 2010) to establish the hardness of approximate counting and sampling antiferromagnetic spin systems. Specifically, each vertex of F is replaced by a gadget H which consists of a (nearly) d-regular random bipartite graph with a relatively small number of trees attached to it. Being more precise, the leaves of each tree will be identified with unique vertices on the same side of the bipartite graph; the precise construction is given in the full version Blanca et al. (2022b). The root of these trees are called *ports* and are used to connect the gadgets as dictated by the edges of F. This results in a simple d-regular graph $\widehat{F}$.

A key feature of the gadget H is that in the tree non-uniqueness region $\beta < \beta_c(d)$, a sample from $\mu_{H,\beta}$ will have mostly $+1$'s on one side of H and mostly -1's on the other, or vice versa. Hence there are two possible "phases" for the gadget which we use to simulate the spin of the corresponding vertex in F; i.e., the phase of the gadget is mapped to the spin of the corresponding vertex of F. Therefore, in a configuration in $\widehat{F}$, the phase of all the gadgets determine a cut for F, and thus one for G. Consequently, the reduction described above from the maximum cut problem to identity testing using F can be done using $\widehat{F}$ instead.

The second technical complication is that we are required to sample from $\mu_{\widehat{F},\beta}(\cdot \mid \Omega_0)$. For this, we observe first that sampling a phase assignment from $\mu_{\widehat{F},\beta}(\Omega_0)$ is straightforward. We then sample the port configuration given the phase vector from Ω_0. This is done via a rejection sampling procedure by noting that the marginal distribution on the ports is within $o(1)$ total variation distance of a suitably defined product distribution. Once the port configuration is sampled within the desired accuracy, we sample the configuration on each gadget (independently) given the configuration of the ports. For this we use a hybrid approach: we use the recent algorithm from (Koehler et al., 2022) for low-rank Ising models for one range of values of β (i.e., when $|\beta|\sqrt{d} = O(1)$) and polymer models—see (Jenssen et al., 2020)—for the other. To use these algorithms, we fleshed out the spectrum of the incidence matrix of the gadget.

Remark 10 *In (Bezáková et al., 2019), hardness of identity testing was established when both the visible and hidden distributions are antiferromagnetic Ising models on graphs of bounded degree also via a reduction from the maximum cut problem. As such, we believe it is meaningful to detail the conceptual and technical differences, as well as some similarities, between the reduction described above and the one from (Bezáková et al., 2019). Conceptually, in (Bezáková et al., 2019) the hidden distribution π is assumed to be from the same class as μ, so the testing problem in consideration*

is easier. In fact, this problem is not hard for all $\beta < \beta_c(d) < 0$ since when $|\beta|d = O(\log n)$ it can be solved using the learning algorithm from (Klivans and Meka, 2017) to learn π. Only when $|\beta|d = \Omega(\log n)$, this variant of identity testing becomes computationally hard, and this is precisely what is established in (Bezáková et al., 2019). Our goal here is to show hardness throughout the entire non-uniqueness regime $\beta < \beta_c(d)$ (not only for $|\beta|d = \Omega(\log n)$), so the hidden distribution in our reduction can not be an Ising model. Our hidden distribution $\mu_{\widehat{F},\beta}(\cdot \mid \Omega_0)$ is instead a conditional *antiferromagnetic Ising distribution, and, as noted, sampling from it is challenging.*

At a technical level, the necessary assumption in (Bezáková et al., 2019) that $|\beta|d = \Omega(\log n)$ simplifies matters significantly. In particular, the degree reducing gadgets there simply consist of random regular bipartite graphs; when $\beta d = \omega(\log n)$, sampling from the antiferromagnetic Ising model on these gadgets is trivial since $1 - o(1)$ of the probability mass is concentrated on two trivial configurations ($+1$ in one side of the bipartite graph, -1 in the other side and vice versa). When $\beta < \beta_c(d)$, the correlations in the model are super-polynomially weaker; i.e., there is no such strong concentration in the ground states. As such, we must use a more sophisticated degree-reducing gadget (the one from (Sly, 2010) as discussed earlier), and consider the phase of the gadget to simulate spin assignments to vertices. In terms of similarities, the construction of the multi-graph F from the max-cut instance detailed above is nearly identical to the construction in (Bezáková et al., 2019), i.e., F is essentially the same, but $\widehat{F}$ is not since we must to use a different gadget.

Acknowledgments

AB research was supported by NSF grant CCF-2143762. EV was supported by NSF grant CCF-2147094.

References

Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. *Advances in Neural Information Processing Systems*, 28, 2015.

Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020.

Tugkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 442–451, 2001.

Ivona Bezáková, Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. Lower bounds for testing graphical models: colorings and antiferromagnetic Ising models. In *Proceedings of the 32nd Annual Conference on Computational Learning Theory (COLT)*, volume 99, pages 283–298, 2019.

Arnab Bhattacharyya, Sutanu Gayen, Saravanan Kandasamy, and NV Vinodchandran. Testing product distributions: A closer look. In *Algorithmic Learning Theory (ALT)*, pages 367–396, 2021.

Arnab Bhattacharyya, Clément L Canonne, and Joy Qiping Yang. Independence testing for bounded degree Bayesian network. *arXiv preprint arXiv:2204.08690*, 2022.

Rishiraj Bhattacharyya and Sourav Chakraborty. Property testing of joint distributions using conditional samples. *ACM Trans. Comput. Theory*, 10(4), 2018. Article No. 16.

Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. Hardness of identity testing for restricted Boltzmann machines and Potts models. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 514–529, 2020.

Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022a.

Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. Identity testing for high-dimensional distributions via entropy tensorization. *arXiv preprint arXiv:2207.09102*, 2022b.

Sergej G Bobkov and Friedrich Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999.

Clément L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing Library Graduate Surveys*, pages 1–100, 2020.

Clément L Canonne and Yucheng Sun. Optimal closeness testing of discrete distributions made (complex) simple. *arXiv preprint arXiv:2204.12640*, 2022.

Clément L Canonne, Dana Ron, and Rocco A. Servedio. Testing probability distributions using conditional samples. *SIAM Journal on Computing*, 44(3):540–616, 2015.

Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing bayesian networks. *IEEE Transactions on Information Theory*, 66(5):3132–3170, 2020.

Clement L. Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, page 321–336, 2021.

Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, Ser. 6, 24(4):691–716, 2015.

Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probability Theory and Related Fields*, 120:569–584, 2001.

Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah. On the power of conditional samples in distribution testing. *SIAM Journal on Computing*, 45(4):1261–1296, 2016.

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1193–1203. SIAM, 2014.

Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Learning and testing junta distributions with subcube conditioning. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, volume 134, pages 1060–1113, 2021a.

Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, 2020.

Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557, 2021b.

Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to Holant-type problems. In *Proceedings of the 62nd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 149–160, 2021c.

Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 1537–1550, 2021d.

Zongchen Chen, Nitya Mani, and Ankur Moitra. From algorithms to connectivity and back: finding a giant component in random k-SAT. *arXiv preprint arXiv:2207.02841*, 2022.

Constantinos Daskalakis and Qinxuan Pan. Square Hellinger subadditivity for Bayesian networks and its applications to identity testing. In *Conference on Learning Theory (COLT)*, pages 697–703, 2017.

Constantinos Daskalakis, Gautam Kamath, and John Wright. Which distribution distances are sublinearly testable? In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2747–2764. SIAM, 2018.

Constantinos Daskalakis, Nishanth Dikkala, and Gautam Kamath. Testing Ising models. *IEEE Transactions on Information Theory*, 65(11):6829–6852, 2019.

Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694, 2016.

Ilias Diakonikolas, Themis Gouleakis, Daniel M Kane, John Peebles, and Eric Price. Optimal testing of discrete distributions with high probability. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 542–555, 2021.

Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Faster algorithms for testing under conditional sampling. In *Proceedings of The 28th Conference on Learning Theory (COLT)*, volume 40, pages 607–636, 2015.

Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1558–1577, 2021.

Tobias Friedrich, Andreas Göbel, Martin S Krejca, and Marcus Pappik. A spectral independence view on hard spheres via block dynamics. *SIAM Journal on Discrete Mathematics*, 36(3):2282–2322, 2022.

Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Andrés Herrera-Poyatos. Fast sampling of satisfying assignments from random k-sat. *arXiv preprint arXiv:2206.15308*, 2022.

Leslie Ann Goldberg and Mark Jerrum. The complexity of ferromagnetic Ising with local fields. *Combinatorics, Probability and Computing*, 16(1):43–61, 2007.

Oded Goldreich. *Foundations of cryptography. II: Basic applications*. Cambridge University Press, 2004.

Oded Goldreich. The uniform distribution is complete with respect to testing identity to a fixed distribution. In *Electron. Colloquium Comput. Complex.*, volume 23, page 15, 2016.

Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020.

Adam Klivans and Raghu Meka. Learning graphical models using multiplicative weights. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 343–354, 2017.

Frederic Koehler, Holden Lee, and Andrej Risteski. Sampling approximately low-rank Ising models: MCMC meets variational methods. In *Proceedings of The 35th Conference on Learning Theory (COLT)*, 2022.

Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. In *APPROX-RANDOM*, 2021.

Katalin Marton. Logarithmic Sobolev inequalities in discrete product spaces. *Combinatorics, Probability and Computing*, 28(6):919–935, 2019.

Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends® in Theoretical Computer Science*, 1(3):237–354, 2006.

Shyam Narayanan. On tolerant distribution testing in the conditional sampling model. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 357–373, 2021.

Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

Allan Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.

Allan Sly and Nike Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6): 2383–2416, 2014.

Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.