Electronic Colloquium on Computational Complexity, Report No. 98 (2019)

Domain Compression and its Application to Randomness-Optimal Distributed Goodness-of-Fit

Jayadev Acharya*
Cornell University
acharya@cornell.edu

Clément L. Canonne†
Stanford University
ccanonne@cs.stanford.edu

Yanjun Han
Stanford University
yjhan@cs.stanford.edu

Ziteng Sun*‡
Cornell University
zs335@cornell.edu

Himanshu Tyagi§
Indian Institute of Science
htyagi@iisc.ac.in

July 19, 2019

Abstract

We study goodness-of-fit of discrete distributions in the distributed setting, where samples are divided between multiple users who can only release a limited amount of information about their samples due to various information constraints. Recently, a subset of the authors showed that having access to a common random seed (*i.e.*, shared randomness) leads to a significant reduction in the sample complexity of this problem. In this work, we provide a complete understanding of the interplay between the amount of shared randomness available, the stringency of information constraints, and the sample complexity of the testing problem by characterizing a tight trade-off between these three parameters. We provide a general distributed goodness-of-fit protocol that as a function of the amount of shared randomness interpolates smoothly between the private- and public-coin sample complexities. We complement our upper bound with a general framework to prove lower bounds on the sample complexity of this testing problems under limited shared randomness. Finally, we instantiate our bounds for the two archetypal information constraints of communication and local privacy, and show that our sample complexity bounds are optimal as a function of all the parameters of the problem, including the amount of shared randomness.

A key component of our upper bounds is a new primitive of *domain compression*, a tool that allows us to map distributions to a much smaller domain size while preserving their pairwise distances, using a limited amount of randomness.

*Supported by NSF-CCF-1846300 (CAREER).
†Supported by a Motwani Fellowship.
‡Supported by NSF-CCF-CRII-1657471.
§Supported by a grant from Robert Bosch Center for Cyber Physical Systems (RBCCPS), Indian Institute of Science.

ISSN 1433-8092

1 Introduction

A prototypical example of statistical inference is that of *goodness-of-fit*, in which one seeks to determine whether a set of observations fits a purported probability distribution. Considered extensively in Statistics and, more recently, in computer science under the name of *identity testing*, the goodness-of-fit question for discrete probability distributions is by now well-understood.

Most of the recent work has focused on the sample complexity of the problem (*i.e.*, the minimum number of observations required to solve the task), and sought to obtain sample-optimal, time-efficient algorithms (see, *e.g.*, [BFR+13, Pan08, ADK15, VV17, DGPP18]). In many emerging settings, however, time or even sample considerations may not be the main bottleneck. Instead, samples may only be partially accessible, or their availability may be subjected to strict information constraints. These constraints may be imposed in form of the number of bits allowed to describe each sample (communication constraints) or privacy constraints for each sample.

In this context, a recent line of work [ACT18, ACT19a] has provided sample-optimal algorithms under such information constraints. An important aspect revealed by this line of work is that shared randomness is very helpful for such problems – public-coin protocols have much lower sample complexity than private-coin protocols. However, shared randomness used by the distributed protocols may itself be an expensive commodity in practice. With an eye towards practical algorithms for deployment of these distributed statistical inference algorithms, we consider the question of randomness-efficient distributed inference algorithms.

Specifically, we consider *public randomness* as a resource. In our setting, n users get independent samples from an unknown k-ary distribution, and each can send a message to a central server in a one-way, non-interactive fashion. Those messages, however, have to comply with a prespecified *local information constraint*, such as communication (each message can be at most ℓ bits long) or local privacy (loosely speaking, messages must not divulge too much about the user's observation.) The server uses the n messages to perform the goodness-of-fit test for the unknown distribution.

Prior work considered two natural classes of protocols: *private-coin*, where users and server are randomized independently; and *public-coin*, where all parties share ahead of time a common random seed that they can leverage to coordinate their messages. Alternatively, one may view shared randomness as the communication sent over the "downlink" channel by the server to the users. In this paper, we significantly generalize prior results, by establishing a tight tradeoff between the number of users n and the number of shared random bits s required for performing inference under local information constraints.

A key component of our distributed protocols is *domain compression*, a new primitive we introduce. Roughly speaking, domain compression allows one to (randomly) map a large domain $[k]$ to a much smaller domain of size $L \ll k$, while ensuring that pairwise distances between probability distributions on $[k]$ are (roughly) preserved when looking at their induced distributions on $[L]$. This notion can then be leveraged to obtain testing protocols from "good" domain compression mappings which use few bits of randomness.

We proceed to describe our results in the next section, before giving an overview of our techniques in the subsequent section. To put our results in context, we then provide a brief overview of prior and related work.

1.1 Our Results

We first provide an informal overview of the setting and our results. We consider *identity testing*, a classic example of goodness-of-fit, where one is given a reference distribution $\mathbf{q}$ over a known domain of size k, as well as a parameter $\varepsilon \in (0,1)$. Upon receiving n i.i.d. samples $X_1, \dots, X_n$ from an unknown distribution $\mathbf{p}$ over the same domain, one must then output accept with high constant probability if $\mathbf{p} = \mathbf{q}$, and reject if the total variation distance between $\mathbf{p}$ and $\mathbf{q}$ is at least ε.

We study a distributed setting where the X_i's are distributed over n users who can only transmit a limited amount of information about their samples to a central server, which then seeks to solve the testing problem from the messages received (see Section 2 for the detailed setup, and Fig. 1 for a pictorial description). For simplicity, we focus on two main applications, communication constraints and local privacy; we point out, however, that our results are more general, and can be leveraged to obtain both upper and lower bounds for the more general class of information constraints described in [ACT18].

The communication-constrained setting. In this setting, each user can communicate at most ℓ bits to the server. We establish the following.

Theorem 1.1 (Informal)**.** *For every $k, \ell \geq 1, s \geq 0$, there exists a protocol for identity testing over $[k]$ with s bits of public randomness, ℓ bits communication per user, and*

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell} \vee 1}\sqrt{\frac{k}{2^{s+\ell}} \vee 1}\right).$$

users. Moreover, this number of users is optimal, up to constant factors, for all values of k, s, ℓ.

Note that for $\ell \geq \log k$, we recover the centralized (unconstrained) sample complexity of $O(\sqrt{k}/\varepsilon^2)$; for $s = 0$ and $s \geq \log k$, the expression matches respectively the public- and private-coin sample complexities established in previous work.

An interesting interpretation of the sample complexity result mentioned above is that "one bit of communication is worth two bits of public randomness." Equivalently, if one interprets the public randomness as an s bit random seed sent over the downlink channel to the users, who then reply with their ℓ-bit message, then improving the capacity of the downlink channel is only half as effective as improving the user-to-server channel capacities.

The locally private setting. In this setting, there is no bound on the length of the message each user can send to the server, but the randomized mechanism W used to decide which message y to send upon seeing sample x has to satisfy ϱ-local differential privacy (ϱ-LDP):

$$\max_{x \neq x'} \max_{y} \frac{W(y \mid x)}{W(y \mid x')} \leq e^{\varrho}. \tag{1}$$

(Equivalently, the probability to send any given message y must stay roughly within a $(1 \pm \varrho)$ multiplicative factor, regardless of which x was observed.) We prove the following.

Theorem 1.2 (Informal)**.** *For every $k \geq 1, \varrho \in (0,1], s \geq 0$, there exists a protocol for identity testing over $[k]$ under ϱ-LDP with s of public randomness, and*

$$n = O\left(\frac{k}{\varepsilon^2\varrho^2}\sqrt{\frac{k}{2^s} \vee 1}\right).$$

users. Moreover, this number of users is optimal, up to constant factors, for all values of k, s*, and* $\varrho \in (0,1]$.

Once again, for $s = 0$ and $s \geq \log k$, this recovers respectively the public- and private-coin sample complexities established in [ACFT19, ACT18]. In order to establish these upper bounds, along the way we provide a sample-optimal private-coin ϱ-LDP identity testing protocol (Lemma 4.7) which only requires one bit of communication per user (improving in this respect on the sample-optimal protocols of [ACFT19]), and may be of independent interest.

General local constraints. Both Theorems 1.1 and 1.2 illustrate the versatility of our approach. To establish our algorithmic upper bounds, we rely on a new primitive we call domain compression (on which we elaborate in the next subsection). Specifically, we show in Theorem 4.1 how to combine as a blackbox this primitive with a *private*-coin protocol for identity testing under any fixed type of local constraint to obtain a protocol for identity testing with s of public randomness, under the same local constraints.

Our proofs of optimality, similarly, are corollaries of a general lower bound framework (Lemma 5.4 and Theorem 5.5) we develop, and which extends that of [ACT18] to handle limited public randomness. We believe that both techniques — the domain compression primitive, and the general lower bound formulation – will find other applications in distributed statistical inference problems.

1.2 Our Techniques

Our proposed scheme has a modular form and, in effect, separates the use of shared randomness from the problem of establishing an information-constrained inference protocol. In particular, we use shared randomness only to enable *domain compression*.

Domain compression. The problem of domain compression is to convert samples from an unknown k-ary distribution $\mathbf{p}$ to samples from $[L]$, while preserving the total variation distances up to a factor of θ. Our main result here is a scheme that reduces the domain-size to roughly $L \approx k\theta^2$ while preserving the total variation distance up to a factor of θ. Furthermore, our randomized scheme does this using the optimal $2\log(1/\theta) + O(1)$ bits of randomness, which will be crucial for our applications. Furthermore, as we will see later, this is the best possible "compression" – the lowest L possible – for a given θ.

In order to come up with this optimal domain compression scheme, we establish first a one-bit ℓ_2 isometry for probability vectors. Namely, we present a random mapping which converts the domain to $\{0,1\}$ while preserving the ℓ_2 distances between pairs of probability vectors. We apply this scheme to non-overlapping parts of our k-ary probability vector to obtain the desired domain compression scheme. Underlying our analysis is a new anti-concentration bound for sub-Gaussian random variables, which maybe of independent interest.

Domain compression to distributed testing. With this general domain compression algorithm at our disposal, we use s bits of randomness to obtain a reduction of the domain size to roughly $k/2^s$, while shrinking the statistical distances by a factor of $1/\sqrt{2^s}$. Now that we have exhausted all our shared randomness in domain compression, we apply the best available private-coin protocol, but one working on domain of size $(k/2^s)$, with new distance parameter $\varepsilon/\sqrt{2^s}$ in place of the original ε.

Interestingly, when instantiating this general algorithm for specific constraints of communication, it is not always optimal to use all the randomness possible. In particular, when we have ℓ bits of communication per sample available, we should compress the domain to 2^ℓ and use the best private-coin protocol for ℓ bits of communication per sample. We formally show that *one bit of communication is worth two bits of shared randomness*. In particular, we should not "waste" any available bit of communication from the users by using too much shared randomness.

However, this only gives us a scheme with failure probability close to $1/2$ at best. To boost the probability of error to an arbitrarily small δ, the standard approach of repeating the protocol independently, unfortunately, is not an option, as we already have exhausted all available public randomness to perform the domain compression. Instead, we take recourse to a deterministic amplification technique [KPS85], which leverages the properties of expander graphs to achieve this failure probability reduction without using any additional random bit.

Optimality. When we instantiate our general algorithm for communication and privacy constraints, we attain performance that is jointly optimal in the information constraint parameter (bits for communication and the LDP parameter for privacy), the number of samples, and the bits of shared randomness. We establish this optimality by showing chi-square fluctuation lower bounds, a technique introduced recently in [ACT18]. This approach considers the interplay between a difficult instance of the problem and the choice of the mappings satisfying information constraints by the users. The main observation is that for public-coin protocols, the users can choose the best mapping for any given instance of the problem by coordinating using shared randomness, resulting in a minmax bottleneck. On the other hand, for private-coin protocols, for each choice of mappings, the users must handle the least favorable instance, resulting in a maxmin bottleneck. To obtain our lower bounds, we need to bridge between these two extremes and provide bounds which seamlessly switch from maxmin to minmax bounds as the number of bits of shared randomness increase. We term this significant generalization of chi-square fluctuation bounds the *semiminmax bound* and use it obtain tight bounds for our setting.

1.3 Prior and Related Work

Goodness-of-fit has a long and rich history in Statistics, starting with the pioneering work of Pearson [Pea00]. More recently, the composite goodness-of-fit question (where one needs to distinguish between the reference distribution, and all distributions sufficiently far in total variation from it) has been investigated in the theoretical computer science community under the name *identity testing* [GR00, BFR+13], with a focus on computational aspects and discrete distributions. This line of work culminated in efficient and sample-optimal testing algorithms [Pan08, VV17, ADK15, Gol16, DGPP18]; we refer the reader to the surveys [Rub12, Can15, BW18], as well as the recent book [Gol17] (Chapter 11) for further details on identity testing, and the more general field of distribution testing.

Recently, there has been a surge of interest in *distributed* statistical inference, focusing on density or parameter estimation under communication constraints [HMÖW18b, HÖW18, HMÖW18a, BHÖ19] or local privacy [DJW17, EPK14, YB18, KBR16, ASZ19, AS19]. The *testing* counterpart, specifically identity testing, was studied in the locally differentially private (LDP) setting by Gaboardi and Rogers [GR18] and Sheffet [She18], followed by [ACFT19]; and in the communication-constrained setting in [ACT19c, ACT19b], as well as by (with a slightly different focus) [FMO18]. The role of public randomness in distributed testing was explicitly studied

in [ACT19c, ACT19b], which showed a quantitative gap between the sample complexities of public- and private-coin protocols; those works, however, left open the fine-grained question of *limited* public randomness we study here.

Related to identity testing, a recent work of [DGKR19] considers identity testing under both memory and communication constraints. Their setting and results, however, are incomparable to ours, as the communication constraints they focus on are global (*i.e.*, the goal is to minimize the total communication between parties), with no hard constraint on any given user's message.

Our domain compression primitive, on the other hand, fits in the area of *dimensionality reduction*, a term encompassing various notions whose common theme is the mapping of high-dimensional objects into lower dimensions, while preserving (approximately) their relevant geometric features. In our case, the objects are elements of the $(k-1)$-dimensional probability simplex, and the geometric features are the pairwise distances (mostly in ℓ_1 distance); this is, especially in view of our use of an ℓ_2 isometry to achieve this goal, reminiscent of the celebrated Johnson-Linderstrauss (JL) lemma and its many applications [JLS86, IM98]. The JL lemma, however, is for general high-dimensional vectors, and does not necessarily map from nor into the probability simplex.

Closest to our primitive is the work of Kyng, Phillips, and Venkatasubramanian [KPV10], which considers a similar question for distributions over $\mathbb{R}^d$ satisfying a smoothness condition. However, their results are not applicable to our setting of finite alphabet. Furthermore, we are interested in preserving the total variation distance, and not Hellinger distance considered in [KPV10]. Finally, our proposed algorithm is randomness efficient, which is crucial for our application. In contrast, the algorithm in [KPV10] for domain compression requires a random mapping similar to the JL lemma construction.

2 Notation and Preliminaries

In what follows, we denote by $\log$ and $\ln$ the binary and natural logarithms, respectively. For an integer $k \geq 1$, we write $[k]$ for the set $\{1, \ldots, k\}$, and $\Delta(k)$ for the $(k-1)$-dimensional probability simplex $\Delta(k) := \{\mathbf{p} \colon [k] \to [0,1] : \sum_{x \in [k]} \mathbf{p}(x) = 1\}$ (where we identify a probability distribution with its probability mass function). For $\mathbf{p}, \mathbf{q} \in \Delta(k)$, recall that $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) := \sup_{S \subseteq [k]} (\mathbf{p}(S) - \mathbf{q}(S))$ is the *total variation distance* between $\mathbf{p}$ and $\mathbf{q}$, which is equal to half their ℓ_1 distance. For our lower bounds, we shall also rely on the chi-square distance between $\mathbf{p}$ and $\mathbf{q}$, defined as $\mathrm{d}_{\chi^2}(\mathbf{p} \,||\, \mathbf{q}) := \sum_{x \in [k]} (\mathbf{p}(x) - \mathbf{q}(x))^2 / \mathbf{q}(x)$. We indicate by $x \sim \mathbf{p}$ that x is a sample drawn from the distribution $\mathbf{p}$.

We will use standard asymptotic notations $O(f)$, $\Omega(f)$, $\Theta(f)$, as well as the (relatively) standard $\tilde{O}(f)$, which hides polylogarithmic factors in its argument.[1] We will, in addition, rely on the notation $a_n \lesssim b_n$ (resp. $a_n \gtrsim b_n$), to indicate there exists an absolute constant $C > 0$ such that $a_n \leq C \cdot b_n$ (resp. $a_n \geq C \cdot b_n$) for all n, and accordingly write $a_n \asymp b_n$ when both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. Finally, for a matrix $M \in \mathbb{R}^{m \times n}$, we denote by $\|M\|_F$ and $\|M\|_*$ the Frobenius and nuclear norms of M, respectively, and by $\rho(M)$ its spectral radius.

[1] Specifically, $g = \tilde{O}(f)$ means that there exists some absolute constant $c > 0$ such that $g = O(f \log^c f)$.

2.1 Setting and problem statement

In the (k, ε, δ)-identity testing problem, given a known reference distribution $\mathbf{q} \in \Delta(k)$, and given i.i.d. samples from $\mathbf{p}$, we seek to test if $\mathbf{p}$ equals $\mathbf{q}$ or if it is ε-far from $\mathbf{q}$ in total variation distance. Specifically, an (n, ε, δ)-test is given by a (randomized) mapping $\mathcal{T} \colon [k]^n \to \{0, 1\}$ such that

$$\Pr_{X^n \sim \mathbf{p}^n}[\,\mathcal{T}(X^n) = 0\,] > 1 - \delta \text{ if } \mathbf{p} = \mathbf{q},$$
$$\Pr_{X^n \sim \mathbf{p}^n}[\,\mathcal{T}(X^n) = 1\,] > 1 - \delta \text{ if } \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon.$$

That is, upon observing independent samples X^n, the algorithm should "accept" with probability at least $1 - \delta$ if the samples come from the reference distribution $\mathbf{q}$ and "reject" with probability at least $1 - \delta$ if they come from a distribution significantly far from $\mathbf{q}$. We will often fix the probability of failure δ to be a small constant, say $1/12$, and write (k, ε)-*identity testing* and (n, ε)-*test* for $(k, \varepsilon, 1/12)$-identity testing and $(n, \varepsilon, 1/12)$-test, respectively.[2] The sample complexity of (k, ε)-identity testing is the minimum n such that we can find an (n, ε)-test, over the worst-case reference distribution $\mathbf{q}$.

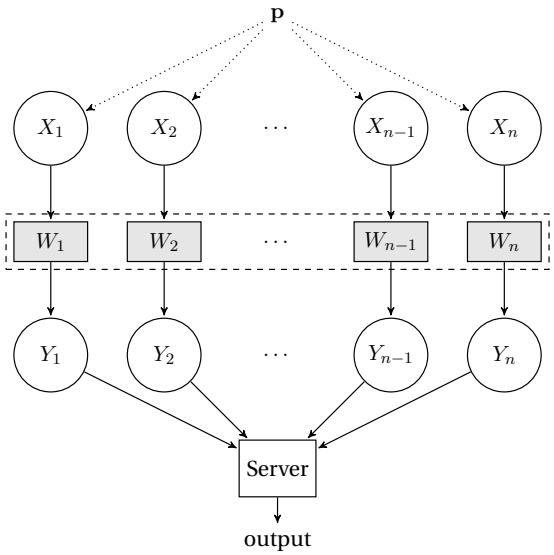

Figure 1: The information-constrained distributed model. In the private-coin setting the channels $W_1, \ldots, W_n$ are independent, while in the public-coin setting they are jointly randomized; in the s-coin setting, they are randomized based on both a joint U uniform on $\{0, 1\}^s$, and on n independent r.v.'s $U_1, \ldots, U_n$.

We work in the following distributed setting: n users each receive an independent sample from an unknown distribution $\mathbf{p} \in \Delta(k)$, and must send a message to a central server, in the simultaneous-message-passing (SMP) setting. The local communication constraints are modeled by a family $\mathcal{W}$ of "allowed" (randomized) channels, such that each user must select a channel $W \in \mathcal{W}$ and, upon seeing their sample x, send the message $y = W(x)$ to the central server. Here, we focus on s-coin SMP protocols, where the users have access to both private randomness, and

[2] Note that the specific choice of $1/12$ is merely for convenience, and any constant less than $1/2$ would do.

a limited number of uniform public random bits. Formally, s-coin SMP protocols are described as follows.

Definition 2.1 (s-coin SMP Protocols)**.** Let U be an s-bit random variable distributed uniformly over $\{0,1\}^s$, independent of $(X_1,\ldots,X_n)$; and let $U_1,\ldots,U_n$ denote independent random variables, which are independent jointly of $(X_1,\ldots,X_n)$ and U. In an *s-coin* SMP protocol, all users are given access to U, and further user i is given access to U_i. For every $i \in [n]$, user i selects the channel $W_i \in \mathcal{W}$ as a function of U and U_i. The central server is given access to the random variable U as well and its estimator and test can depend on U; however, it does not have access to the realization of $(U_1,\ldots,U_n)$.

In particular, for $s = 0$ we recover the private-coin setting, while for $s = \infty$ we obtain the public-coin setting. We then say an SMP protocol Π with n users is an (k,ε)-*identity testing s-coin protocol using $\mathcal{W}$ with n users* (resp. *public-coin*, resp. *private-coin*) if it is an s-coin SMP protocol (resp. *public-coin*, resp. *private-coin*) using channels from $\mathcal{W}$ which, as a whole, constitutes an (n,ε)-test.

The communication-constrained and LDP channel families. Two specific families of constraints we will consider throughout this paper are those of communication constraints, where each user can send at most ℓ bits to the server, and those of ϱ-LDP channels, where the users' channels must satisfy the definition of local differentially privacy given in (1). We denote those two families, respectively, by $\mathcal{W}_\ell$ and $\mathcal{W}_\varrho$:

$$\mathcal{W}_\ell := \{W\colon [k] \to \{0,1\}^\ell\}, \qquad \mathcal{W}_\varrho := \{W\colon [k] \to \{0,1\}^* : W \text{ satisfies } (1)\}\,.$$

A useful simplification. Throughout the paper, we will assume that the domain size k is a power of two. This can be done without loss of generality and does not restrict the scope of our results; we establish this reduction formally in Appendix B.

3 Domain Compression from Shared Randomness

We now introduce our main algorithmic tool – a new primitive called *domain compression*. We believe that the application of domain compression will go beyond this work. At a high-level, the domain compression problem requires us to convert statistical inference problems over large domain size to those over a small domain size. This problem is an instance of *universal compression*, since it is clear that we cannot assume the knowledge of the generating distribution of the samples. We present a simple formulation which can have applications for a variety of statistical tasks. Specifically, we require that pairwise distances be preserved between the distributions induced over the smaller domain. For our work, we only formulate a specific instance of the problem; it is easy to formulate a more general version which will have applications beyond the identity-testing problem that we consider, *e.g.*, to continuous distributions or other distance measures.

For a mapping $f\colon [k] \to [L]$ and $\mathbf{p} \in \Delta(k)$, denote by $\mathbf{p}^f$ the distribution of $f(X) \in [L]$.

Definition 3.1 (Domain compression)**.** For $L < k$, $\mathcal{U} := \{0,1\}^s$, and a mapping $\Psi\colon \mathcal{U} \times [k] \to [L]$, denote by Ψ_u, $u \in \mathcal{U}$, the mapping $\Psi_u(x) = \Psi(u,x)$. For $\theta \in (0,1)$, the mapping Ψ constitutes

an (L,θ,δ)-*domain compression mapping* ((L,θ,δ)-DCM) for $\Delta(k)$ if for all $\mathbf{p},\mathbf{q}\in\Delta(k)$ such that $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q})\geq\varepsilon$, the mapping satisfies

$$\Pr\Big[\,\mathrm{d}_{\mathrm{TV}}\Big(\mathbf{p}^{\Psi_U},\mathbf{q}^{\Psi_U}\Big)\geq\theta\cdot\varepsilon\,\Big]\geq 1-\delta, \tag{2}$$

where the randomness is over U which is distributed uniformly over $\mathcal{U}$. Furthermore, we say that this domain compression mapping uses s bits of randomness.

In effect, we are asking that a DCM preserves separation in total variation distance up to a loss-factor of θ while compressing the domain-size to L. For brevity, we shall say that such a DCM *compresses the domain-size* to L with a *loss-factor* of θ.

Our main result in this section, stated below, shows that we can compress the domain-size to $k\theta^2$ with a loss-factor of θ. Furthermore, we can do so using $2\log(1/\theta)$ bits of randomness.

Theorem 3.2. *Suppose $k=2^t$ for some $t\in\mathbb{N}$. Then, there exist positive constants c,δ_0 and c_0 such that, for every $\theta\in(\sqrt{c/k},\sqrt{c/2})$ and every $L\geq k\theta^2/2c$, there is an an (L,θ,δ_0)-DCM for $\Delta(k)$. Furthermore, this domain compression mapping uses at most $2\log(1/\theta)+c_0$ bits of randomness.*

Stated differently, we have a DCM that compresses the domain-size to L with a loss-factor of $\sqrt{L/k}$. In fact, this is the minimum loss-factor we must incur to compress the domain-size to L. Indeed, by choosing $L=2^\ell$, we can use the output of an (L,θ,δ)-DCM to enable uniformity testing using ℓ bits of communication. This output will be distributed over $[2^\ell]$ and the induced distribution will be separated from the uniform distribution by at least $\theta\varepsilon$ in total variation distance. Thus, using *e.g.,* the (non-distributed) uniformity test of [Pan08], we can complete uniformity testing using $\sqrt{2^\ell}/(\theta^2\varepsilon^2)$ samples. But this must exceed the lower bound of $k/(\varepsilon^2\sqrt{2^\ell})$ shown in [ACT18] for public-coin protocols. Therefore, θ must be less than $\sqrt{2^\ell/k}$. We will formalize this proof of optimality later (see Section 5), when we will show that the randomness of $2\log(1/\theta)$ bits that we use for attaining this corner-point of L versus θ tradeoff is optimal, too. Note that we can only achieve a constant δ from our scheme, which suffices for our purpose. A more general treatment of the domain-compression problem, with optimal tradeoff for all range of parameters, is an intriguing research direction.

As described, the domain compression problem requires us to preserve distances in total variation distance, which is equivalent to the ℓ_1 metric. We have setup this definition keeping in view the application of domain compression in identity-testing. In general, we can consider some other metrics. For instance, in place of Eq. (2) we can require

$$\Pr\Big[\,\|\mathbf{p}^{\Psi_U}-\mathbf{q}^{\Psi_U}\|_2\geq\theta\cdot\varepsilon\,\Big]\geq 1-\delta. \tag{3}$$

This is a stricter requirement since $\|x\|_1\geq\|x\|_2$, and would imply Eq. (2). In fact, using a random partition of the domain into $[L]$ parts, it was shown in [ACT19a, Theorem VI.2] that a loss-factor of roughly $1/\sqrt{k}$ can be attained for the definition of separation in Eq. (3). This in turn implies a scheme to compress domain-size to L with a loss-factor of $1/\sqrt{k}$, even for the definition of separation in Eq. (2). Comparing this with the result of Theorem 3.2, we find that the performance of this random partition based DCM is off by a $\sqrt{L}$ factor from the loss-factor of $\sqrt{L/k}$ attained by our proposed DCM in this paper. However, there is a simple modification than can help: Instead of applying this scheme to the entire domain, we can divide the domain into smaller parts and

ensure ℓ_2 separation for each part. If we divide the domain $[k]$ into equal parts and attain ℓ_2 separation loss-factor of θ for each part, this implies an overall loss-factor of θ in ℓ_1 as well.

To enable this approach, in the result below we establish a "one-bit isometry" for ℓ_2 distances between distributions. That is, we show that a random mapping Ψ with one-bit output exists such that the ℓ_2 distance between the distribution of output is at least a constant times the ℓ_2 distance between the distribution of input. Since the output is only binary, we can express the result in terms of difference between probabilities of sets that map to 1. Note that we need this isometry not only for distribution vectors $\mathbf{p}$ and $\mathbf{q}$, but also for subvectors of distribution vectors.

Theorem 3.3 (One-bit isometry)**.** *There exist absolute constants α, δ_0, c_0 and subsets $\{S_u\}_{u\in\mathcal{U}}$ of $[2^s]$ with $|\mathcal{U}| = 2^{s+c_0}$ and such that for every $\mathbf{p}, \mathbf{q} \in [0,1]^{2^s}$ we have*

$$\Pr[\,|\mathbf{p}(S_U) - \mathbf{q}(S_U)| \geq \alpha \,\|\mathbf{p}-\mathbf{q}\|_2\,] \geq 1-\delta_0, \tag{4}$$

where U is distributed uniformly over $\mathcal{U}$ and $\mathbf{p}(S) := \sum_{i\in S} \mathbf{p}_i$.

In other words, there is a randomized ℓ_2 isometry for distributions over $[2^s]$ that uses $s+c_0$ bits of randomness. The most significant aspect of the previous result, which is the main workhorse for this work, is that the sets $\{S_u\}_{u\in\mathcal{U}}$, are fixed and do not depend on vectors $\mathbf{p}$ and $\mathbf{q}$.

As outlined above, we want to apply our one-bit isometry to parts of domain. But there is one difficulty still left in implementing this idea to obtain our desired DCM: the guarantees are only for each part and the randomness requirement to make it work for all the parts simultaneously maybe higher. The following simple, but useful, observation comes to the rescue.

Lemma 3.4 (Additivity of tails)**.** *Let $a_1, \ldots, a_m \geq 0$, and $Y_1, \ldots, Y_m$ be non-negative random variables such that for some $c \in (0,1)$, $\Pr[\,Y_i \geq a_i\,] \geq c$ for every $1 \leq i \leq m$. Then,*

$$\Pr\left[\, Y_1 + \ldots + Y_m \geq c \cdot \frac{a_1 + \ldots + a_m}{2} \,\right] \geq \frac{c}{2-c}\,.$$

We defer the proof of this lemma to the appendix and of Theorem 3.3 to the end of this section. For now, we complete the proof of Theorem 3.2, our main theorem, using these results.

Proof of Theorem 3.2. Consider distributions $\mathbf{p}$ and $\mathbf{q}$ from $\Delta(k)$. Set $s = \lceil \log(c/\theta^2) \rceil$; then by our assumption, $s \leq t$. Further, denoting $J := 2^{t-s}$, for $0 \leq j \leq J-1$ define the vectors $\mathbf{p}^j$ and $\mathbf{q}^j$ in $[0,1]^{2^s}$ as $\mathbf{p}^j_i := \mathbf{p}_{j\cdot 2^s+i}$ and $\mathbf{q}^j_i := \mathbf{q}_{j\cdot 2^s+i}$ for all $i \in [2^s]$. We apply Theorem 3.3 to $\mathbf{p}^j$ and $\mathbf{q}^j$ to get

$$\Pr\left[\, |\mathbf{p}^j(S_U) - \mathbf{q}^j(S_U)| \geq \alpha \sqrt{\sum_{i\in S_U} (\mathbf{p}^j_i - \mathbf{q}^j_i)^2} \,\right] \geq 1-\delta_0, \quad 0 \leq j \leq J-1.$$

which together with the Cauchy–Schwarz inequality yields

$$\Pr\left[\, |\mathbf{p}^j(S_U) - \mathbf{q}^j(S_U)| \geq \alpha \cdot \frac{1}{\sqrt{2^s}} \cdot \sum_{i=1}^{2^s} |\mathbf{p}_{j\cdot 2^s+i} - \mathbf{q}_{j\cdot 2^s+i}| \,\right] \geq 1-\delta_0, \quad 0 \leq j \leq J-1,$$

We apply the "additivity of tails" property (Lemma 3.4) to arrive at

$$\Pr\left[\, \sum_{j=0}^{J-1} |\mathbf{p}^j(S_U) - \mathbf{q}^j(S_U)| \geq \frac{2\alpha}{\sqrt{2^s}} \cdot \mathrm{d_{TV}}(\mathbf{p},\mathbf{q}) \,\right] \geq \frac{1-\delta_0}{1+\delta_0}. \tag{5}$$

Consider the following function Ψ with range $\{0,\dots,2J-1\}$: For every $u \in \mathcal{U} = \{0,1\}^{s+c_0}$ and $i \in [k]$, let

$$\Psi(u,i) := \begin{cases} 2j, & i - j\cdot 2^s \in S_u, \\ 2j+1, & i - j\cdot 2^s \in [2^s] \setminus S_u, \end{cases} \qquad 0 \le j \le J-1.$$

Note that $\mathrm{d_{TV}}\left(\mathbf{p}^{\Psi_u}, \mathbf{q}^{\Psi_u}\right)$ equals $\sum_{j=0}^{J-1} |\mathbf{p}^j(S_U) - \mathbf{q}^j(S_U)|$. Then, Eq. (5) implies that Ψ constitutes a $(2J, 2\alpha/\sqrt{2^s}, 2\delta_0/(1+\delta_0))$-DCM. The proof is completed by setting $\theta := \sqrt{4\alpha^2/2^s}$ and noting that $2J = k\theta^2/(2\alpha^2)$. □

Proof of Theorem 3.3. Denote $x := \mathbf{p} - \mathbf{q}$ and consider a subset $S \subseteq [2^s]$. With these notations, the event we seek to handle is $(\sum_{i\in S} x_i)^2 \ge \alpha^2\|x\|_2^2$. We associate with S a vector $u \in \{0,1\}^{2^s}$ with ith entry given by $\mathbb{1}_{\{i\in S\}}$. Then, our of interest can be expressed as $x^\intercal(uu^\intercal)x \ge \alpha^2\|x\|_2^2$, where $\intercal$ denotes the transpose. Thus, we can associate a collection of vectors $S_1,\dots,S_m$ with a collection $u_1,\dots,u_m$. Then, our claim can be cast as the existence of $u_1,\dots,u_m$ such that

$$\frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\left\{x^\intercal\left(uu^\intercal\right)x < \alpha^2\|x\|_2^2\right\}} \le \delta_0.$$

Consider the set $\mathcal{J}$ of indices $j \in [m]$ given by $\mathcal{J} := \{j \in [m] : x^\intercal(u_j u_j^\intercal)x < \alpha^2\|x\|_2^2\}$. It is easy to see that by definition of $\mathcal{J}$, we have $x^\intercal\left(\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} u_j u_j^\intercal\right)x < \alpha^2\|x\|_2^2$, which further implies

$$\lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} u_j u_j^\intercal\right) < \alpha^2. \tag{6}$$

The main technical component of our proof is the following result.

Theorem 3.5 (Spectrum of outer products)**.** *For $n \in \mathbb{N}$, there exist constants $c_0 \in \mathbb{N}$, $c_1, c_2 \in (0,1)$ and vectors $u_1,\dots,u_m \in \{0,1\}^n$ with $m = 2^{c_0}n$ such that for every $\mathcal{J} \subseteq [m]$ with $|\mathcal{J}| \ge (1-c_1)m$ we must have*

$$\lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} u_j u_j^\intercal\right) \ge c_2.$$

Specifically, we show that random binary vectors $V_1,\dots,V_m$ will do the job. The proof is quite technical and requires a careful analysis of the spectrum of the random matrix $\sum_{j=1}^m V_j V_j^\intercal$. In particular, effort is required to handle entries of V_j with nonzero mean; we provide the complete proof in Appendix A.

We use vectors of Theorem 3.5, which implies that for vectors $u_1,\dots,u_m$ of Theorem 3.5 inequality (6) can hold only for $|\mathcal{J}| < (1-c_1)m$. Therefore,

$$\frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\left\{x^\intercal\left(u_j u_j^\intercal\right)x < c_2\|x\|_2^2\right\}} \le c_1,$$

whereby the claim follows for sets S_i, $i \in [m]$, given by $S_i = \mathtt{supp}(u_i)$ with $\delta_0 := c_1$ and $c_2 := \alpha^2$. □

4 Applications: Distributed Testing via Domain Compression

In this section, we show how the notion of domain compression developed in Section 3 yields distributed protocols for identity testing under local information constraints. Specifically, we show in Section 4.1 how to combine any private-coin identity testing protocol using $\mathcal{W}$ with an s-coin domain compression scheme to obtain an s-coin identity tester using $\mathcal{W}$. Then, in Sections 4.2 and 4.3, we instantiate this general algorithm with $\mathcal{W} = \mathcal{W}_\ell$ and $\mathcal{W} = \mathcal{W}_\varrho$ to obtain s-coin identity testing protocols under communication and local privacy constraints, respectively.

4.1 The General Algorithm

We establish the following result characterizing the performance of our general algorithm.

Theorem 4.1. *Let $\mathcal{W} \subseteq \{W : [k] \to \mathcal{Y}\}$ be a family of channels. Suppose there exists a (k,ε)-identity testing private-coin protocol using $\mathcal{W}$ with $n(k,\varepsilon)$ players. Then, for every $s_0 < s \leq \log k - c_0$, there exists a (k,ε)-identity testing s-public-coin protocol using $\mathcal{W}$ with $C \cdot n(ck/2^s, c'\varepsilon/2^{s/2})$ players, where $s_0, c_0, c, c' > 0$ are absolute constants and $C > 0$ is a constant depending on the desired probability of error.*

A few remarks are in order. First, we may view (and we will illustrate this in the following sections) this statement as saying that "an optimal private-coin testing protocol under local constraints yields, *as a blackbox*, an optimal s-coin testing protocol under the same local constraints, using domain compression." Second, in some cases (such as Section 4.2), it is beneficial to use this blackbox method, with a number of public coins s strictly smaller than the number of available public coins. Namely, we do better by ignoring some of the shared randomness resource. This is seemingly paradoxical, but the following heuristic may help resolve this conundrum: reducing the domain "too much" may prevent the private-coin tester from using fully what the local constraints allow. Concretely, in the case of communication constraints where each player can send ℓ bits, reducing the domain size below 2^ℓ means that some bits of communication cannot be utilized. Third, and foremost, this theorem hints at the versatility of our notion of domain compression and the simplicity of its use: (i) use public coins to reduce the domain while preserving the pairwise distances; (ii) run a private-coin protocol on the induced distributions, on the smaller domain.

Overview of the proof. Before delving into the details of the proof, we provide an outline of the argument. Suppose we have an identity testing private-coin protocol Π using $\mathcal{W}$. Given s of public randomness, we use the domain compression protocol from the previous section to reduce the domain size from k to $L \approx k/2^s$, while shrinking the total variation distances by a factor $\theta \approx 1/\sqrt{2^s}$. This entirely uses the s bits of public randomness, after which it suffices to use the private-coin Π to test identity of the induced distribution $\mathbf{p}' \in \Delta(L)$ to the induced reference distribution $\mathbf{q}' \in \Delta(L)$ with distance parameter $\theta \cdot \varepsilon \approx \varepsilon/\sqrt{2^s}$. Note that $\mathbf{q}'$ is known by all parties, as it is solely a function of $\mathbf{q}$ and the public randomness; and the players, after the domain compression, hold i.i.d. samples from $\mathbf{p}'$. Since the only communication between the parties occur when running the protocol Π (which by assumption uses channels from $\mathcal{W}$), the resulting protocol satisfies the local constraints modeled by $\mathcal{W}$.

This clean approach is indeed the main element of our algorithm. The issue, however, is that the domain compression only guarantees distance preservation with some constant probability

δ_0. Therefore, when $\mathbf{p}$ is ε-far from $\mathbf{q}$, the approach above can only guarantee correctness of the overall protocol with probability at most δ_0. In other words, the proposed protocol has low *soundness*. When $\mathbf{p} = \mathbf{q}$, however, the domain compression obviously yields $\mathbf{p}' = \mathbf{q}'$ with probability one, so the *completeness* guarantee holds. A standard approach to handle this would be to amplify the success probability by independent parallel repetitions, costing only a small constant factor overhead in the number of players. However, this is not an option for our setting, since independent repetitions would require fresh public randomness, *which we do not have anymore*. Further, dividing the public randomness in different random seeds and using these disjoint seeds to run this amplification-by-repetition idea would be suboptimal, as d repetitions would result in weaker domain compression – we will get domain of cardinality $k/2^{s/d}$ instead of the desired $k/2^s$.

To circumvent this issue, we use a different approach, that of *deterministic amplification* introduced in [KPS85]. The idea is indeed to run the protocol several times, say d, to amplify the probability of success, but carefully reusing the *same* s bit public randomness $U = r$ for all the d runs. Namely, we can find suitable mappings $\pi_1, \dots, \pi_d \colon \{0,1\}^s \to \{0,1\}^s$ such that upon running a protocol separately for (correlated) random seeds $\pi_1(r), \pi_2(r), \dots, \pi_d(r)$ and aggregating the results of the d distinct runs, we can amplify the success probability from $1/3$ to $\approx 1 - 1/d$. Specifically, we rely on the deterministic amplification lemma below, which guarantees that we can drive the error from any given constant to δ paying a factor $\tilde{O}(1/\delta)$ penalty in the runtime (*i.e.*, the number of parallel runs of the protocol, and therefore also number of players), but without using a *single* extra bit of public randomness.

Lemma 4.2 (Deterministic Amplification for One-Sided Error)**.** *For any $s \in \mathbb{N}$ and $\eta, \gamma \in (0,1)$, there exist $d = d(\eta, \gamma)$ and (time-efficiently computable) functions $\pi_1, \dots, \pi_d \colon \{0,1\}^s \to \{0,1\}^s$ such that the following holds. Suppose $\mathcal{X}_0 \subseteq \mathcal{X}$ and $A \colon \mathcal{X} \times \{0,1\}^s \to \Omega$ and $\mathcal{E} \subseteq \Omega$ satisfy the following:*

(i) If $x \in \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, A(x,\sigma) \in \mathcal{E}\,] = 1$; *(Perfect completeness)*

(ii) If $x \notin \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, A(x,\sigma) \notin \mathcal{E}\,] \geq 1 - \eta$. *(Low soundness)*

Then, we have

(i) If $x \in \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, \forall i \in [d],\ A(x, \pi_i(\sigma)) \in \mathcal{E}\,] = 1$; *(Perfect completeness)*

(ii) If $x \notin \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, \exists i \in [d],\ A(x, \pi_i(\sigma)) \notin \mathcal{E}\,] \geq 1 - \gamma$. *(High soundness)*

Moreover, on can take $d = \tilde{O}\Big(\frac{\eta}{(1-\eta)^2\gamma}\Big)$.

For completeness, we provide a self-contained proof of this result in Appendix C. Using the lemma above, we can provide a straightforward algorithm to increase soundness: given public randomness $r \in \{0,1\}^s$, we can divide the players in d disjoint groups for some suitable (constant) d. Group i then runs the natural protocol we discussed, using $\pi_i(r) \in \{0,1\}^s$ as its public randomness; and the server, upon seeing the outcomes of these d not-quite-independent protocols, aggregates them to produce the final outcome.

Remark 4.3 (Universality of our algorithm). Our proposed algorithm is universal in that the players are not required to know the reference distribution $\mathbf{q}$ (in contrast to previous work [ACT19a, ACFT19], which relied on a reduction to uniformity testing). The same protocol for choosing Ws from $\mathcal{W}$ works for *any* identity testing problem: the knowledge of $\mathbf{q}$ is only required for the center to complete the test.

We are now in position to provide the detailed proof of Theorem 4.1; the pseudocode of the resulting protocol is given in Algorithm 2.

Proof of Theorem 4.1. We hereafter set the constants c_0, c, δ_0 to be as in the statement of Theorem 3.2. Fix a reference distribution $\mathbf{q} \in \Delta(k)$, and let PRIVATEIDENTITYTESTING$_{\mathcal{W}}$ be a (k,ε)-identity testing private-coin protocol using $\mathcal{W}$ with $n(k,\varepsilon)$ players; with a slight abuse of notation, we will use the same name to invoke it with probability of failure δ for any chosen $\delta \in (0,1)$, using $n(k,\varepsilon,\delta) := n(k,\varepsilon)\cdot 72\ln(1/\delta)$ players.[3] Further, denote by $\Psi\colon \Delta(k)\times\{0,1\}^s \to \Delta(L)$ the (L,θ,δ_0)-domain compression mapping from Theorem 3.2, where[4] $\theta := 1/\sqrt{2^{s-c_0}}$ and $L := k\theta^2/2c$.

By Lemma 4.2 invoked with $\eta := 1-\delta_0$, $\gamma = 1/24$, there exist $d = \Theta(1)$ and (efficiently computable) $\pi_1,\dots,\pi_d\colon \{0,1\}^s \to \{0,1\}^s$ satisfying the conclusion of the lemma. We will apply it to the mapping $A\colon \Delta(k)\times\{0,1\}^s \to \{0,1\}$ defined by

$$A(\mathbf{p},r) := \mathbb{1}_{\left\{\mathrm{d_{TV}}(\mathbf{p}^{\Psi_r},\mathbf{q}^{\Psi_r}) \geq \theta\cdot \mathrm{d_{TV}}(\mathbf{p},\mathbf{q})\right\}}$$

where the event $\mathcal{E}$ considered is $\mathcal{E} := \{1\}$, *i.e.*, the event that the domain compression mapping is successful. Define $\delta' := \min(1-(11/12)^{1/d}, 1/24) = \Theta(1)$.

The protocol. Partition the n players into d groups $\Gamma_1,\dots,\Gamma_d$ of $N := n/d$ players, where by our setting of n we have $N \geq n(L,\theta\cdot\varepsilon,\delta')$. Given the public randomness $r \in \{0,1\}^s$, the N players in group Γ_i compute their "new" public randomness $r_i := \pi_i(r)$, and use it to run the domain compression Ψ. The N players in group i therefore obtain i.i.d. samples from a distribution $\mathbf{p}^{(i)} \in \Delta(L)$; moreover, both players and server can compute the induced reference distribution $\mathbf{q}^{(i)}$ (obtained by running the domain compression Ψ on $\mathbf{q}$ and randomness r_i). The players from Γ_i then run the protocol PRIVATEIDENTITYTESTING$_{\mathcal{W}}$ on their samples from $\mathbf{p}^{(i)}$, to test identity to $\mathbf{q}^{(i)}$, with parameters L, $\theta\cdot\varepsilon$ and failure probability δ'. This results in d bits $\nu_1,\dots,\nu_d \in \{0,1\}$ at the server, where ν_i is 0 if the protocol run by group i returned accept. The server then outputs 0 (accept) if, and only if, all ν_i's are equal to 0.

Correctness. First, observe that if $\mathbf{p} = \mathbf{q}$, then with probability one we have that $\mathbf{p}^{(i)} = \mathbf{q}^{(i)}$ for all $i \in [d]$, and therefore the probability that all d protocols return 0 (accept) is at least $(1-\delta')^d \geq 11/12$. This establishes the completeness.

Suppose now that $\mathrm{d_{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$. By definition of the domain compression protocol and our choice of A,η, we have that $\Pr_{r\sim\{0,1\}^s}[\,A(\mathbf{p},r) \notin \mathcal{E}\,] \geq \delta_0 = 1-\eta$. Lemma 4.2 then guarantees that, with probability at least $1-\gamma = 23/24$, there exists $i^* \in [d]$ such that $\mathrm{d_{TV}}(\mathbf{p}^{(i^*)},\mathbf{q}^{(i^*)}) > \theta\cdot\varepsilon$. When this happens, for this i^* the protocol run by the players in Γ_{i^*} will output $\nu_{i^*} = 1$ (reject) with probability at least $1-\delta'$, and therefore by a union bound the server outputs 1 (reject) with probability at least $1-(1/24+\delta') \geq 1-1/12$.

Number of samples. The analysis above requires that $N \geq n(L,\theta\cdot\varepsilon,\delta') \asymp n(L,\theta\cdot\varepsilon)\log(1/\delta')$. Recalling that $n = d\cdot N$ and that d, δ' are constant,[5] we have $n \asymp n(k/(2c2^{s-c_0}),\varepsilon/\sqrt{2^{s-c_0}})$ as claimed. □

[3] This is possible by the standard amplification trick: running the protocol independently several times and taking a majority vote. Crucially, this uses no shared randomness to perform as PRIVATEIDENTITYTESTING$_{\mathcal{W}}$ is private-coin.

[4] Recall that the conditions of Theorem 3.2 mandate $2-\log c^2 \leq s-c_0 \leq \log k - \log c^2$, for some constants c_0 and c.

[5] Specifically, if one aims for non-constant error probability $\delta \in (0,1)$ instead of $1/12$, we have $d = \tilde{O}(1/\delta)$ and $\delta' = \tilde{O}(\delta^2)$.

Algorithm 1 Domain compression protocol DOMAINCOMPRESSION

Require: Parameters $k > 1$, $s \geq 1$ with k a power of two; $X_1, \ldots, X_n \in [k]$ distributed among n players, random seed $u \in \{0,1\}^s$ available to all players.
Ensure: All players compute values L, θ, and obtain independent samples $X'_1, \ldots, X'_n \in [L]$

Set $\sigma \leftarrow s - c_0$, $\theta \leftarrow 1/\sqrt{2^\sigma}$, and $L \leftarrow k\theta^2/2c$ ▷ c_0, c are as in Theorem 3.2.
All players compute Ψ, the (L, θ, δ_0)-DCM for $\Delta(k)$ guaranteed by Theorem 3.2.
for $j \in [n]$ **do** ▷ As $2^s = c_0 \cdot 2^\sigma$, the players interpret u as a random seed for Ψ.
    Player j maps their sample $X_j \in [k]$ to $X'_j \leftarrow \Psi_U(X_j)$ in $[k]$.
return $L \in \mathbb{N}$, $\theta \in (0,1]$

Algorithm 2 The full (k, ε, δ)-identity testing protocol

Require: Parameters $k > 1$, $s \geq 0$ with k a power of two; $\varepsilon, \delta \in (0,1)$
Require: Private-coin protocol PRIVATEIDENTITYTESTING$_{\mathcal{W}}$ using $\mathcal{W}$ with $n(k, \varepsilon, \delta)$ players
Ensure: This is a (k, ε, δ)-identity testing protocol as long as $n \geq C_\delta \cdot n(k, \varepsilon, \delta)$

Deterministic error reduction
    Apply Lemma 4.2 to $\eta := 3/4$, $\gamma = \delta/2$, to obtain mappings $\pi_1, \ldots, \pi_d \colon \{0,1\}^s \to \{0,1\}^s$
    Partition the n players into d groups $\Gamma_1, \ldots, \Gamma_d \subseteq [n]$ of $N \leftarrow n/d$ players
    Set $\delta' \leftarrow \min(1 - (1-\delta)^{1/d}, \delta/2)$
Domain compression
▷ The constants c_0, c are as in Theorem 3.2.
    if $s > c_0$ **then** ▷ Enough public coins are available.
        All players agree on a uniformly random $R \in \{0,1\}^s$. ▷ This uses s public coins.
        for $i \in [d]$ **do** ▷ players in group i run the protocol on randomness $\pi_i(R)$
            $(L, \theta) \leftarrow$ DOMAINCOMPRESSION$(k, s, (X_j)_{j \in \Gamma_i}, \pi_i(R))$
    else ▷ If too few public coins are available, use directly the private-coin protocol.
        $(L, \theta) \leftarrow (k, 1)$
        for $j \in [n]$ **do** player j sets $X'_j \leftarrow X_j$. ▷ Keep same sample
Private-Coin Tester
    for $i \in [d]$ **do** ▷ Each group runs the private-coin identity testing protocol using $\mathcal{W}$
        Let $\mathbf{q}^{(i)}$ be the reference distribution induced by DOMAINCOMPRESSION run on $\pi_i(R)$.
        $\nu_i \leftarrow$ PRIVATEIDENTITYTESTING$_{\mathcal{W}}(\mathbf{q}^{(i)}, L, \theta \cdot \varepsilon, \delta', (X'_j)_{j \in \Gamma_i})$ ▷ $\nu_i = 0$ if the test accepts
return 0 (accept) if $\nu_i = 0$ for all $i \in [d]$, 1 (reject) otherwise

In the two next subsections, we will illustrate the versatility of Theorem 4.1 by applying it to ℓ-bit local communication constraints and ϱ-local privacy constraints, respectively, to obtain sample-optimal protocols.

4.2 Communication-Constrained Testing

In the communication-constrained setting each player can only send $\ell < \log k$ bits to the server: *i.e.*, $\mathcal{W} = \mathcal{W}_\ell$, where $\mathcal{W}_\ell = \{W \colon [k] \to \{0,1\}^\ell\}$. We establish the following theorem:

Theorem 4.4. *For any integers $\ell \geq 1, s \geq 0$, there exists an ℓ-bit communication protocol with s bits of public randomness using*

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell} \vee 1}\sqrt{\frac{k}{2^{s+\ell}} \vee 1}\right)$$

players to perform (k,ε)-identity testing. In particular, for $\ell+s \leq \log k$, this becomes $O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\sqrt{\frac{k}{2^{s+\ell}}}\right)$.

As we shall see in Section 5, this is sample-optimal.

Proof of Theorem 4.4. We note first that for $\ell \geq \log k$, the setting becomes equivalent to the centralized setting, and the claimed expression becomes $O(\sqrt{k}/\varepsilon^2)$, the (known) tight centralized sample complexity. Thus, it is sufficient to focus on $1 \leq \ell < \log k$, which we hereafter do. To apply Theorem 4.1, we utilize the *simulate-and-infer* private-coin identity testing protocol of [ACT19a]. Specifically, we invoke the following result from [ACT19a], which gives a sample-optimal private-coin identity testing protocol Π_ℓ using $\mathcal{W}_\ell$:

Theorem 4.5 ([ACT19a, Corollary IV.3])**.** *For any integer $\ell \geq 1$, there exists a private-coin (k,ε,δ)-identity testing protocol using $\mathcal{W}_\ell$ and*

$$n = O\left(\frac{k}{2^\ell\varepsilon^2}\left(\sqrt{k\log\frac{1}{\delta}} + \log\frac{1}{\delta}\right)\right)$$

players. In particular, for constant δ this becomes $n(k,\varepsilon) = O\left(\frac{k^{3/2}}{2^\ell\varepsilon^2}\right)$.

Armed with this protocol Π_ℓ, we proceed as follows. Set $\bar{s} \leftarrow \min(\log(k) - \ell, s)$ to be the "effective" number of usable public coins (intuitively, if more than $\log k - \ell$ public coins are available, it is not worth using them all, as compressing the domain below 2^ℓ would render some of the ℓ available bits of communication useless).

- If $\bar{s} \leq c_0$ (where c_0 is the constant from the statement of Theorem 3.2), then we simply run the private-coin protocol Π_ℓ. This requires

$$n \geq n(k,\varepsilon) \asymp \frac{k^{3/2}}{2^\ell\varepsilon^2} = \frac{\sqrt{k}}{\varepsilon^2}\cdot\sqrt{\frac{k}{2^\ell}}\cdot\sqrt{\frac{k}{2^\ell}} \asymp \frac{\sqrt{k}}{\varepsilon^2}\cdot\sqrt{\frac{k}{2^\ell}}\cdot\sqrt{\frac{k}{2^{\ell+s}} \vee 1},$$

  since either $s \leq c_0$ (in which case $\frac{k}{2^\ell} \asymp \frac{k}{2^{\ell+s}}$) or $\log(k) - \ell \leq c_0$ (in which case $\frac{k}{2^{\ell+s}} \leq \frac{k}{2^\ell} \lesssim 1$).
- Else, we apply Theorem 4.1 with $\bar{s}$ bits of public randomness and private-coin identity testing protocol Π_ℓ. This can be done as long as

$$\begin{aligned} n \geq C\cdot n(ck/2^{\bar{s}}, c'\varepsilon/2^{\bar{s}/2}) \asymp \frac{(k/2^{\bar{s}})^{3/2}}{2^\ell(\varepsilon/2^{\bar{s}/2})^2} = \frac{k^{3/2}}{2^\ell\varepsilon^2 2^{\bar{s}/2}} &= \frac{\sqrt{k}}{\varepsilon^2}\cdot\sqrt{\frac{k}{2^\ell}}\cdot\sqrt{\frac{k}{2^{\ell+\bar{s}}}} \\ &= \frac{\sqrt{k}}{\varepsilon^2}\cdot\sqrt{\frac{k}{2^\ell}}\cdot\sqrt{\frac{k}{2^{\ell+s}} \vee 1}, \end{aligned}$$

  where the last identity holds since $\bar{s} = (\log(k) - \ell) \wedge s$.

This concludes the proof. □

4.3 Locally Differentially Private Testing

In this section, we consider the locally private channel family, where each player can only send a message that is ϱ-LDP. That is, recalling Eq. (1), we consider the channel family

$$\mathcal{W} = \mathcal{W}_\varrho = \{\, W : [k] \to \mathcal{Y} \,:\, \forall y \in \mathcal{Y}, \forall x_1, x_2 \in [k], W(y \mid x_2) \le e^\varrho W(y \mid x_1) \,\} \,.$$

We establish the following result for performance of our proposed general algorithm for testing under privacy constraints. It will be seen in the next section that, much like the communication-constrained setting, for the privacy-constrained setting as well our general algorithm is optimal.

Theorem 4.6. *For any integers $k \ge 1, s \ge 0$, and parameter $\delta \in (0,1), \varrho > 0$, there exists a one-bit communication ϱ-LDP protocol with s bits of public randomness using*

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{\sqrt{k}}{\varrho^2}\sqrt{\frac{k}{2^s} \vee 1}\right)$$

players to perform (k,ε,δ)-identity testing. When $s > \log k$, this becomes $O\left(\frac{k}{\varepsilon^2\varrho^2}\right)$.

In [ACFT19], it was shown that the sample complexity for identity testing with ϱ-local differential privacy constraints is $\Theta(k^{3/2}/(\varepsilon^2\varrho^2))$ using only private randomness and $\Theta(k/(\varepsilon^2\varrho^2))$ with (unlimited) public randomness.[6] Theorem 4.6 matches these bounds in both cases. Moreover, we note here that for private-coin schemes, we can achieve the optimal sample complexity with a one-bit communication protocol. This is in contrast with the private-coin protocols of [ACFT19] which require $\Omega(\log k)$ bits of communication per player. This also shows that, unlike the communication-constrained setting, under LDP constraints there is no tradeoff between the number of available bits of communication and sample complexity.

Proof. We will rely on the following lemma, which improves on the private-coin protocol of [ACFT19] in terms of communication complexity (while achieving the same sample complexity). The protocol is inspired by that of [AS19], which provides a one-bit LDP protocol for distribution *learning*.

Lemma 4.7. *There exists a one-bit communication private-coin ϱ-LDP protocol that uses*

$$n = O\left(\frac{k^{3/2}}{\varepsilon^2\varrho^2}\log\frac{1}{\delta}\right)$$

players to perform (k,ε,δ)-identity testing. For constant δ this becomes $n_\varrho(k,\varepsilon) = O\left(\frac{k^{3/2}}{\varepsilon^2\varrho^2}\right)$.

We defer the proof for this intermediate result to Section 4.3.1, and continue the proof of the theorem assuming the statement. Let us denote by Π_ϱ the protocol from Lemma 4.7; we then proceed as follows:

- If $s \le c_0$ (where c_0 is the constant from the statement of Theorem 3.2), then we just run the private-coin protocol Π_ϱ.

[6] Although, as the authors showed, their protocol could be made to work with $O(\log k)$ bits of public randomness.

- Else, we apply Theorem 4.1 with $\bar{s} = \min(\log k, s)$ bits of public randomness and private-coin identity testing protocol Π_ϱ. This can be done as long as

$$n \geq C \cdot n_\varrho(ck/2^{\bar{s}}, c'\varepsilon/2^{\bar{s}/2}) \asymp \frac{(k/2^{\bar{s}})^{3/2}}{\varrho^2(\varepsilon/2^{\bar{s}/2})^2} = \frac{k^{3/2}}{\varrho^2\varepsilon^2 2^{\bar{s}/2}} = \frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2} \cdot \sqrt{\frac{k}{2^s} \vee 1}$$

  the last equality recalling that $\bar{s} = (\log k) \wedge s$.

□

4.3.1 Proof of Lemma 4.7

It only remains to prove Lemma 4.7, our intermediary result giving a communication-efficient private-coin protocol for identity testing under LDP. We emphasize that the main advantage of this protocol is that we require only one bit of communication per player as compared to $\Omega(\log k)$ for those of [ACFT19], while in terms of sample complexity both protocols are optimal.

Proof of Lemma 4.7. We use the same response scheme as in [AS19]. The scheme is the following. Let $K := 2^{\lceil \log_2(k+1) \rceil}$, which is the smallest power of two larger than k. Let H_K be the $K \times K$ Hadamard matrix. Without loss of generality, we assume K divides n (as otherwise we can ignore the last $(n - K\lfloor \frac{n}{K} \rfloor)$ players). Deterministically partition divide the players into K disjoint blocks of equal size $B_1, B_2, \ldots, B_K$. Each player $i \in B_j$ is assigned the jth column of the Hadamard matrix. Let C_j be the location of $+1$'s on the jth column; the channel used by player $i \in B_j$ is given by

$$W_i(1 \mid x) = \begin{cases} \frac{e^\varrho}{e^\varrho+1}, & \text{if } x \in C_j, \\ \frac{1}{e^\varrho+1}, & \text{otherwise.} \end{cases} \tag{7}$$

Then, following the same computations as in [AS19], we have that for all $j \in [K]$,

$$p_{C_j} := \Pr[\, Y_i = 1 \mid i \in B_j \,] = \frac{e^\varrho - 1}{e^\varrho + 1}\mathbf{p}(C_j) + \frac{1}{e^\varrho + 1}\,.$$

Taking one player from each block and viewing the resulting collection of messages as a length-K vector, we thus get n/K samples from a product distribution on $\{0,1\}^K$ with mean vector $p_C := (p_{C_1}, \ldots, p_{C_K})$. From a Parseval-based argument analogous to [ACFT19], we then know that

$$\|p_C - q_C\|_2^2 = \frac{K(e^\varrho - 1)^2}{4(e^\varrho + 1)^2}\|\mathbf{p} - \mathbf{q}\|_2^2\,,$$

where $q_C \in [0,1]^K$ is the mean vector obtained as above when the input distribution is $\mathbf{q}$ instead of $\mathbf{p}$. (Note that q_C can be explicitly computed given knowledge of $\mathbf{q}$.) Therefore, when $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon$, $\|p_C - q_C\|_2^2 > \alpha := \frac{(e^\varrho-1)^2}{(e^\varrho+1)^2}\varepsilon^2$, while $\|p_C - q_C\|_2^2 = 0$ when $\mathbf{p} = \mathbf{q}$. Since, for product distributions over $\{0,1\}^K$, the problem of testing whether the mean vector is either (i) a prespecified vector $\mu \in \mathbb{R}^K$ or (ii) at ℓ_2 distance at least α from μ has sample complexity $O\big(\sqrt{K}\log(1/\delta)/\alpha^2\big)$,[7] having $n/K \gtrsim \sqrt{K}\log(1/\delta)/\alpha^2$ suffices, *i.e.*,

$$n = O\left(k^{3/2}\frac{(e^\varrho+1)^2}{4(e^\varrho-1)^2\varepsilon^2}\log\frac{1}{\delta}\right) = O\left(\frac{k^{3/2}}{\varepsilon^2\varrho^2}\log\frac{1}{\delta}\right),$$

[7] This is more or less folklore; see *e.g.*, [CDKS17, Section 2.1], or [CKM+19, Lemma 4.2].

as claimed. Finally, the fact that this protocol does, indeed, satisfy the ϱ-LDP constraints is immediate from Eq. (7). □

5 Lower Bounds

Our lower bounds consist of the following ingredients. In Section 5.1, we introduce the notion of *semimaxmin chi-square fluctuation* of a family of channels $\mathcal{W}$, which will be central to our results. In Theorem 5.5 we provide an upper bound on the semimaxmin chi-square fluctuation as a function of $\|H(W)\|_*$. We then, in the corresponding following sections, use Lemma 5.4, in conjunction with the bounds on $\|H(W)\|_*$ for communication-constrained and locally private channels, to prove our lower bounds in those two settings and establish the lower bound part of Theorems 1.1 and 1.2.

As we aim to prove a lower bound on the sample complexity of identity testing (for general reference distribution $\mathbf{q}$), it is enough to show a lower bound on its special case of *uniformity* testing. This is a sensible choice, as the uniform distribution $\mathbf{u}_k$ is the "hardest" instance of identity testing (see *e.g.*, [Pan08, Gol16]).

5.1 The General Formulation: Semimaxmin decoupled chi-square fluctuation

We build on the notions of maxmin and minmax *decoupled chi-square fluctuations*, introduced in [ACT18] to prove lower bounds on the sample complexity of SMP protocols with and without public randomness, respectively. The maxmin fluctuation results in a bottleneck for private-coin protocols and the minmax for public-coin protocols. To obtain our lower bounds, we generalize these and define the notion of *semimaxmin* decoupled chi-square fluctuation, which interpolates between the maxmin and minmax fluctuations and captures the setting of *limited* public randomness.

In order to do so, we first recall the definition of perturbations around a fixed distribution $\mathbf{q} \in \Delta(k)$.

Definition 5.1 ([ACT18, Definition IV.4])**.** Consider $0 < \varepsilon < 1$, a family of distributions $\mathcal{P} = \{\mathbf{p}_z, z \in \mathcal{Z}\}$, and a distribution ζ on $\mathcal{Z}$. The pair $\mathcal{P}_\zeta = (\mathcal{P}, \zeta)$ is an *almost ε-perturbation (around $\mathbf{q}$)* if

$$\Pr[\, \mathrm{d}_{\mathrm{TV}}(\mathbf{p}_Z, \mathbf{q}) \geq \varepsilon \,] \geq \alpha,$$

for some $\alpha \geq 1/10$. We denote the set of all almost ε-perturbations by Υ_ε. Moreover, for $\alpha = 1$ we refer to $\mathcal{P}$ as a *perturbed family*.

For a channel $W\colon [k] \to \mathcal{Y}$, $z \in \mathcal{Z}$, and a symbol $y \in \mathcal{Y}$, we denote by $\mathbf{q}^W$ the distribution on $\mathcal{Y}$ induced by $\mathbf{q}$ and W (so that $\mathbf{q}^W(y) = \sum_{x\in[k]} W(y \mid x)\mathbf{q}(x)$), and let $\delta_z^W(y) := (\mathbf{p}_z^W(y) - \mathbf{q}^W(y))/\mathbf{q}^W(y)$. Also, for a family of channels $\mathcal{W}$, denote by $\overline{\mathcal{W}}$ its convex hull. We now recall the definition of decoupled chi-square fluctuation, and provide an operational meaning for it.

Definition 5.2 ([ACT18, Definition IV.3])**.** Consider a perturbed family $\mathcal{P} = \{\mathbf{p}_z : z \in \mathcal{Z}\}$ and a family of channels $\mathcal{W}$. The n-fold *induced decoupled chi-square fluctuation* of $\mathcal{P}$ for $W^n \in \mathcal{W}^n$ is given by

$$\chi^{(2)}\left(W^n \mid \mathcal{P}\right) := \ln \mathbb{E}_{ZZ'}\left[\exp\left(\sum_{i=1}^n \left\langle \delta_Z^{W_i}, \delta_{Z'}^{W_i} \right\rangle\right)\right],$$

where $\left\langle \delta_z^W, \delta_{z'}^W \right\rangle = \mathbb{E}_{Y\sim \mathbf{p}^W}\left[\delta_z^W(Y)\delta_{z'}^W(Y)\right]$.

It was shown in previous work that $\chi^{(2)}\left(W^n \mid \mathcal{P}\right)$ is an upper bound on the chi-square distance over the n channel output distributions induced by the almost ε-perturbation, and $\mathbf{q}$; in particular, for any testing protocol to be successful, this quantity must be bounded away from zero. After these definitions, we are now in position to introduce the main tool underlying our randomness tradeoff lower bound, the new notion of *semimaxmin* fluctuation:

Definition 5.3 (Semimaxmin Chi-square Fluctuation)**.** For a family of channels $\mathcal{W}$ and $s \in \mathbb{N}$, the (n,ε,s)*-semimaxmin decoupled chi-square fluctuation for* $\mathcal{W}$ is given by

$$\underline{\overline{\chi}}^{(2)}(\mathcal{W}^n,\varepsilon,s) := \sup_{\substack{\mathcal{W}_s\subseteq\overline{\mathcal{W}}^n \\ |\mathcal{W}_s|\leq 2^s}} \inf_{\mathcal{P}_\zeta\in\Upsilon_\varepsilon} \mathbb{E}\left[\chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right)\wedge 1\right],$$

where the supremum is over all multisets $\mathcal{W}_s$ of $\overline{\mathcal{W}}^n$ of size at most 2^s, the infimum is over all almost ε-perturbations $\mathcal{P}_\zeta$, and the expectation over the uniform choice of W^n from $\mathcal{W}_s$.

One may observe that when $s=0$ and $s=\infty$, respectively, replacing the expectation by a supremum yields the maxmin and minmax formulations from previous work. Here, we consider instead an inner expectation, as it makes it easier to bound the resulting quantity in practice – while making the proof of Lemma 5.4 only slightly more technical. Note that in the definition we take a supremum over 2^s choices of W^n to capture the fact that there are s public bits which determine the distribution over the channels. If only s bits of public randomness are available, we will show that any test using channels from $\mathcal{W}$ will err with large constant probability if the above quantity $\underline{\overline{\chi}}^{(2)}(\mathcal{W}^n,\varepsilon,s)$ is upper bounded by a sufficiently small constant.

Lemma 5.4 (Semimaxmin decoupled chi-square fluctuation bound for testing)**.** *For $0<\varepsilon<1$, $s\in\mathbb{N}$, and a k-ary reference distribution $\mathbf{p}$, the sample complexity $n=n(k,\varepsilon,s)$ of (k,ε)-identity testing with s bits of public randomness using $\mathcal{W}$ must satisfy*

$$\underline{\overline{\chi}}^{(2)}(\mathcal{W}^n,\varepsilon,s) \geq c\,, \tag{8}$$

for some constant $c>0$ depending only on the probability of error.

Proof. The proof uses Le Cam's two-point method. Consider an almost ε-perturbation $\mathcal{P}_\zeta$: we note first that, due the use of private coins, the effective channel used by each user is a convex combination of channels from $\mathcal{W}$, namely it is a channel from $\overline{\mathcal{W}}$. Thus, when X^n has distribution either $\mathbf{p}^n$ and $\mathbf{p}_z^n$, respectively, Y^n has distribution $\mathbf{p}^{W^n}$ and $\mathbf{p}_z^{W^n}$ with $W^n\in\overline{\mathcal{W}}^n$. The public randomness then allow the users to jointly sample from any distribution on $\overline{\mathcal{W}}^n$ which can be sampled by s independent unbiased bits, that is from any uniform distribution on a multiset $\mathcal{W}_s\subseteq\overline{\mathcal{W}}^n$ of size (at most) 2^s.

Now, for every choice of channels $W^n=(W_1,\ldots,W_n)\in\overline{\mathcal{W}}^n$, by Pinsker's inequality and the concavity of logarithm,

$$\mathrm{d_{TV}}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right],\mathbf{p}^{W^n}\right)^2 \leq \frac{1}{2}\mathrm{KL}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right]\,||\,\mathbf{p}^{W^n}\right) \leq \frac{1}{2}\ln\left(1+\mathrm{d}_{\chi^2}\left(\mathbb{E}\left[\mathbf{p}_Z^{W^n}\right]\,||\,\mathbf{p}^{W^n}\right)\right).$$

Also, we have the trivial bound $\mathrm{d_{TV}}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W^n}\Big], \mathbf{p}^{W^n}\Big)^2 \le 1$. Fix any multiset $\mathcal{W}_s \subseteq \overline{\mathcal{W}}^n$. Over the uniformly random choice of $W_U^n \in \mathcal{W}_s$ (using the public randomness U), we then have using the concavity of square roots,

$$\begin{aligned}\mathbb{E}_U\Big[\mathrm{d_{TV}}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big], \mathbf{p}^{W_U^n}\Big)\Big]^2 &\le \mathbb{E}_U\left[1 \wedge \sqrt{\frac{1}{2}\ln\Big(1+\mathrm{d}_{\chi^2}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big] \,||\, \mathbf{p}^{W_U^n}\Big)\Big)}\right]^2 \\ &\le \mathbb{E}_U\left[1 \wedge \frac{1}{2}\ln\Big(1+\mathrm{d}_{\chi^2}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big] \,||\, \mathbf{p}^{W_U^n}\Big)\Big)\right].\end{aligned}$$

We then bound the right-side further using [ACT18, Lemma III.V] with θ replaced by z, $Q_\vartheta^n = \mathbf{p}_z^{W_U^n}$ and $P_i = \mathbf{p}^{W_U^n}$ to get

$$\begin{aligned}\mathbb{E}_U\Big[1 \wedge \ln\Big(1+\mathrm{d}_{\chi^2}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big] \,||\, \mathbf{p}^{W_U^n}\Big)\Big)\Big] &\le \mathbb{E}_U\left[1 \wedge \ln \mathbb{E}_{ZZ'}\left[\prod_{i=1}^n (1+H_i^U(Z,Z'))\right]\right] \\ &\le \mathbb{E}_U\Big[1 \wedge \ln \mathbb{E}_{ZZ'}\Big[e^{\sum_{i=1}^n H_i^U(Z,Z')}\Big]\Big] \\ &= \mathbb{E}_U\Big[1 \wedge \chi^{(2)}\left(W_U^n \mid \mathcal{P}_\zeta\right)\Big],\end{aligned}$$

since $H_i^U(Z,Z') = \left\langle \delta_Z^{W_{U,i}}, \delta_{Z'}^{W_{U,i}} \right\rangle$. That is, we have[8]

$$\mathbb{E}_U\Big[\mathrm{d_{TV}}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big], \mathbf{p}^{W_U^n}\Big)\Big]^2 \le \mathbb{E}_U\Big[1 \wedge \chi^{(2)}\left(W_U^n \mid \mathcal{P}_\zeta\right)\Big]. \tag{9}$$

Consider an (n,ε)-test $\mathcal{T}$ using a public-coin protocol. Denote by U the public randomness and by $Y_1,\dots,Y_n$ the messages from each user and by $\mathcal{Z}_0$ the set of z such that $\mathrm{d_{TV}}(\mathbf{p}_z, \mathbf{p}) \ge \varepsilon$. Since $\mathcal{P}_\zeta$ is an almost ε-perturbation, $\Pr[\,Z \in \mathcal{Z}_0\,] \ge \alpha \ge 1/10$. Also, for the test $\mathcal{T}$ we have $\Pr_{X^n\sim\mathbf{p}^n}[\,\mathcal{T}(U,Y^n)=1\,] \ge 11/12$ and $\Pr_{X^n\sim\mathbf{p}_z^n}[\,\mathcal{T}(U,Y^n)=0\,] \ge 11/12$ for every $z \in \mathcal{Z}_0$. Thus, we obtain

$$\frac{1}{2}\Pr_{X^n\sim\mathbf{p}^n}[\,\mathcal{T}(U,Y^n)=1\,] + \frac{1}{2}\Pr_{X^n\sim\mathbb{E}[\mathbf{p}_Z^n]}[\,\mathcal{T}(U,Y^n)=0\,] \ge \frac{11(1+\alpha)}{24} \ge \frac{121}{240},$$

where the last inequality relies on the fact that $\alpha \ge 1/10$. Equivalently,

$$\frac{1}{2}\Pr_{X^n\sim\mathbf{p}^n}[\,\mathcal{T}(U,Y^n)\ne 1\,] + \frac{1}{2}\Pr_{X^n\sim\mathbb{E}[\mathbf{p}_Z^n]}[\,\mathcal{T}(U,Y^n)\ne 0\,] \le \frac{119}{240}. \tag{10}$$

An important remark here is that the distribution of W_U^n (that is, the choice of $\mathcal{W}_s \subseteq \overline{\mathcal{W}}^n$) does not depend on $\mathcal{P}_\zeta$. The left-hand-side of Eq. (10) above coincides with the Bayes error for test $\mathcal{T}$ for the simple binary hypothesis testing problem of $\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big]$ versus $\mathbb{E}_U\Big[\mathbf{p}^{W_U^n}\Big]$, which must be at least

$$\frac{1}{2}\left(1 - \mathbb{E}_U\Big[\mathrm{d_{TV}}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big], \mathbf{p}^{W_U^n}\Big)\Big]\right).$$

Thus, we can find $\mathcal{W}_s$ such that for W_U^n distributed uniformly on $\mathcal{W}_s$ and any almost ε-perturbations $\mathcal{P}_\zeta$

$$\mathbb{E}_U\Big[\mathrm{d_{TV}}\Big(\mathbb{E}\Big[\mathbf{p}_Z^{W_U^n}\Big], \mathbf{p}^{W_U^n}\Big)\Big] \ge \frac{1}{120},$$

[8] Dropping the constant $1/2$ for simplicity of the resulting bound.

which along with Eq. (9) yields

$$\mathbb{E}_U\left[1 \wedge \chi^{(2)}\left(W_U^n \mid \mathcal{P}_\zeta\right)\right] \geq c, \tag{11}$$

where $c = 1/14400$. The result follows upon taking minimum over all almost ε-perturbations $\mathcal{P}_\zeta$ and the maximum over all multisets $\mathcal{W}_s \in \overline{\mathcal{W}}^n$ of size at most 2^s. □

In view of Lemma 5.4, it then suffices to come up with a particular reference distribution $\mathbf{q}$ of our choosing, and, for any type of constraint $\mathcal{W}$, to upper bound $\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon, s)$ as a function of k, ε, s and (some quantity of) $\mathcal{W}$. To do so, recalling the definition of semimaxmin decoupled chi-square fluctuation (Definition 5.3), it suffices to do the following: for each fixed $\mathcal{W}_s \subseteq \overline{\mathcal{W}}^n$ of size at most 2^s, construct an almost ε-perturbation $\mathcal{P}_\zeta = (\mathcal{P}, \zeta)$ around our $\mathbf{q}$ such that $\mathbb{E}\left[\chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right)\right]$ is small enough. As previously mentioned, we will choose our reference distribution $\mathbf{q}$ to be the uniform distribution $\mathbf{u}_k$. Our almost perturbations will consist of "small local perturbations" around uniform, and be of the form

$$\mathbf{p}_Z = \frac{1}{k}(1 + Z_1\varepsilon, 1 - Z_1\varepsilon, \ldots, 1 + Z_{k/2}\varepsilon, 1 - Z_{k/2}\varepsilon)\,, \tag{12}$$

where Z is drawn for a suitably chosen distribution ζ on $\mathbb{R}^{k/2}$. Note that taking ζ to be uniform on $\{-1, 1\}^{k/2}$, we retrieve the "Paninski construction" [Pan08], widely used to prove lower bounds in the centralized, unconstrained setting. Unfolding the definition of decoupled chi-square perturbation, the form chosen in (12) for our perturbation then naturally leads to the following channel-dependent matrix $H(W)$, which will guide the choice of the "worst possible mixture ζ over $\mathcal{Z}$" for a given family of channels. For each channel $W \in \mathcal{W}$, let the $(k/2)$-by$(k/2)$ positive semidefinite matrix $H(W)$ be defined as

$$H(W)_{i_1,i_2} := \sum_{y \in \mathcal{Y}} \frac{(W(y \mid 2i_1 - 1) - W(y \mid 2i_1))(W(y \mid 2i_2 - 1) - W(y \mid 2i_2))}{\sum_{x \in [k]} W(y \mid x)}, \quad i_1, i_2 \in [k/2]\,. \tag{13}$$

This matrix will, loosely speaking, capture the ability of channel W to discriminate between even and odd inputs, and thus to distinguish the reference uniform distribution from such a mixture of perturbed distributions. Our bounds will rely on the nuclear norm $\|H(W)\|_*$ of the matrix $H(W)$. In effect, our results characterize the *informativeness* of a channel W for testing in terms of the nuclear norm of $H(W)$. Channels with larger nuclear norms provide more information, and the channel constraints impose a bound on the nuclear norms, which leads to our result:

Theorem 5.5. *Given $n \in \mathbb{N}$, $\varepsilon \in (0, 1)$, $s \in \mathbb{N}$, for a channel family $\mathcal{W}$ the (n, ε, s)-semimaxmin chi-square fluctuation is bounded as*

$$\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon, s) = O\left(\frac{n^2\varepsilon^4}{k^3} \cdot 2^s \cdot \max_{W \in \mathcal{W}} \|H(W)\|_*^2\right),$$

whenever

$$n \leq \frac{k^{3/2}}{C\varepsilon^2 2^{s/2} \max_{W \in \mathcal{W}} \|H(W)\|_*}, \tag{14}$$

where $C > 0$ is a constant.

The proof of this theorem is quite technical, and is provided in Appendix D. We here give an outline of the argument.

Proof of Theorem 5.5 (Sketch). In view of the discussion above, we would like, given any multiset $\mathcal{W}_s$ of 2^s n-fold channels W^n, to design a suitable distribution for our perturbation Z which "fools" all (or most) of the 2^s channels. Loosely speaking, we would like to construct a distribution for which (informally) most of variance falls in subspaces corresponding to small eigenvectors for a large fraction of the matrices $H(W_i)$. To do so, we proceed along the same lines as the proof of [ACT18, Theorem IV.18] (hereafter denoted $(\star)$), reducing the problem to finding a distribution of the perturbation vector Z such that, for any fixed (multi)set $\mathcal{W}_s \subseteq \mathcal{W}^n$ of size at most 2^s, the expectation

$$\mathbb{E}_{W^n}\left[\ln \mathbb{E}_{ZZ'}\left[e^{\frac{\beta^2 n^2 \varepsilon^2}{k} Z^\intercal \bar{H}(W^n) Z'}\right]\right]$$

(where $\beta > 0$ is a constant, and $\bar{H}(W^n) := \frac{1}{n}\sum_{i=1}^n H(W_i)$), is small. Using a similar argument, it suffices to find a matrix V such that (i) $\|V\|_F^2 \gtrsim k$, (ii) each row of V has 2-norm at most 1, and (iii) the average (over $W^n \in \mathcal{W}_s$) Frobenius norm $\mathbb{E}_{W^n}\left[\|V^\intercal \overline{H}(W^n) V\|_F\right]$ is small.

Since all the matrices $\bar{H}(W^n)$ (and therefore all $V^\intercal \bar{H}(W^n) V$'s) are symmetric positive semi-definite matrices, one can then show that

$$\frac{1}{2^s}\|V^\intercal\left(\sum_{W^n \in \mathcal{W}_s} \bar{H}(W^n)\right) V\|_F^2 \geq \mathbb{E}_{W^n}\left[\|V^\intercal \bar{H}(W^n) V\|_F\right]^2. \tag{15}$$

Using a construction from $(\star)$ applied to $\tilde{H}(\mathcal{W}_s) := \sum_{W^n \in \mathcal{W}_s} \bar{H}(W^n)$, we obtain a matrix V satisfying the above conditions (i) and (ii), and such that we have the following analogue of (iii):

$$\|V^\intercal \tilde{H}(\mathcal{W}_s) V\|_F^2 \lesssim \frac{1}{k}\|\tilde{H}(\mathcal{W}_s)\|_*^2. \tag{16}$$

Combining this inequality with (15) and the triangle inequality, this leads to

$$\mathbb{E}_{W^n}\left[\|V^\intercal \bar{H}(W^n) V\|_F\right]^2 \lesssim \frac{2^s}{k} \max_{W \in \mathcal{W}} \|H(W)\|_*^2. \tag{17}$$

From (17), we can finally derive the desired bound in a manner analogous to the end of $(\star)$. This is however not entirely immediate, as (by our very construction), we can only guarantee small Frobenius norms and spectral radius *on average* for the $V^\intercal \bar{H}(W^n) V$'s. The original argument of $(\star)$, however, crucially requires during its last step a pointwise guarantee; to conclude, we thus must resort to a careful averaging argument over these spectral radii to ensure *most* of them are under control, and handle the small remaining "bad" fraction separately. More specifically, this last part hinges on the inner $\min$ in the definition of semimaxmin fluctuation: when bounding the quantity $\mathbb{E}_{W^n}[\chi^{(2)}(W^n \mid \mathcal{P}_\zeta) \wedge 1]$ in the end, we control the pointwise contribution of the "good" W^n's via the term $\chi^{(2)}(W^n \mid \mathcal{P}_\zeta)$ (which we show is then $\ll 1$), and the contribution of the "bad" W^n's via the term 1 (which, while large, is weighted by the fraction of "bad" channels, which is itself small enough). □

5.2 Communication-Constrained and LDP Testing

We now instantiate the general lower bound result established in the previous section to the two specific settings we consider, communication and local privacy constraints. For communication-constrained and LDP channels the nuclear norms of the H matrices can be uniformly bounded as follows.

Lemma 5.6 ([ACT18, Lemmas V.1 and V.5])**.** *For $\ell \geq 1$, and $\varrho \in (0,1]$, $\max_{W \in \mathcal{W}_\ell} \|H(W)\|_* \leq 2^\ell$ and $\max_{W \in \mathcal{W}_\varrho} \|H(W)\|_* = O(\varrho^2)$.*

Using these bounds, we readily obtain our sample complexity results for both communication-constrained and LDP channels.

Theorem 5.7. *For $0 < \varepsilon < 1$ and $\ell, s \in \mathbb{N}$, the sample complexity of (k,ε)-uniformity testing with s bits of public randomness using $\mathcal{W}_\ell$ is at least*

$$\Omega\left(\frac{\sqrt{k}}{\varepsilon^2}\sqrt{\frac{k}{2^\ell} \vee 1}\sqrt{\frac{k}{2^{s+\ell}} \vee 1}\right).$$

Theorem 5.8. *For $0 < \varrho < 1$, and $s \in \mathbb{N}$ the sample complexity of (k,ε)-uniformity testing with s bits of public randomness using $\mathcal{W}_\varrho$ is at least*

$$\Omega\left(\frac{\sqrt{k}}{\varepsilon^2}\frac{\sqrt{k}}{\varrho^2}\sqrt{\frac{k}{2^s} \vee 1}\right).$$

Indeed, from Lemma 5.4, we get that $\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon, s)$ must be lower bounded by a constant for n samples to be sufficient for testing. Plugging in the bounds from Lemma 5.6 in Theorem 5.5 yields the two above results.

References

[ACFT19] Jayadev Acharya, Clément L Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 2067–2076. PMLR, 16–18 Apr 2019. 1.1, 1.3, 4.3, 4.3, 4.3.1, 4.3.1

[ACT18] Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints I: lower bounds from chi-square contraction. *CoRR*, abs/1812.11476, 2018. In submission. Full version of [ACT19b]. 1, 1.1, 1.1, 1.2, 3, 5.1, 5.1, 5.2, 5.1, 5.1, 5.6, D, D.2, 9, D.3

[ACT19a] Jayadev Acharya, Clément L Canonne, and Himanshu Tyagi. Inference under information constraints II: communication constraints and shared randomness. *CoRR*, abs/1905.08302, 2019. In submission. Full version of [ACT19c]. 1, 3, 4.3, 4.2, 4.5

[ACT19b] Jayadev Acharya, Clément L Canonne, and Himanshu Tyagi. Inference under information constraints: Lower bounds from chi-square contraction. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 3–17, Phoenix, USA, 25–28 Jun 2019. PMLR. 1.3, 5.2

[ACT19c] Jayadev Acharya, Clément L Canonne, Canonne, and Himanshu Tyagi. Communication-constrained inference and the role of shared randomness. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 30–39, Long Beach, California, USA, 09–15 Jun 2019. PMLR. 1.3, 5.2

[ADK15] Jayadev Acharya, Constantinos Daskalakis, and Gautam C. Kamath. Optimal Testing for Properties of Distributions. In C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3577–3598. Curran Associates, Inc., 2015. 1, 1.3

[AS19] Jayadev Acharya and Ziteng Sun. Communication complexity in locally private distribution estimation and heavy hitters. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 51–60, Long Beach, California, USA, 09–15 Jun 2019. PMLR. 1.3, 4.3, 4.3.1, 4.3.1

[ASZ19] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1120–1129. PMLR, 16–18 Apr 2019. 1.3

[BFR+00] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000*, pages 189–197, 2000. 5.2

[BFR+13] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4:1–4:25, 2013. This is a long version of [BFR+00]. 1, 1.3

[BHÖ19] Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Learning distributions from their samples under communication constraints. *CoRR*, abs/1902.02890, 2019. 1.3

[BW18] Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for high-dimensional multinomials: A selective review. *The Annals of Applied Statistics*, 12(2):727–749, 2018. 1.3

[Can15] Clément L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, 2015. 1.3

[CDKS17] Clément L. Canonne, Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Testing bayesian networks. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 370–448, Amsterdam, Netherlands, 07–10 Jul 2017. PMLR. 7

[CG89] Benny Chor and Oded Goldreich. On the power of two-point based sampling. *J. Complexity*, 5(1):96–106, 1989. C

[CKM+19] Clément L. Canonne, Gautam Kamath, Audra McMillan, Jonathan Ullman, and Lydia Zakynthinou. Private identity testing for high-dimensional distributions. *CoRR*, abs/1905.11947, 2019. 7

[CW89] Aviad Cohen and Avi Wigderson. Dispersers, deterministic amplification, and weak random sources (extended abstract). In *FOCS*, pages 14–19. IEEE Computer Society, 1989. C

[DGKR19] Ilias Diakonikolas, Themis Gouleakis, Daniel M. Kane, and Sankeerth Rao. Communication and memory efficient testing of discrete distributions. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1070–1106, Phoenix, USA, 25–28 Jun 2019. PMLR. 1.3

[DGPP18] Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. Sample-optimal identity testing with high probability. In *Proceedings of ICALP*, volume 107 of *LIPIcs*, pages 41:1–41:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. 1, 1.3

[DJW17] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 2017. 1.3

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, 2014. ACM. 1.3

[FMO18] Orr Fischer, Uri Meir, and Rotem Oshman. Distributed uniformity testing. In Calvin Newport and Idit Keidar, editors, *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018, Egham, United Kingdom, July 23-27, 2018*, pages 455–464. ACM, 2018. 1.3

[Gol16] Oded Goldreich. The uniform distribution is complete with respect to testing identity to a fixed distribution. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:15, 2016. 1.3, 5

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. 1.3

[GR00] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity (ECCC), 2000. 1.3

[GR18] Marco Gaboardi and Ryan Rogers. Local private hypothesis testing: Chi-square tests. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1626–1635, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. 1.3

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006. C, C.3

[HMÖW18a] Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman. Distributed statistical estimation of high-dimensional and non-parametric distributions. In *Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT'18)*, pages 506–510, 2018. 1.3

[HMÖW18b] Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman. Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints, February 2018. Talk given at ITA 2018. 1.3

[HÖW18] Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Proceedings of the 31st Conference on Learning Theory, COLT 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 3163–3188. PMLR, 2018. 1.3

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *STOC*, pages 604–613. ACM, 1998. 1.3

[JLS86] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, Jun 1986. 1.3

[KBR16] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2436–2444. JMLR.org, 2016. 1.3

[KPS85] Richard Karp, Nicholas Pippenger, and Michael Sipser. A time-randomness tradeoff, 1985. Oral presentation given at the *AMS Conference on Probabilistic Computational Complexity*. 1.2, 4.1, C

[KPV10] Rasmus J Kyng, Jeff M Phillips, and Suresh Venkatasubramanian. Johnson-Lindenstrauss dimensionality reduction on the simplex. In *20th Annual Fall Workshop on Computational Geometry*, 2010. 1.3

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008. 1, 1.3, 3, 5, 5.1

[Pea00] Karl Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900. 1.3

[Pis89] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989. A

[Rub12] Ronitt Rubinfeld. Taming big probability distributions. *XRDS: Crossroads, The ACM Magazine for Students*, 19(1):24, sep 2012. 1.3

[She18] Or Sheffet. Locally private hypothesis testing. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning (ICML)*, volume 80 of *Proceedings of Machine Learning Research*, pages 4612–4621, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. 1.3

[VV14] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *55th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2014*, 2014. 5.2

[VV17] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017. Journal version of [VV14]. 1, 1.3

[YB18] Min Ye and Alexander Barg. Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676, 2018. 1.3

A Spectrum of outer products result

In this appendix we prove Theorem 3.5 and Lemma 3.4. Theorem 3.5 is restated below:

Theorem A.1 (Spectrum of outer products). *For $n \in \mathbb{N}$, there exist constants $c_0 \in \mathbb{N}$, $c_1, c_2 \in (0,1)$ and vectors $u_1, \ldots, u_{m_0} \in \{0,1\}^n$ with $m_0 = 2^{c_0} n$ such that for every $\mathcal{J} \subseteq [m_0]$ with $|\mathcal{J}| \geq (1-c_1)m_0$ we must have*

$$\lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} u_j u_j^{\intercal}\right) \geq c_2.$$

Consider random, independent binary vectors $V_1, \ldots, V_{m_0} \in \{0,1\}^n$, with each V_i drawn uniformly from the set of all binary vectors of length n. We establish Theorem A.1 using probablistic argument. It would be enough to show that:

$$\Pr\left[\exists\, \mathcal{J} \subseteq [m_0], |\mathcal{J}| \geq (1-\theta)m_0 \text{ s.t. } \lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{i\in\mathcal{J}} V_i V_i^{\intercal}\right) < c_2\right] < 1.$$

First, for any $\mathcal{J}$ with $|\mathcal{J}| = m \geq (1-\theta)m_0$, we will derive an exponential upper bounds for the probability,

$$\Pr\left[\lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{i\in\mathcal{J}} V_i V_i^{\intercal}\right) < t\right].$$

Without loss of generality, we can assume $\mathcal{J} = [m]$. Since

$$\lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{i\in\mathcal{J}} V_i V_i^{\intercal}\right) = \min_x \{\frac{\frac{1}{m}\sum_{j=1}^m x^{\intercal} V_j V_j^{\intercal} x}{\|x\|_2^2}\},$$

we first establish an exponential upper bound for

$$\Pr\left[\frac{1}{m}\sum_{j=1}^m x^{\intercal} V_j V_j^{\intercal} x < t\|x\|_2^2\right].$$

We derive this bound using a general anti-concentration bound for subgaussian random variables, which may be of independent interest.

Theorem A.2 (An anti-concentration bound). *Consider independent random variables $Y_1, \ldots, Y_m$ such that each Y_i is zero-mean and subgaussian with variance parameter σ^2, i.e., $\mathbb{E}\left[e^{\lambda Y_i}\right] \leq e^{\lambda^2\sigma^2/2}$ for all $\lambda \in \mathbb{R}$. Suppose further that, for all i, $\mathbb{E}[Y_i^2] \geq \eta\sigma^2$ for some $\eta \in (0,1)$. Then, there exist positive constants c_1 and c_2 such that for every $\mu \in \mathbb{R}$,*

$$\Pr\left[\frac{1}{m}\sum_{i=1}^m (Y_i+\mu)^2 \geq c_1\eta^2\Big(\min_{1\leq i\leq m}\mathbb{E}\left[Y_i^2\right] + \mu^2\Big)\right] \geq 1 - \exp(-c_2 m\eta^4).$$

To prove this result, we take recourse to the following "clipped-tail" version of Hoeffding bound, which allows us to obtain exponential anti-concentration bounds using anti-concentration bounds.

Lemma A.3 (Clipped-tail Hoeffding bound). *For $t > 0$, let $X_1, \ldots, X_m$ be nonnegative, independent random variables satisfying*

$$\Pr[\,X_i \geq t\,] \geq \alpha, \qquad 1 \leq i \leq m,$$

Then,

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} X_i \geq \frac{t\alpha}{2}\right] \geq 1 - \exp\left(-m\frac{\alpha^2}{2}\right).$$

Proof. Since X_is are nonnegative, $\sum_{i=1}^m X_i \geq t\sum_{i=1}^m \mathbb{1}_{\{X_i > t\}}$. It follows that

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m} X_i \leq \frac{t\alpha}{2}\right] = \Pr\left[\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{X_i > t\}} \leq \frac{\alpha}{2}\right],$$

where the right-side is bounded above further by $\exp(-m\alpha^2/2)$ using Hoeffding's inequality and the assumption of the lemma. □

We use this bound to now complete the proof of Theorem A.2.

Proof of Theorem A.2. Let Y be zero-mean and subgaussian with variance parameter σ^2. Then, for $X = Y + \mu$, we get $\mathbb{E}[X^4] \leq 8\mathbb{E}[Y^4] + 8\mu^4$. Also, since Y is subgaussian with variance parameter σ^2, it is easy to show that $\mathbb{E}[Y^4] \leq 8\sigma^2$, whereby we get $\mathbb{E}[X^4] \leq 64\sigma^4 + 8\mu^4$. Since by our assumption $\mathbb{E}[X^2] = \mathbb{E}[Y^2] + \mu^2 \geq \eta\sigma^2 + \mu^2$, it follows that $\mathbb{E}[X^2]^2 \geq \eta^2\sigma^4 + \mu^4$. Upon combining this with the previous bound, we obtain $\mathbb{E}[X^4] \leq \frac{64}{\eta^2}\mathbb{E}[X^2]^2$. We now invoke the Paley–Zygmund inequality to get

$$\Pr\left[X^2 \geq \frac{1}{2}\left(\mathbb{E}\left[Y^2\right] + \mu^2\right)\right] \geq \frac{\eta^2}{256}.$$

Finally, an application of Lemma A.3 yields

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m}(Y_i + \mu)^2 \geq \frac{\eta^2}{1024}\left(\min_{1\leq i\leq m}\mathbb{E}\left[Y_i^2\right] + \mu^2\right)\right] \geq 1 - \exp\left(-\frac{m\eta^4}{256^2}\right),$$

which completes the proof. □

Proof of Theorem A.1. Let $\mathbf{1}$ be the all one vector in R^n. We apply Theorem A.2 to $Y_i = x^\intercal V_i - (\mathbf{1}^\intercal x)/2$, $1 \leq i \leq m$, with $\mu = (\mathbf{1}^\intercal x)/2$. Note that the Y_i's are zero-mean, and by Hoeffding's lemma, they are subgaussian with variance parameter $\|x\|_2^2/4$. Furthermore, it is easy to verify that $\mathbb{E}[Y_i^2] = \|x\|_2^2/4$. Thus, the condition of Theorem A.2 holds with $\eta = 1$, which gives

$$\Pr\left[\frac{1}{m}\sum_{i=1}^{m}(x^\intercal V_i)^2 \geq c_1\,\frac{\|x\|_2^2 + (\mathbf{1}^\intercal x)^2}{4}\right] \geq 1 - \exp\left(-c_2 m\right). \tag{18}$$

Denote by A_m the random matrix $\frac{1}{m}\sum_{i=1}^m V_i V_i^\intercal$. Our goal is to bound $\lambda_{\min}(A_m)$. It will be convenient to introduce a new norm $\|\cdot\|_\star$ on $\mathbb{R}^n$: for $x \in \mathbb{R}^n$,

$$\|x\|_\star := \sqrt{\|x\|_2^2 + (\mathbf{1}^\intercal x)^2}.$$

Clearly, $\|\cdot\|_\star$ is a norm, as $\|x\|_\star = \|L(x)\|_2$ where $L(x) := (x_1, \ldots, x_n, \sum_{i=1}^n x_i) \in \mathbb{R}^{n+1}$ is linear.

Now, if we can find an x such that $x^\intercal A_m x < \lambda \|x\|_2^2$, then $y = x/\|x\|_\star$ has $\|y\|_\star = 1$ and satisfies $y^\intercal A_m y < \lambda$. Therefore,

$$\Pr\left[\min_{x:\|x\|_2=1} x^\intercal A_m x < \lambda\right] \leq \Pr\left[\min_{y:\|y\|_\star=1} y^\intercal A_m y < \lambda\right] \tag{19}$$

We use Eq. (18) to obtain this bound, together with an appropriate netting argument. Specifically, let $\mathcal{N}$ be a δ-net of the sphere $\{\, y \in \mathbb{R}^n : \|y\|_\star = 1 \,\}$ in the norm $\|\cdot\|_\star$. We can find such a net with $|\mathcal{N}| \leq (1+\frac{2}{\delta})^n \leq e^{2n/\delta}$ (see, *e.g.*, [Pis89, Lemma 4.16]), which is the net we use. By a union bound applied to Eq. (18), we get

$$\Pr\left[\min_{x\in\mathcal{N}} x^\intercal A_m x < c_1 \frac{\|x\|_2^2 + (\mathbf{1}^\intercal x)^2}{4}\right] < \exp\left(\frac{2n}{\delta} - c_2 m\right). \tag{20}$$

We bound $y^\intercal A_m y$ for a y with $\|y\|_\star = 1$ by relating it to $x^\intercal A_m x$ for a vector $x \in \mathcal{N}$ such that $\|x-y\|_\star$. While this is he standard netting argument, there is added complication since we need to work with the norm $\|\cdot\|_\star$.

In particular, for a y such that $\|y\|_\star$ consider an $x \in \mathcal{N}$ satisfying $\|x-y\|_\star \leq \delta$. Denoting $z := y - x$, we decompose $z = z_\| + z_\perp$, where $z_\| \in \mathrm{span}_{\mathbb{R}}(\mathbf{1})$, and $z_\perp^\intercal \mathbf{1} = 0$. By definition, $z_\| = \frac{(z^\intercal \mathbf{1})}{n}\mathbf{1}$ and $z_\perp = z - z_\perp$. Using the inequality $(a+b)^2 \geq a^2/2 - b^2$, for every $i \in [m]$ we have

$$(V_i^\intercal y)^2 = (V_i^\intercal x + V_i^\intercal z)^2 \geq \frac{1}{2}(V_i^\intercal x)^2 - (V_i^\intercal z)^2 \geq \frac{1}{2}(V_i^\intercal x)^2 - 2(V_i^\intercal z_\|)^2 - 2(V_i^\intercal z_\perp)^2.$$

Summing over i and using the expression for $z_\|$, we get

$$y^\intercal A_m y \geq \frac{1}{2}\cdot x^\intercal A_m x - \frac{2(z^\intercal \mathbf{1})^2}{n^2}\cdot(\mathbf{1}^\intercal A_m \mathbf{1}) - 2(z_\perp^\intercal A_m z_\perp). \tag{21}$$

To proceed further, we derive bounds for random variables $(\mathbf{1}^\intercal A_m \mathbf{1})$ and $(z_\perp^\intercal A_m z_\perp)$. For the first term, we can show

$$\Pr\left[\,(\mathbf{1}^\intercal A_m \mathbf{1}) > 5n^2\,\right] \leq 2\exp(-m/2). \tag{22}$$

We provide a proof at the end. For the second term, we observe that

$$z_\perp^\intercal A_m z_\perp = \frac{1}{m}\sum_{i=1}^m (z_\perp^\intercal V_i)^2 = \frac{1}{m}\sum_{i=1}^m \left(z_\perp^\intercal\left(V_i - \frac{1}{2}\cdot\mathbf{1}\right)\right)^2.$$

Denote by $\overline{V}_i$ a random variable which takes values $1/2$ and $-1/2$ with equal probabilities, and by $\overline{A}_m$ the random matrix $(1/m)\sum_{i=1}^m \overline{V}_i \overline{V}_i^\intercal$, we get

$$z_\perp^\intercal A_m z_\perp \leq \lambda_{\max}(\overline{A}_m)\|z_\perp\|_2^2.$$

The next result, whose proof is standard and will be given later, provides a bound for $\lambda_{\max}(\overline{A}_m)$.

Lemma A.4. *There exist constants c_2, c_3 such that*

$$\Pr\left[\,\lambda_{\max}(\overline{A}_m) > c_2\,\right] \leq \exp\left(c_3 n - \frac{m}{2}\right).$$

This result, together with Eq. (21) and Eq. (22), yields

$$\begin{aligned}
&\Pr\left[\min_{y:\|y\|_\star=1} y^\mathsf{T} A_m y \geq t\right] \\
&\geq \Pr\left[\min_{x\in\mathcal{N}} x^\mathsf{T} A_m x \geq 2t + 20(z^\mathsf{T}\mathbf{1})^2 + 4c_2\|z_\perp\|_2^2\right] - 2\exp\left(-\frac{m}{2}\right) - \exp\left(c_3 n - \frac{m}{2}\right) \\
&\geq 1 - \Pr\left[\min_{x\in\mathcal{N}} x^\mathsf{T} A_m x \geq 2t + c_4\delta^2\right] - 2\exp\left(-\frac{m}{2}\right) - \exp\left(c_3 n - \frac{m}{2}\right)
\end{aligned}$$

where we used $\|z\|_\star^2 \geq \|z_\perp\|_2^2 + (z^\mathsf{T}\mathbf{1})^2$. We set $t = \delta^2$ and note that for any $x \in \mathcal{N}$ we must have $\|x\|_\star \leq 1+\delta$. Therefore,

$$\Pr\left[\min_{x\in\mathcal{N}} x^\mathsf{T} A_m x < 2t + c_4\delta^2\right] = \Pr\left[\exists\, x \in \mathcal{N} \text{ s.t. } x^\mathsf{T} A_m x < c_5 \frac{\delta^2}{(1+\delta)^2}\|x\|_\star^2\right].$$

Setting δ such that $c_5\delta^2(1+\delta)^2 = c_2/4$, it follows from Eq. (18) that

$$\Pr\left[\min_{x\in\mathcal{N}} x^\mathsf{T} A_m x < 2t + c_4\delta^2\right] \leq |\mathcal{N}|\exp(-c_2 m) \leq \exp\left(\frac{2n}{\delta} - c_2 m\right).$$

Upon combining the bounds above, we get

$$\Pr\left[\min_{y:\|y\|_\star=1} y^\mathsf{T} A_m y < \delta^2\right] \leq \exp\left(\frac{2n}{\delta} - c_2 m\right) + 2\exp\left(-\frac{m}{2}\right) + \exp\left(c_3 n - \frac{m}{2}\right)$$

where δ, c_2, c_3 are constants. Recalling Eq. (19), we have obtained

$$\Pr\left[\lambda_{\min}\left(\frac{1}{m}\sum_{i=1}^m V_i V_i^\mathsf{T}\right) < \delta^2\right] \leq \exp\left(c_6 n - c_7 m\right)$$

Finally, by a union bound of all subsets of $[m_0]$ with size larger than $(1-\theta)m_0$, we get

$$\Pr\left[\exists\, \mathcal{J} \subseteq [m_0], |\mathcal{J}| \geq (1-\theta)m_0 \text{ s.t. } \lambda_{\min}\left(\frac{1}{|\mathcal{J}|}\sum_{i\in\mathcal{J}} V_i V_i^\mathsf{T}\right) < \delta^2\right] \leq m_0 2^{m_0 h(\theta)} \exp\left(c_6 n - c_7 m_0(1-\theta)\right),$$

where $h(\cdot)$ denotes the binary entropy function, and we have used the fact that the number of subsets of $[m_0]$ of cardinality greater than $(1-\theta)m_0$, $\theta \in (0, 1/2)$, is at most $m_0 2^{m_0 h(\theta)}$. The proof is completed by ensuring that the exponent on right-side above is negative. □

It only remains to prove Eq. (22) and Lemma A.4, which we do next.

Proof of Eq. (22). Consider random variables $\xi_i := (V_i^\mathsf{T}\mathbf{1})$, $i \in [m]$. Note that $\mathbb{E}[\xi_i] = n/2$ and each ξ_i is subgaussian with variance parameter $n/4$. Therefore, $\Pr\left[\frac{1}{m}\sum_{i=1}^m \xi_i > n\right] \leq \exp(-mn/2)$. Furthermore, since $\mathbb{E}[(\xi_i - n/2)^2] = n/4$, the random variable $(\xi_i - n/2)^2 - n/4$ is subexponential with parameter $4n$, which gives $\Pr\left[\frac{1}{m}\sum_{i=1}^m (\xi_i - n/2)^2 > 17n/4\right] \leq \exp(-m/2)$. Thus,

$$\begin{aligned}
\Pr\left[\frac{1}{m}\sum_{i=1}^m \xi_i^2 > \frac{3}{4}\cdot n^2 + \frac{17}{4}\cdot n\right] &\leq \Pr\left[\frac{1}{m}\sum_{i=1}^m \xi_i > n\right] + \Pr\left[\frac{1}{m}\sum_{i=1}^m (\xi_i - n/2)^2 > 5n/4\right] \\
&\leq 2\exp\left(-\frac{m}{2}\right),
\end{aligned}$$

which leads to the claimed bound. □

Proof of Lemma A.4. For a fixed $x \in \mathbb{R}^n$, consider random variables $\zeta_i := (\overline{V}^\intercal x)$, $i \in [m]$. They are all zero-mean and are subgaussian with variance parameter $\|x\|_2^2/4$. Furthermore, their second moment $\mathbb{E}[\zeta_i^2]$ equals $\|x\|_2^2/4$. Therefore, the random variable $\zeta_i^2 - \|x\|_2^2/4$ is subexponential with parameter $4\|x\|_2^2$, and we have $\Pr\left[\frac{1}{m}\sum_{i=1}^m \zeta_i^2 > \frac{17}{4}\cdot\|x\|_2^2\right] \le \exp\left(-\frac{m}{2}\right)$.

Next, consider a δ-net $\mathcal{N}_2$ of the unit ball under $\|\cdot\|_2$ of cardinality $|\mathcal{N}_2| \le e^{2n/\delta}$. For a y such that $\|y\|_2 = 1$ and $y^\intercal \overline{A}_m y = \lambda_{\max}(\overline{A}_m)$, consider the $x \in \mathcal{N}_2$ such that $\|y - x\|_2 \le \delta$. Then, since $y^\intercal \overline{A}_m y = x^\intercal \overline{A}_m x + 2(y-x)^T \overline{A}_m y$, we have

$$\lambda_{\max}(\overline{A}_m) = y^\intercal \overline{A}_m y \le x^\intercal \overline{A}_m x + 2\delta\lambda_{\max}(\overline{A}_m),$$

which further gives

$$(1-2\delta)\lambda_{\max}(\overline{A}_m) \le \max_{x\in\mathcal{N}_2} x^\intercal \overline{A}_m x.$$

Also, every $x \in \mathcal{N}_2$ satisfies $\|x\|_2 \le 1+\delta$, a and so, by the tail-probability bound for $\sum_{i=1}^m \zeta_i^2$ that we saw above, we get $\Pr\left[\frac{1}{m}\sum_{i=1}^m \zeta_i^2 > \frac{17(1+\delta)^2}{4}\right] \exp\left(2n/\delta - m/2\right)$. Therefore, we obtain

$$\Pr\left[\lambda_{\max}(\overline{A}_m) > \frac{17(1+\delta)^2}{4(1-2\delta)}\right] \le \exp\left(-\frac{m}{2}\right).$$

In particular, we can set $\delta = 1/4$ to get the claimed result with $c_2 = 425/32$ and $c_3 = 8$. □

We close with a proof of Lemma 3.4, which we recall below for easy reference.

Lemma A.5 (Additivity of tails Lemma 3.4, restated)**.** *Let $a_1, \ldots, a_m \ge 0$, and suppose $Y_1, \ldots, Y_m$ are non-negative random variables with $\Pr[\, Y_i \ge a_i \,] \ge c$ for every $1 \le i \le m$, for some $c \in (0,1)$. Then,*

$$\Pr\left[Y_1 + \ldots + Y_m \ge c \cdot \frac{a_1 + \ldots + a_m}{2}\right] \ge \frac{c}{2-c}.$$

Proof. Let $a_1, \ldots, a_m \ge 0$ and $Y_1, \ldots, Y_m$ be as in the statement, and define $Z_i := a_i \mathbb{1}_{\{Y_i \ge a_i\}}$ for $i \in [m]$. Then $Z_1, \ldots, Z_m$ satisfy the assumptions of the lemma as well, namely $\Pr[\, Z_i \ge a_i \,] \ge c$. Further, $\Pr[\, Y_1 + \ldots + Y_m \ge \alpha \,] \ge \Pr[\, Z_1 + \ldots + Z_m \ge \alpha \,]$ for all α. Thus it suffices to prove the statement for the Z_i's, which are supported on two points, which is what we do.

Let $Z := Z_1 + \ldots + Z_m$. By the assumption, we have $\mathbb{E}[Z] \ge c(a_1 + \ldots + a_m)$, and $0 \le Z \le a_1 + \ldots + a_m$. By Markov's inequality applied to $\sum_{i=1}^m a_i - Z \ge 0$, for $\gamma \in (0,1)$,

$$\Pr\left[Z < \gamma c \sum_{i=1}^m a_i\right] = \Pr\left[\sum_{i=1}^m a_i - Z > (1-\gamma c)\sum_{i=1}^m a_i\right] \le \frac{\sum_{i=1}^m a_i - \mathbb{E}[Z]}{(1-\gamma c)\sum_{i=1}^m a_i} \le \frac{1-c}{1-\gamma c} = 1 - \frac{(1-\gamma)c}{1-\gamma c}.$$

Taking $\gamma := 1/2$ yields the claim. □

B Miscellaneous: some useful lemmas

We provide in this appendix two simple results, mentioned in the preliminaries. We begin with a simple proposition, which allowed us throughout the paper on to assume that one can partition the domain $[k]$ into any number L of equal-sized sets. Indeed, as shown below, when aiming to perform (k,ε)-identity testing this can always be achieved at the cost of only a constant multiplicative factor in the distance parameter ε (and only requires private randomness, as well as knowledge of k and L, from the n users).

Proposition B.1. *Let $k, L \geq 1$ be two integers with $1 \leq L \leq k$, and define $k' := L\lceil k/L \rceil$. There exists an explicit mapping $\Phi_{k,L} \colon \Delta(k) \to \Delta(k')$ such that (i) the uniform distribution is mapped to the uniform distribution, i.e., $\Phi_{k,L}(\mathbf{u}_k) = \mathbf{u}_{k'}$; and (ii) distances are approximately preserved: for every $\mathbf{p}, \mathbf{q} \in \Delta(k)$,*

$$\mathrm{d}_{\mathrm{TV}}(\Phi_{k,L}(\mathbf{p}), \Phi_{k,L}(\mathbf{q})) = \frac{k}{k'}\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \geq \frac{1}{2}\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \,.$$

Further, there exists a randomized mapping $\Psi_{k,L}$ such that, for every $\mathbf{p} \in \Delta(k)$, $\Psi_{k,L}(X) \sim \Phi_{k,L}(\mathbf{p})$ whenever $X \sim \mathbf{p}$.

Proof. We define $\Phi_{k,L}$ as a mixture of the input and the uniform distribution on $[k] \setminus [k']$: for any $\mathbf{p} \in \Delta(k)$, $\Phi_{k,L}(\mathbf{p}) := \frac{k}{k'}\mathbf{p} + \frac{k'-k}{k'}\mathbf{u}_{[k]\setminus[k']}$. Recalling that $k \leq k' < k + L$, is immediate to verify that all the claimed properties hold. □

Applying the above with $L := 2^{\lfloor \log k \rfloor}$, we in particular get the following:

Corollary B.2. *Let $k \geq 1$ be any integer, and define $k' := 2^{\lceil \log k \rceil} \in [k, 2k)$. There exists an explicit mapping $\Phi_k \colon \Delta(k) \to \Delta(k')$ such that, for every $\mathbf{p}, \mathbf{q} \in \Delta(k)$,*

$$\frac{1}{2}\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \leq \mathrm{d}_{\mathrm{TV}}(\Phi_k(\mathbf{p}), \Phi_k(\mathbf{q})) \leq \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \,.$$

Further, there exists a randomized mapping Ψ_k such that, for every $\mathbf{p} \in \Delta(k)$, $\Psi_k(X) \sim \Phi_k(\mathbf{p})$ whenever $X \sim \mathbf{p}$.

In view of this corollary, we without loss of generality can assume throughout that k is a power of two.

C Omitted proof: Deterministic Amplification

In this appendix, we provide for completeness a proof of Lemma 4.2, the "deterministic error reduction" lemma we used in the argument of Theorem 4.1. The idea underlying this deterministic error reduction for RP is well-known, and was introduced by Karp, Pippenger, and Sipser in 1985 [KPS85]. The gist is to see the random string σ as the index of a vertex is a d-regular graph on 2^s vertices, and then run the algorithm on all d neighbors of this random vertex v_r. If the graph is a good enough expander, doing so will ensure not all d neighbors cause the algorithm to err. (For more on deterministic amplification for RP (one-sided) and BPP (two-sided) algorithms, as well as the related notion of exponential error amplification with *few* extra random bits, see, *e.g.*, [CW89, CG89], or [HLW06, Sections 1.3.3 and 3.3]).

We begin by recalling some definitions and a useful lemma. Fix $n, d \in \mathbb{N}$ and $\lambda \geq 0$. We say that a d-regular graph $G = (V, E)$ on n vertices with (normalized) adjacency matrix A has *spectral expansion* λ if $\lambda(G) \leq \lambda$, where $\lambda(G) := \max(|\lambda_2|, |\lambda_n|)$ and $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$ are the eigenvalues of A.

Theorem C.1 (Expander Mixing Lemma)**.** *Let $G = (V, E)$ be a d-regular graph on n vertices with spectral expansion λ. Then, for every $S, T \subseteq V$,*

$$\left|\frac{|e(S,T)|}{dn} - \frac{|S|}{n} \cdot \frac{|T|}{n}\right| \leq \lambda\sqrt{\frac{|S|}{n} \cdot \frac{|T|}{n}} \,,$$

where $e(S,T) = \{\ (u,v) \in E :\ u \in S, v \in T\ \}$.

We are now ready to prove Lemma 4.2, restated below.

Lemma C.2 (Deterministic Amplification for One-Sided Error (RP)). *For any $s \in \mathbb{N}$ and $\eta, \gamma \in (0,1)$, there exist $d = d(\eta,\gamma)$ and (time-efficiently computable) functions $\pi_1, \ldots, \pi_d \colon \{0,1\}^s \to \{0,1\}^s$ such that the following holds. Suppose $\mathcal{X}_0 \subseteq \mathcal{X}$ and $A \colon \mathcal{X} \times \{0,1\}^s \to \Omega$ and $\mathcal{E} \subseteq \Omega$ satisfy*

(i) If $x \in \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, A(x,\sigma) \in \mathcal{E}\,] = 1$ *(Perfect completeness)*

(ii) If $x \notin \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, A(x,\sigma) \notin \mathcal{E}\,] \geq 1 - \eta$ *(Low soundness)*

Then we have

(i) If $x \in \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, \forall i \in [d],\ A(x,\pi_i(\sigma)) \in \mathcal{E}\,] = 1$ *(Perfect completeness)*

(ii) If $x \notin \mathcal{X}_0$, $\Pr_{\sigma \sim \{0,1\}^s}[\, \exists i \in [d],\ A(x,\pi_i(\sigma)) \notin \mathcal{E}\,] \geq 1 - \gamma$ *(High soundness)*

Moreover, on can take $d = \tilde{O}\left(\frac{\eta}{(1-\eta)^2\gamma}\right)$.

Proof. Fix A as in the statement, and let $G = (V,E)$ be a d-regular graph on $n := 2^s$ vertices with spectral expansion $\lambda \leq (1-\eta)\sqrt{\gamma/\eta}$, for some d. We define $\pi_1, \ldots, \pi_d$ as follows: fixing any labeling of the vertices of G, we see $r \in \{0,1\}^s$ as a vertex $v_r \in V$ and let $\pi_1(r), \ldots, \pi_d(r)$ be the labels of the d neighbors of v_r in G.

To see why the claimed properties hold, first note that whenever $x \in \mathcal{X}_0$, then as A has one-sided error we have $A(x,\pi_i(\sigma)) \in \mathcal{E}$ for all i with probability one. To establish the second item, fix $x \notin \mathcal{X}_0$, and define $B_x \subseteq \{0,1\}^s$ as the set of "bad" random seeds, *i.e.*, those on which A errs:

$$B_x := \{\ \sigma \in \{0,1\}^s\ :\ A(x,\sigma) \in \mathcal{E}\ \}\ .$$

By assumption, $|B_x| \leq \eta \cdot 2^s$. Now, consider the set $\tilde{B}_x$ of random seeds for which *all* neighbors are bad seeds, that is those seeds for which $A(x,\pi_i(\sigma))$ fails for all $i \in [d]$:

$$\tilde{B}_x := \{\ \sigma \in \{0,1\}^s\ :\ \forall i \in [d]\,, A(x,\pi_i(\sigma)) \in \mathcal{E}\ \}\ .$$

Since every $\sigma \in \tilde{B}_x$ has d "bad" neighbors, we must have $|e(\tilde{B}_x, B_x)| \geq d|\tilde{B}_x|$. Applying the Expander Mixing Lemma (Theorem C.1), we get

$$\frac{|e(\tilde{B}_x, B_x)|}{dn} \leq \frac{|B_x|}{n} \cdot \frac{|\tilde{B}_x|}{n} + \lambda\sqrt{\frac{|B_x|}{n} \cdot \frac{|\tilde{B}_x|}{n}}$$

which implies, recalling the above bounds on both $|\tilde{B}_x|$ and $|B_x|$, $\frac{d|\tilde{B}_x|}{dn} \leq \eta \cdot \frac{|\tilde{B}_x|}{n} + \lambda\sqrt{\eta \cdot \frac{|\tilde{B}_x|}{n}}$. Rearranging,

$$\frac{|\tilde{B}_x|}{n} \leq \lambda^2 \frac{\eta}{(1-\eta)^2}$$

which is at most γ by our choice of λ. Therefore, for every $x \notin \mathcal{X}_0$, $\Pr_\sigma\left[\, x \in \tilde{B}_x\,\right] \leq \gamma$, establishing the high-soundness statement.

The bound on d, as well as the time efficiency statement, finally follow from the following construction of expanders, due to Bilu and Linial:

Theorem C.3 ([HLW06, Theorem 6.12]). *For every $d \geq 3$, and every $n \geq 1$, there exists an explicit d-regular graph G on n vertices with spectral expansion $\lambda = O((\log^{3/2} d)/\sqrt{d})$. Moreover, G can be constructed in time polynomial in n and d.*

To achieve the desired bound on λ^2, it therefore suffices to have $d = \tilde{O}\left(\frac{\eta}{(1-\eta)^2\gamma}\right)$. □

Remark C.4. By a probabilistic argument, for all $n, d \geq 1$, and every constant $\delta > 0$, there exist d-regular graphs on n vertices with spectral expansion $\lambda \leq (2+\delta)\sqrt{d-1}/d$ (more precisely, almost all d-regular graph on n vertices have spectral expansion at most $(2+\delta)\sqrt{d-1}/d$). Therefore, if one does not insist on being able to efficiently construct such a graph, the bound on d in Lemma C.2 can be improved to $d \geq \frac{4.1\eta}{(1-\eta)^2\gamma}$.

D Omitted proof: Proof of Theorem 5.5

In this appendix, we prove of Theorem 5.5, restated below.

Theorem D.1 (Theorem 5.5, restated)**.** *Given $n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $s \in \mathbb{N}$, for a channel family $\mathcal{W}$ the (n, ε, s)-semimaxmin chi-square fluctuation is bounded as*

$$\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon, s) = O\left(\frac{n^2\varepsilon^4}{k^3} \cdot 2^s \cdot \max_{W\in\mathcal{W}} \|H(W)\|_*^2\right),$$

whenever

$$n \leq \frac{k^{3/2}}{C\varepsilon^2 2^{s/2} \max_{W\in\mathcal{W}} \|H(W)\|_*}, \tag{23}$$

where $C > 0$ is a constant.

Proof. To obtain the desired bound for semimaxmin chi-square fluctuation, we fix an arbitrary multiset $\mathcal{W}_s \subseteq \bar{\mathcal{W}}^n$ of size at most 2^s, and bound the average (over W^n in $\mathcal{W}_s$) decoupled chi-square fluctuation for a suitable almost ε-perturbation $\mathcal{P}_\zeta$. With this in mind, suppose we have a random variable $Z = (Z_1, \ldots, Z_{k/2})$ taking values in $[-1,1]^{k/2}$ and with distribution ζ such that

$$\Pr\left[\|Z\|_1 \geq \frac{k}{\beta}\right] \geq \alpha \tag{24}$$

for some constants $\alpha \geq 1/10$ and $\beta > 0$. For $\varepsilon \in (0, \beta^{-1})$, consider the perturbed family around $\mathbf{u}_k$ consisting of elements $\mathbf{p}_z$, $z \in [-1,1]^{k/2}$, given by

$$\mathbf{p}_z = \frac{1}{k}\left(1+\beta\varepsilon z_1, 1-\beta\varepsilon z_1, \ldots, 1+\beta\varepsilon z_t, 1-\beta\varepsilon z_t, \ldots, 1+\beta\varepsilon z_{k/2}, 1-\beta\varepsilon z_{k/2}\right). \tag{25}$$

By our assumption on Z (Eq. (24)), $\mathbf{p}_Z$ then satisfies $\mathrm{d_{TV}}(\mathbf{p}_Z, \mathbf{u}_k) = \frac{\beta\varepsilon}{k}\|Z\|_1 \geq \varepsilon$ with probability at least α. Consider any $W^n \in \mathcal{W}_s^n$. From the same steps as in [ACT18, Theorem IV.14], we get

$$\chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right) = \ln \mathbb{E}_{ZZ'}\left[\exp\left(\frac{\beta^2 n\varepsilon^2}{k} \cdot Z^\intercal \bar{H}(W^n) Z'\right)\right], \tag{26}$$

where Z, Z' are independent random variables with common distribution ζ, $\bar{H}(W^n) := \frac{1}{n}\sum_{j=1}^n H(W_j)$, and $H(W_j)$ is defined as in Eq. (13). Now, to bound the semimaxmin chi-square fluctuation, we must handle $\mathbb{E}_{W^n}[\chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right)]$, for W^n drawn uniformly at random from $\mathcal{W}_s$. We thus define the new "aggregate" matrix

$$\tilde{H}(\mathcal{W}_s) := \sum_{W^n\in\mathcal{W}_s} \overline{H}(W^n)$$

to which we apply a construction of Acharya, Canonne, and Tyagi, whose properties we summarize below.

Lemma D.2 (Implicit in the proof of [ACT18, Theorem IV.18]). *Let $A \in \mathbb{R}^{(k/2)\times(k/2)}$ be a p.s.d. matrix. Then, there exists a matrix $V \in \mathbb{R}^{(k/2)\times(k/4)}$ such that the following holds.*

(i) Each row vector of V has ℓ_2 norm at most 1, and V has Frobenius norm $\|V\|_F \geq \sqrt{k}/2$.

(ii) Let $Y = (Y_1 \dots Y_{k/4})$ be a vector of i.i.d. Rademacher random variables. Then,[9]

$$\Pr\left[\|VY\|_1 \geq \frac{k}{12\sqrt{2}}\right] \geq \frac{1}{9}.$$

(iii) We have $\|V^\intercal AV\|_F^2 \leq \frac{4}{k}\|A\|_*^2$.

We invoke this lemma on $\tilde{H}(\mathcal{W}_s)$, and denote by $V \in \mathbb{R}^{(k/2)\times(k/4)}$ the resulting matrix. Letting ζ be the distribution of the random variable $Z := VY$, where Y is a vector of $k/4$ i.i.d. Rademacher random variables, item (ii) implies that ζ satisfies the condition from Eq. (24), for $\alpha := 1/9$ and $c := 1/(12\sqrt{2})$. Moreover, since all $\overline{H}(W^n)$ (and therefore all $V^\intercal\overline{H}(W^n)V$'s) are symmetric p.s.d. matrices, we have[10]

$$\|V^\intercal \sum_{W^n\in\mathcal{W}_s} \overline{H}(W^n)V\|_F^2 \geq \sum_{W^n\in\mathcal{W}_s} \|V^\intercal\overline{H}(W^n)V\|_F^2,$$

or, equivalently, $2^{-s}\|V^\intercal\tilde{H}(\mathcal{W}_s)V\|_F^2 \geq \mathbb{E}_{W^n}\left[\|V^\intercal\overline{H}(W^n)V\|_F^2\right]$. However, by item (iii), $\|V^\intercal\tilde{H}(\mathcal{W}_s)V\|_F^2 \leq (4/k)\|\tilde{H}(\mathcal{W}_s)\|_*^2$. Since, by the triangle inequality, we further have

$$\|\tilde{H}(\mathcal{W}_s)\|_* \leq 2^s \max_{W^n\in\mathcal{W}_s}\|\overline{H}(W^n)\|_* \leq 2^s \max_{W\in\mathcal{W}}\|H(W)\|_*$$

we obtain

$$\mathbb{E}_{W^n}\left[\|V^\intercal\overline{H}(W^n)V\|_F^2\right] \leq \frac{4\cdot 2^s}{k}\max_{W\in\mathcal{W}}\|H(W)\|_*^2. \tag{27}$$

We can now bound $\mathbb{E}_{W^n}[\chi^{(2)}(W^n \mid \mathcal{P}_\zeta)]$. Let $c > 0$ be the constant from the statement of Lemma 5.4. Setting $\lambda := (\beta^2 n\varepsilon^2)/k$ and recalling our assumption (Eq. (23)) on n, we have

$$1 \geq \frac{16\beta^2 n\varepsilon^2 2^{s/2}\cdot\max_{W\in\mathcal{W}}\|H(W)\|_*}{ck^{3/2}} \geq \frac{8\lambda}{c}\sqrt{\mathbb{E}_{W^n}\left[\|V^\intercal\overline{H}(W^n)V\|_F^2\right]} \geq \frac{8\lambda}{c}\mathbb{E}_{W^n}\left[\|V^\intercal\overline{H}(W^n)V\|_F\right]$$
$$\geq \frac{8\lambda}{c}\mathbb{E}_{W^n}\left[\rho(V^\intercal\bar{H}(W^n)V)\right],$$

where the second inequality is Eq. (27) and $\rho(A)$ denotes the spectral norm of matrix A. By Markov's inequality, we have that

$$\Pr_{W^n}\left[\rho(V^\intercal\bar{H}(W^n)V) > \frac{1}{4\lambda}\right] \leq \frac{c}{2}.$$

[9] We note that this second item is a consequence of the first, along with Khintchine's inequality and an anticoncentration argument; see [ACT18, Claim IV.21]. For clarity, we nonetheless explicitly state both here.

[10] This follows from the fact that $\operatorname{Tr} AB \geq 0$ for two p.s.d. matrices A, B; and thus $\|A+B\|_F^2 = \operatorname{Tr}[(A+B)^2] = \operatorname{Tr}[A^2] + \operatorname{Tr}[B^2] + 2\operatorname{Tr}[AB] \geq \|A\|_F^2 + \|B\|_F^2$.

Let $\mathcal{G} \subseteq \mathcal{W}_s$ (for "good") be the multiset of W^n's such that $\rho(V^\intercal \bar{H}(W^n)V) \leq 1/(4\lambda)$, which by the above has size at least $(1-c/2)\cdot 2^s$. Upon reorganizing, for any $W^n \in \mathcal{G}$ we have

$$\lambda^2/(1-4\lambda^2\rho(V^\intercal \bar{H}(W^n)V)^2) \leq 4\lambda^2/3\,.$$

We can then apply the lemma below on the MGF of a Rademacher chaos to i.i.d. Rademacher random variables Y and the symmetric matrix $V^\intercal \bar{H}(W^n)V \in \mathbb{R}^{k/4\times k/4}$:

Lemma D.3 ([ACT18, Claim IV.17])**.** *For random vectors $\theta, \theta' \in \{-1,1\}^{k/2}$ with each θ_i and θ'_i distributed uniformly over $\{-1,1\}$, independent of each other and independent for different i's. Then, for a positive semi-definite matrix H,*

$$\ln \mathbb{E}_{\theta\theta'}\left[e^{\lambda\theta^\intercal H\theta'}\right] \leq \frac{\lambda^2}{2}\cdot\frac{\|H\|_F^2}{1-4\lambda^2\rho(H)^2}\,, \quad \forall\, 0\leq\lambda<\frac{1}{2\rho(H)},$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\rho(\cdot)$ the spectral radius.

This gives, for any $W^n \in \mathcal{G}$,

$$\mathbb{E}_{ZZ'}\left[\exp\left(\tfrac{\beta^2 n\varepsilon^2}{k}\cdot Z^\intercal \bar{H}(W^n)Z'\right)\right] = \mathbb{E}_{YY'}[e^{\frac{\beta^2 n\varepsilon^2}{k}Y^\intercal V^\intercal \bar{H}(W^n)VY'}] \leq e^{\frac{2\beta^4 n^2\varepsilon^4}{3k^2}\|V^\intercal \bar{H}(W^n)V\|_F^2}\,.$$

From there, by concavity and using Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}_{W^n}[1\wedge\chi^{(2)}(W^n\mid\mathcal{P}_\zeta)] &\leq \mathbb{E}_{W^n}[\chi^{(2)}(W^n\mid\mathcal{P}_\zeta)\,\mathbb{1}_{\mathcal{G}}(W^n)+\mathbb{1}_{\mathcal{G}^c}(W^n)]\\
&\leq \mathbb{E}_{W^n}[\chi^{(2)}(W^n\mid\mathcal{P}_\zeta)\,\mathbb{1}_{\mathcal{G}}(W^n)]+\frac{c}{2}\\
&\leq \mathbb{E}_{W^n}[\ln(e^{\frac{2\beta^4 n^2\varepsilon^4}{3k^2}\|V^\intercal \bar{H}(W^n)V\|_F^2})]+\frac{c}{2}\\
&\leq \ln(e^{\frac{2\beta^4 n^2\varepsilon^4}{3k^2}\mathbb{E}_{W^n}[\|V^\intercal \bar{H}(W^n)V\|_F^2]})+\frac{c}{2}\,.
\end{aligned}$$

The above, along with Eq. (27), then finally yields

$$\mathbb{E}_{W^n}[1\wedge\chi^{(2)}(W^n\mid\mathcal{P}_\zeta)] \leq \frac{8\beta^4 n^2\varepsilon^4 2^s}{3k^3}\max_{W\in\mathcal{W}}\|H(W)\|_*^2+\frac{c}{2}\,,$$

which, invoking Lemma 5.4, completes the proof. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il