Distributed Gaussian Mean Testing Under Communication Constraints: Messages, Samples, and Coins

Clément L. Canonne
University of Sydney, Australia

Nimitt
IIT Gandhinagar, India

Abstract

We revisit the problem of Gaussian mean testing in a distributed, communication constrained setting, where each of n users independently observes samples from an unknown d-dimensional spherical Gaussian distribution $\mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$, and can communicate up to ℓ bits to a central referee. The referee's goal is then to distinguish between cases (i) $\|\boldsymbol{\mu}\|_2 = 0$ versus (ii) $\|\boldsymbol{\mu}\|_2 \geq \varepsilon$. This problem has been considered in the private- and public-coin settings, when each user holds exactly *one* sample, or more generally when each holds exactly m samples. In this work, we significantly generalize the question in three directions: when the users only share a small number s of random bits, when each user holds a different number of samples m_k, and when each user can send a different number of bits ℓ_k to the referee.

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Distributed algorithms; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Distribution Testing, Property Testing, Distributed Algorithms, Communication Constraints

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.36

Category RANDOM

Related Version *Preprint*: `https://arxiv.org/abs/2605.29426`

Acknowledgements The authors thank the anonymous reviewers for their helpful suggestions.

1 Introduction

You have been put in charge of a new program, to monitor underwater volcanic activity in a large sector of the ocean. To maximize coverage, you have identified a set of n locations where to deploy buoys equipped with high-tech measuring devices, strategically chosen to maximize coverage. Each of these sensors will measure the volcanic activity at regular intervals, and send hourly the result to the central station for processing and analysis, to detect any risk. But choosing the location of these buoys was only the first step: the ocean is a harsh and big place, and as a result these sensors are not only quite distant from the station, they also have strong limitations on how many good-quality measurements they can make and how much data they can accurately transmit before running out of energy for the day. The measurements they make are subject to significant background noise; and even broadcasting data *to* the network of buoys to synchronize them is complex, given how remote they all are. So you are left with a daunting, highly constrained task:

> *How to design a good detection protocol under heterogeneous constraints, where each sensor can make a different number of measurements, send a different amount of data back to the center, and they all only share a limited amount of coordination?*

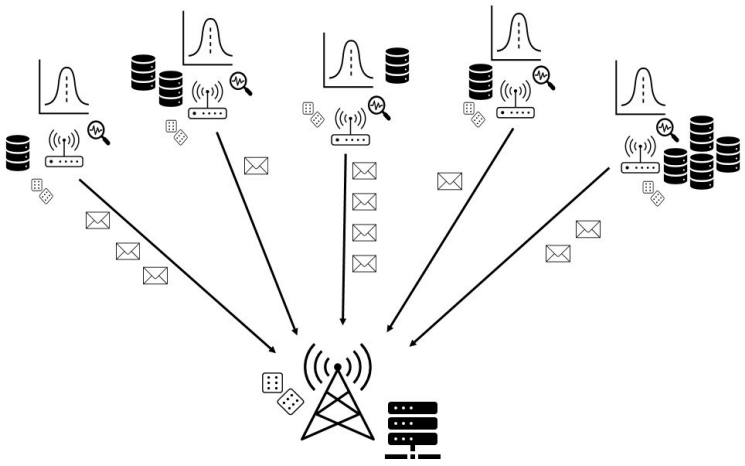

Figure 1 A depiction of the setting considered in this work: several distributed machines ("users") are measuring the same signal (vector $\boldsymbol{\mu} \in \mathbb{R}^d$) subject to white noise (spherical Gaussian). They all share a short random seed, and have different capabilities (amount of data collected, amount of communication they can send): by sending a message to a central server ("referee"), they must enable it to test whether the underlying signal measure exists ($\|\boldsymbol{\mu}\|_2 > \varepsilon$) or is pure noise ($\boldsymbol{\mu} = \mathbf{0}$).

Problem Statement

We formulate the problem as a composite hypothesis testing task. Let $\mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$ denote a d-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance $\mathbb{I}_d$. Define the following hypothesis classes for a pre-specified threshold $\varepsilon \in (0, 1]$:

$$\mathcal{H}_0 := \{\mathcal{G}(\mathbf{0}, \mathbb{I}_d)\} \quad \mathcal{H}_1 := \{\mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d) : \|\boldsymbol{\mu}\|_2 \geq \varepsilon\}$$

In a distributed setting, n users observe i.i.d. samples from an unknown distribution $G \in \mathcal{H}_0 \cup \mathcal{H}_1$ and transmit ℓ-bit messages to a central referee. Specifically, user $k \in [n]$ observes $m_k \geq 1$ independent samples from G:

$$X^{(k)} = \left(\boldsymbol{X}^{(k,1)}, \boldsymbol{X}^{(k,2)}, \cdots, \boldsymbol{X}^{(k,m_k)}\right), \quad \boldsymbol{X}^{(k,i)} \sim G.$$

and applies a (possibly randomized) encoding function $W^{(k)} \colon \mathbb{R}^{d \cdot m_k} \to \mathcal{Y}_k := \{0,1\}^{\ell_k}$ to it, before sending the resulting message $\boldsymbol{Y}^{(k)} = W^{(k)}\left(X^{(k)}\right)$ to the referee. The referee then applies a test $T \colon \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n \to \{0,1\}$ to the vector of received messages $Y^n = \left(\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \cdots, \boldsymbol{Y}^{(n)}\right)$, and accepts $\mathcal{H}_0$ if, and only if, $T(Y^n) = 0$.

We allow the users to share $s \geq 0$ bits of *public randomness*, which are used to jointly sample the tuple encoding functions $W^n = (W^{(1)}, W^{(2)}, \cdots, W^{(n)})$.[1] We define the *two-sided error* for protocol $\Pi := (W^n, T)$ as

$$\delta(\Pi) = \max\left(\sup_{G \in \mathcal{H}_0} \Pr\left[T(Y^n) = 1\right], \sup_{G \in \mathcal{H}_1} \Pr\left[T(Y^n) = 0\right]\right)$$

The goal is to find required amount of *resources* (amount of shared randomness s, numbers of samples $(m_k)_k$, individual communication budgets $(\ell_k)_k$) such that there exists a protocol Π with two-sided error bounded by a target error probability δ (unless specified otherwise, fixed as a small constant $1/10$). For simplicity, we decompose the problem into following three subproblems:

[1] That is, the encoding functions (or "channels") are randomized using both $s \geq 0$ bits of public randomness, and an arbitrary number of bits of private randomness (independent across users).

1. **Limited public coin**: Each user observes only one sample, and sends ℓ bits ($m_k = 1$, $\ell_k = \ell$ for all k); the users are allowed to share s bits of public randomness. What is the *sample complexity*, that is, the minimum $n = n(d, s, \ell, \varepsilon)$ such that $\inf_\Pi \delta(\Pi) \le 1/10$?
2. **Heterogeneous samples**: The users share $s = \Omega(\log(d/\ell))$ bits of public randomness, and each sends ℓ bits ($\ell_k = \ell$ for all k). What are sufficient conditions on $\boldsymbol{m} = (m_1, \dots, m_n)$ such that there exists a protocol Π with $\delta(\Pi) \le 1/10$?
3. **Heterogeneous communication**: The users share $s = \Omega(\log(d/\ell))$ bits of public randomness, and each observes one sample ($m_k = 1$ for all k). What are sufficient conditions on $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n)$ such that there exists a protocol Π with $\delta(\Pi) \le 1/10$?

Our results

We now summarize our results for these three subproblems.

▶ **Theorem 1** (Limited public-coin). *For $\ell \in [d]$ and $s \ge 0$, there exists an s-public coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraint with*

$$O\left(\frac{d}{\ell\varepsilon^2} \cdot \sqrt{\max\left(\frac{d}{2^{\Theta(s)}}, \ell\right)}\right)$$

users, each holding one sample.

Note that for $s = 0$ and $s = O(\log(d/\ell))$, this matches the *private-* and *public-coin* bounds of $O(d^{3/2}/(\ell\varepsilon^2))$ and $O(d/(\sqrt{\ell}\varepsilon^2))$ from [3], known to be tight [12].

▶ **Theorem 2** (Heterogeneous samples). *For $\ell \in [r, rd]$, $s = \Omega(\log(d/\ell))$, and $\boldsymbol{m} = (m_1, \dots, m_n) \in \mathbb{Z}_+^n$ satisfying $m_k \in [r, r\varepsilon^{-2}]$ for all $k \in [n]$, there exists a protocol for distributed Gaussian mean testing under ℓ-bit communication constraint with n users provided that*

$$\frac{\sum_{1 \le k_1 \ne k_2 \le n} \sqrt{m_{k_1} m_{k_2}}}{n} \ge C \cdot \frac{d}{\sqrt{\ell}\varepsilon^2}$$

where $C, r > 0$ are absolute constants.

To interpret this result, observe that, at one extreme, with $m_k = \Theta(1)$ for all k, the expression matches the *public-coin* bound. Moreover, we note that for a fixed "budget" of samples $m = \sum_{k=1}^n m_k$, a "balanced" allocation of samples across users maximizes the left-hand side, and hence is better (at least for our testing protocol).

▶ **Theorem 3** (Heterogeneous communication). *For $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n) \in \mathbb{Z}_+^n$ and $s = \Omega(\log(d/\|\boldsymbol{\ell}\|_\infty))$, there exists a protocol for distributed Gaussian mean testing under $\boldsymbol{\ell}$-bit communication constraint with n users, where user k has one sample and can send ℓ_k bits to the referee, provided that*

$$\frac{\|\boldsymbol{\ell}\|_1}{\sqrt{\|\boldsymbol{\ell}\|_\infty}} \ge C \cdot \frac{d}{\varepsilon^2}$$

where $C > 0$ is an absolute constant.

Letting $\tilde{\ell} = \frac{1}{n}\sum_{k=1}^{n} \ell_k$ be the average communication budget, one can rephrase the sufficient condition as

$$n \cdot \frac{\tilde{\ell}}{\sqrt{\max_{1\le k\le n} \ell_k}} \ge C \cdot \frac{d}{\varepsilon^2}$$

which, for a fixed value of $\tilde{\ell}$, is easiest to satisfy when all communication constraints are the same ("balanced").

We emphasize that this separation in three subproblems is for clarity and simplicity of exposition: we also provide a protocol involving a "mix-and-match" of all parameters, where public randomness s, sample counts $\boldsymbol{m} = (m_1, \dots, m_n)$, and individual communication budgets $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n)$ are all kept as free parameters. The guarantees of this protocol are summarized in the following theorem:

▶ **Theorem 4** (Everything, everywhere, all at once)**.** *For $s \ge 0$, $\boldsymbol{m} = (m_1, \dots, m_n) \in \mathbb{Z}_+^n$, and $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n) \in \mathbb{Z}_+^n$, there exists a protocol for distributed Gaussian mean testing with n users sharing s bits of public randomness, where user k has m_k samples and can send ℓ_k bits to the referee, provided that there exists a partition $\mathcal{P} = \{P_1, \dots, P_K\}$ of the n users such that*

$$\frac{\sum_{j_1 \ne j_2 \in [K]} \sqrt{\min_{i_1 \in P_{j_1}}(m_{i_1}) \cdot \min_{i_2 \in P_{j_2}}(m_{i_2})}}{K} \ge C \cdot \frac{d}{\varepsilon^2 \sqrt{\max\left(\frac{d}{2^{\Theta(s)}}, \|\boldsymbol{\ell}\|_\infty\right)}}$$

and, for every $P \in \mathcal{P}$, $\sum_{i\in P} \ell_i \ge r \cdot \max\left(\frac{d}{2^{\Theta(s)}}, \|\boldsymbol{\ell}\|_\infty\right)$ and $\min_{i\in P}(m_i) \in [r, r\varepsilon^{-2}]$; where $C, r > 0$ are absolute constants.

Related work

In the centralized setting where all samples are fully available to the referee, the tight sample complexity of for Gaussian mean testing is known, and, while considered "folklore", can for instance be found in [7, 10] (as a special case of more general problem). Subsequent work has expanded on this foundation, exploring Gaussian mean testing under communication constraints [3] and truncation models [9]. [3] studied the problem in two variants of homogeneous regime (one sample and same communication budget ℓ for all users), namely the *private-coin* ($s = 0$) and *public-coin* ($s = \infty$), and showed sample complexities of

$$O\left(\frac{d^{3/2}}{\ell\varepsilon^2}\right) \quad (\textit{private coin}), \; O\left(\frac{d}{\sqrt{\ell}\varepsilon^2}\right) \quad (\textit{public coin})$$

While the authors only established upper bounds and (limited) lower bounds, follow-up work [11, 12] showed these two sample complexities to be tight. (Moreover, exploiting standard properties of the Gaussian distribution, the above easily generalizes to the case where each of the n users holds the same number of samples $m \ge 1$.[2])

The above works reveal a $\sqrt{d/\ell}$ gap between the private- and public-coin sample complexities, highlighting the value of shared randomness. However, the public-coin protocol of [3] (as well as the similar protocol of [12]) rely on jointly sampling a uniformly random rotation

[2] Namely, if $\boldsymbol{X}^{(k,1)}, \boldsymbol{X}^{(k,2)}, \cdots, \boldsymbol{X}^{(k,m_k)} \sim \mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$, then $\tilde{\boldsymbol{X}}^{(k)} := \frac{1}{\sqrt{m}}\sum_{i=1}^{m} \boldsymbol{X}^{(k,i)} \sim \mathcal{G}(\sqrt{m}\boldsymbol{\mu}, \mathbb{I}_d)$, so one can use the one-sample protocol with new distance parameter $\varepsilon' = \sqrt{m}\varepsilon$ and "single sample per user" $\tilde{\boldsymbol{X}}^{(k)}$.

of the space, which technically requires infinitely many shared random bits – even with adequate discretization, the amount of shared randomness necessary is still impractical. One can thus wonder what can be achieved with a *smaller* "shared randomness budget:" similarly, what happens if each user holds a different number of samples m_k, or can communicate a different number of bits ℓ_k. To the best of our knowledge, all of these remain essentially understudied and open in the context of Gaussian mean testing, and even distribution testing at large, with the two exceptions outlined below.

On the "limited randomness" side, our techniques and results for the s-public coin setting are reminiscent of the work of [1], which study the problem of *identity testing for discrete distributions* under communication constraints. They provide an optimal protocol for identity testing of distributions over a finite domain $[k]$ under ℓ-bit communication constraint with

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

users, giving clear interpolation between *private-* and *public-coin*. Our Blockwise Randomized Hadamard Transform (BRHT) primitive (see Section 3) for compression is similar to their technique "Domain Compression", showing an exponential saving (in the number of shared random bits) in the sample complexity of the problem. Interestingly, however, both the actual techniques used, and the resulting tradeoffs, are quite incomparable: the upshot of [1] was that "one bit of communication is equivalent to two bits of shared randomness" (in terms of sample complexity), which does no longer hold in the Gaussian mean testing setting.

On the "heterogeneous communication" side, while we are not aware of previous work for the case of Gaussian mean testing, we note that in the case of private-coin protocols for testing of discrete distributions, a baseline upper bound is hinted at in [6, Exercise 4.4], while *lower bounds* for *estimation* (not testing) are derived in [2].

We conclude by mentioning that, in the (as discussed above, analogous, yet orthogonal) setting of identity testing of discrete distributions, under local privacy constraints, the recent work [8] states that *"it is not at all obvious how this [heterogeneous sample sizes] could be handled neatly, and general results for this model could greatly help practical implementations."* While our results do not directly address their question (as we deal with mean testing and not discrete distributions, and communication instead of local privacy constraints), we hope that some of our ideas can help in this regard, especially given the observed similarities between communication and local privacy constraints [2].

2 Preliminaries

Notation

Throughout, we write $[n] := \{1, 2, \ldots, n\}$, and denote by $\boldsymbol{u}_d := (1/2, 1/2, \ldots, 1/2) \in \mathbb{R}^d$ the mean vector of a d-dimensional binary product uniform distribution over $\{0,1\}^d$, and by $\mathbb{I}_d$ the d-dimensional identity matrix. We use the standard asymptotic notations $O(\cdot), \Omega(\cdot)$ and $\Theta(\cdot)$.

Simplification

For clarity of exposition, we assume throughout that $d = 2^{k_1}$ and $\ell = 2^{k_2}$ for some integers $k_1, k_2 \geq 0$. This assumption is without loss of generality up to constant factors. Indeed, for arbitrary d, one can embed into dimension d' with

$$d \leq d' < 2d, \qquad d' = 2^{k_1}$$

via zero-padding. Similarly, for arbitrary ℓ, one can use ℓ' bits with

$$\ell/2 < \ell' \leq \ell, \qquad \ell' = 2^{k_2}.$$

We will heavily rely on the following tools and techniques from the literature, which we recall below.

▶ **Definition 5** (Normalized Sylvester Hadamard Matrix). *The* normalized Sylvester Hadamard matrix $H_d \in \mathbb{R}^{d\times d}$ *is recursively defined by*

$$H_1 := [1],$$

and for every integer $d \geq 1$

$$H_{2d} := \frac{1}{\sqrt{2}}\begin{bmatrix} H_d & H_d \\ H_d & -H_d \end{bmatrix}$$

▶ **Lemma 6** (Paley–Zygmund inequality). *For a non-negative random variable* Z *and* $0 \leq \theta \leq 1$

$$\Pr[Z > \theta\mathbb{E}[Z]] \geq (1-\theta)^2\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}$$

Binary product mean testing

Let P be an unknown d-dimensional binary product distribution over $\{0,1\}^d$ with mean vector $\boldsymbol{p} \in (0,1)^d$. Given i.i.d. samples from P and a threshold $\varepsilon > 0$, the binary product mean testing problem is to distinguish between cases $\boldsymbol{p} = \boldsymbol{u}_d$ versus $\|\boldsymbol{p} - \boldsymbol{u}_d\|_2 \geq \varepsilon$ with probability at least 99/100. It is known that the optimal sample complexity for this problem is $\Theta(\sqrt{d}/\varepsilon^2)$, and can be achieved by Algorithm 1 below. For completeness, we provide the analysis of Algorithm 1 in Appendix, Theorem 21.

Algorithm 1 Binary product mean testing.

Require: Samples $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)} \in \{0,1\}^d$; threshold $\varepsilon > 0$.

$\tau \leftarrow \varepsilon^2/2$
for $i \in [d]$ **do**
  $T_i \leftarrow \sum_{k_1 \neq k_2 \in [n]} (X_i^{(k_1)} - 1/2)(X_i^{(k_2)} - 1/2)$
$T \leftarrow \frac{1}{n(n-1)} \sum_{i=1}^d T_i$
if $T > \tau$ **then**
  return reject
else
  return accept

Private-coin

As discussed above, the private-coin setting, which corresponds to $s = 0$ bits of shared randomness, was studied for Gaussian mean testing in [3]. We briefly describe their protocol, as we will use it later on, which uses two main ingredients: a reduction to the binary product case, and a simple private-coin algorithm to solve the latter.

▶ **Lemma 7** (Reduction to binary product mean testing). *Let* $\varepsilon \in (0,1)$. *If there exists an algorithm for a binary product mean testing with* $n(d,\varepsilon)$ *samples, there exists an algorithm for Gaussian mean testing with* $n(d, \varepsilon/\sqrt{8})$.

Proof. Let $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ be i.i.d. samples from $\mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$. Define

$$\boldsymbol{Y}^{(k)} := \mathbb{I}(\boldsymbol{X}^{(k)} > 0) \in \{0,1\}^d.$$

Then the $\boldsymbol{Y}^{(k)}$ are i.i.d. samples from a binary product distribution P over $\{0,1\}^d$ with mean vector $\boldsymbol{p} \in (0,1)^d$. We distinguish the two cases,

Case 1: $\|\boldsymbol{\mu}\|_2 = 0$. Then, for all i, $p_i = \Pr[\boldsymbol{Y}_i^{(k)} = 1] = 1/2$. Thus, $\boldsymbol{p} = \boldsymbol{u}_d$.

Case 2: if $\|\boldsymbol{\mu}\|_2 \geq \varepsilon$. Then, for all i, $p_i = \Pr[\boldsymbol{Y}_i^{(k)} = 1] = \frac{1}{2}\operatorname{Erfc}\left(-\mu_i/\sqrt{2}\right)$. Using standard properties of Erfc, we have

$$\sum_{i=1}^{d}(p_i - 1/2)^2 \geq \frac{1}{8}\min\left(\sum_{i=1}^{d}\mu_i^2, d/2\right) = \frac{1}{8}\min\left(\|\boldsymbol{\mu}\|_2^2, d/2\right) \geq \frac{\varepsilon^2}{8}$$

Thus, $\|\boldsymbol{p} - \boldsymbol{u}_d\|_2 \geq \varepsilon/\sqrt{8}$.

Hence, to distinguish between the two cases, it is sufficient to test if $\boldsymbol{p} = \boldsymbol{u}_d$ versus $\|\boldsymbol{p} - \boldsymbol{u}_d\|_2 \geq \varepsilon/\sqrt{8}$, which is exactly the binary product mean testing problem. ◀

▶ **Theorem 8** ([3] Theorem 1). *For all $\ell \in [d]$, there exists a private-coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraint with $O(d^{3/2}/\ell\varepsilon^2)$ users.*

Proof. Let $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)}$ be i.i.d. samples from $\mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$ where user k observes $\boldsymbol{X}^{(k)}$. Define $\boldsymbol{Y}^{(k)} := \mathbf{1}(\boldsymbol{X}^{(k)} > 0) \in \{0,1\}^d$. Then $\boldsymbol{Y}^{(k)}$ are i.i.d. samples from a binary product distribution P over $\{0,1\}^d$ with mean vector $\boldsymbol{p} \in (0,1)^d$. First, note that if $\ell = d$, each user can transmit the entire vector $\boldsymbol{Y}^{(k)}$, yielding a protocol with $\Theta(\sqrt{d}/\varepsilon^2)$ users using Lemma 7.

When $\ell < d$, we use the "*simulate and infer*" technique of [4]: d/ℓ users can simulate *one* sample of the d-dimensional product distribution, by having send (a different set of) ℓ coordinates. Combining this with the $\Theta(\sqrt{d}/\varepsilon^2)$ centralized sample complexity of binary product mean testing, this gives a protocol with sample complexity

$$\frac{d}{\ell} \cdot O\left(\frac{\sqrt{d}}{\varepsilon^2}\right) = O\left(\frac{d^{3/2}}{\ell\varepsilon^2}\right)$$

as claimed. ◀

3 Blockwise Randomized Hadamard Transform

In this section, we define and analyze the properties of a key primitive used throughout the paper, the *Blockwise Randomized Hadamard Transform (BRHT)*. At a high level, the idea is to replace a uniformly random rotation of $\mathbb{R}^d$ by a discrete analogue, which can be sampled using a *finite* number of random bits. An obvious candidate would be to use either a matrix of uniformly random ± 1; however, our aim is to preserve the ℓ_2 geometry, and this fails to do so. Indeed, we want an isometry, something which would play the same role as a uniformly random rotation, and in particular a transformation R which, when applied to a spherical Gaussian, would still result in a spherical Gaussian: which requires $R^\top R = \mathbb{I}_d$. This makes using a Hadamard matrix a very natural idea: however, with such a deterministic construction, we lack the randomness which enables more efficient testing protocols. To remedy this while still using a minimum amount of randomness, we consider a structured random orthogonal transform derived from the Hadamard matrix, which we define Blockwise Randomized Hadamard Transform (BRHT), and provides strong sketching guarantees with control over randomness requirement.

We note that structural randomized Hadamard transforms have previously been studied in areas such as fast dimensionality reduction and randomized linear algebra (e.g., the Fast Johnoson–Lindenstrauss Transform [5]). Our own primitive, BRHT, is tailored to the Gaussian mean testing setting, where the objective is not Johnson–Lindenstrauss embeddings but rather finite-randomness compression satisfying the block-wise energy preservation property established in Lemma 11. We formally define the BRHT below.

▶ **Definition 9** (Blockwise Randomized Hadamard Transform (BRHT))**.** *Let $d, L \in \mathbb{N}$ satisfy $L \mid d$, and define $b := d/L$. A (d, L)-*Blockwise Randomized Hadamard Transform (BRHT)* is the random matrix*

$$R := H_d D$$

where $H_d \in \mathbb{R}^{d\times d}$ is a normalized Sylvester Hadamard matrix and $D \in \mathbb{R}^{d\times d}$ is a blockwise diagonal matrix:

$$\mathrm{diag}(D) = \left(\boldsymbol{\delta}^{(1)}, \boldsymbol{\delta}^{(2)}, \cdots, \boldsymbol{\delta}^{(b)}\right)$$

where for every $i \in [b]$, $\boldsymbol{\delta}^{(i)} = (\delta_i, \delta_i, \cdots, \delta_i) \in \{-1, 1\}^L$ and $\delta_1, \ldots, \delta_b$ are 4-wise independent Rademacher random variables.

▶ **Lemma 10** (Orthogonality)**.** *Let R be a (d, L)-BRHT. Then $RR^\top = \mathbb{I}_d$.*

Proof. Since H_d is orthogonal and D is diagonal with entries in $\{\pm 1\}$, $RR^\top = (H_dD)(H_dD)^\top = H_dDD^\top H_d^\top = H_dH_d^\top = \mathbb{I}_d$. ◀

With this in hand, we are ready to state and prove the key property of BRHT our algorithms will hinge upon.

▶ **Lemma 11** ((d, L)-compression)**.** *Let R be a (d, L)-BRHT. Then, for every $\boldsymbol{\mu} \in \mathbb{R}^d$ and every $t \geq 1$,*

$$\Pr\left[\left\|(R\boldsymbol{\mu})_{[1:tL]}\right\|_2^2 \geq \frac{1}{100} \cdot \frac{tL}{d} \cdot \|\boldsymbol{\mu}\|_2^2\right] \geq \frac{8}{25},$$

where the probability is over the choice of R.

Proof. Let $b := d/L$ and $\rho := \|\boldsymbol{\mu}\|_2$. By Definition 9,

$$R = H_dD.$$

Define

$$Z := \left\|(R\boldsymbol{\mu})_{[1:tL]}\right\|_2^2.$$

Partition

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_b), \qquad \boldsymbol{\mu}_i \in \mathbb{R}^L.$$

Using the Kronecker decomposition

$$H_d = H_b \otimes H_L,$$

the r-th block of $R\boldsymbol{\mu}$ satisfies

$$(R\boldsymbol{\mu})_{[(r-1)L+1:rL]} = \sum_{i=1}^{b} (H_b)_{r,i}\delta_i H_L \boldsymbol{\mu}_i.$$

Define

$$Z_r := \left\|(R\boldsymbol{\mu})_{[(r-1)L+1:rL]}\right\|_2^2,$$

so that

$$Z = \sum_{r=1}^{t} Z_r.$$

We first compute $\mathbb{E}[Z]$. Expanding Z_r gives

$$\begin{aligned} Z_r &= \left\| \sum_{i=1}^{b} (H_b)_{r,i}\delta_i H_L \boldsymbol{\mu}_i \right\|_2^2 \\ &= \sum_{i_1,i_2\in[b]} (H_b)_{r,i_1}(H_b)_{r,i_2}\delta_{i_1}\delta_{i_2} \left\langle H_L\boldsymbol{\mu}_{i_1}, H_L\boldsymbol{\mu}_{i_2}\right\rangle. \end{aligned}$$

Since H_L is orthogonal,

$$\langle H_L\boldsymbol{\mu}_{i_1}, H_L\boldsymbol{\mu}_{i_2}\rangle = \langle \boldsymbol{\mu}_{i_1}, \boldsymbol{\mu}_{i_2}\rangle.$$

Therefore,

$$Z_r = \sum_{i=1}^{b} (H_b)_{r,i}^2 \|\boldsymbol{\mu}_i\|_2^2 + \sum_{i_1\neq i_2} (H_b)_{r,i_1}(H_b)_{r,i_2}\delta_{i_1}\delta_{i_2}\langle \boldsymbol{\mu}_{i_1}, \boldsymbol{\mu}_{i_2}\rangle.$$

Since $(H_b)_{r,i}^2 = 1/b$,

$$Z_r = \frac{\rho^2}{b} + \sum_{i_1\neq i_2} (H_b)_{r,i_1}(H_b)_{r,i_2}\delta_{i_1}\delta_{i_2}\langle \boldsymbol{\mu}_{i_1}, \boldsymbol{\mu}_{i_2}\rangle.$$

By pairwise independence and symmetry of the Rademacher variables,

$$\mathbb{E}[\delta_{i_1}\delta_{i_2}] = 0 \qquad (i_1 \neq i_2),$$

and hence

$$\mathbb{E}[Z_r] = \frac{\rho^2}{b}.$$

Summing over $r \in [t]$,

$$\mathbb{E}[Z] = \frac{t\rho^2}{b}.$$

We now bound $\mathbb{E}[Z^2]$. For $r_1, r_2 \in [t]$,

$$\mathbb{E}[Z_{r_1}Z_{r_2}] = \frac{\rho^4}{b^2} + \sum_{\substack{i_1\neq i_2\\ j_1\neq j_2}} (H_b)_{r_1,i_1}(H_b)_{r_1,i_2}(H_b)_{r_2,j_1}(H_b)_{r_2,j_2} \cdot \mathbb{E}[\delta_{i_1}\delta_{i_2}\delta_{j_1}\delta_{j_2}]\langle \boldsymbol{\mu}_{i_1}, \boldsymbol{\mu}_{i_2}\rangle\langle \boldsymbol{\mu}_{j_1}, \boldsymbol{\mu}_{j_2}\rangle.$$

By 4-wise independence and symmetry of the Rademacher variables, only terms in which every δ_i appears with even multiplicity contribute to the expectation. Thus,

$$\mathbb{E}[Z_{r_1}Z_{r_2}] = \frac{\rho^4}{b^2} + 2\sum_{i_1\neq i_2} (H_b)_{r_1,i_1}(H_b)_{r_1,i_2}(H_b)_{r_2,i_1}(H_b)_{r_2,i_2}\langle \boldsymbol{\mu}_{i_1}, \boldsymbol{\mu}_{i_2}\rangle^2.$$

Using

$$(H_b)_{r,i} = \pm\frac{1}{\sqrt{b}}$$

and Cauchy–Schwarz,

$$\begin{aligned}\mathbb{E}[Z_{r_1}Z_{r_2}] &\le \frac{\rho^4}{b^2} + \frac{2}{b^2}\sum_{i_1\neq i_2}\|\boldsymbol{\mu}_{i_1}\|_2^2\|\boldsymbol{\mu}_{i_2}\|_2^2 \\ &\le \frac{3\rho^4}{b^2},\end{aligned}$$

where we used

$$\sum_{i=1}^{b}\|\boldsymbol{\mu}_i\|_2^2 = \rho^2.$$

Summing over $r_1, r_2 \in [t]$ yields

$$\mathbb{E}[Z^2] \le \frac{3t^2\rho^4}{b^2}.$$

Applying the Paley–Zygmund (Lemma 6),

$$\Pr[Z > \theta\mathbb{E}[Z]] \ge (1-\theta)^2\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \ge \frac{(1-\theta)^2}{3}.$$

Choosing $\theta = 1/100$ gives

$$\Pr\left[Z > \frac{t\rho^2}{100b}\right] \ge \frac{8}{25}.$$

Since

$$b = \frac{d}{L},$$

the claim follows. ◀

The compression guarantees established in Lemma 11 are the key property of BRHT used throughout the remainder of the paper. Roughly speaking, this guarantees that, with constant probability, the retained block preserves $\Omega(L/d)$ fraction of the signal energy, allowing the original Gaussian mean testing problem to be compressed while retaining sufficient signal for testing. To conclude, we analyze the amount of randomness required to sample a BRHT.

▶ Remark 12. A (d, L)-BRHT requires at most $4\log(d/L)$ bits of randomness as (d/L) 4-wise Rademacher random variables can be generated using a standard construction involving a degree-3 polynomial over an appropriate field (see, e.g., [13, Corollary 3.34]).

▶ Remark 13 (Computational complexity). For $\boldsymbol{v} \in \mathbb{R}^d$, $H_d\boldsymbol{v}$ can be computed in $O(d\log d)$ time via the Fast Walsh–Hadamard transform. This computation dominates the running time of our protocols.

4 Limited randomness

In this section, we prove Theorem 1, by providing and analyzing an explicit protocol for the problem. This protocol is quite simple: it first uses the shared randomness to jointly sample a BRHT R with suitable parameters; each user then applies R to their samples to obtain an instance of Gaussian mean testing in a lower-dimensional space. Having used all the shared randomness available, they then run the private-coin protocol of Theorem 8 on this lower-dimensional problem.

▶ **Theorem 14** (Theorem 1, restated). *For $\ell \in [d]$ and $s \geq 0$, there exists an s-public coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraint with*

$$O\left(\frac{d}{\ell\varepsilon^2} \cdot \sqrt{\max\left(\frac{d}{2^{\Theta(s)}}, \ell\right)}\right)$$

users, each holding one sample.

Proof. The users use $s/7$ bits of shared randomness to sample a (d, d_s)-BRHT R, where

$$d_s := d/2^{\lfloor s/28 \rfloor}\,.$$

This is possible by Remark 12, as sampling R requires at most $4\log(d/d_s)$ random bits. (We will explain the reason for the factor 7 in $s/7$ momentarily.)

Let $\boldsymbol{\mu} \in \mathbb{R}^d$ be an unknown mean vector and for all $k \in [n]$, $\boldsymbol{X}^{(k)} \sim \mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$ be the sample observed by user k. Define $L := d_s \vee \ell$ and $\tilde{\boldsymbol{\mu}} = (R\boldsymbol{\mu})_{[1:L]}$. Each user applies R to its sample to get $\tilde{\boldsymbol{X}}^{(k)} := (R\boldsymbol{X}^{(k)})_{[1:L]}$, so that $\tilde{\boldsymbol{X}}^{(k)} \sim \mathcal{G}\left(\tilde{\boldsymbol{\mu}}, \mathbb{I}_L\right)$. By Lemma 11, with probability at least $8/25$,

$$\|\tilde{\boldsymbol{\mu}}\|_2^2 \geq \frac{1}{100} \cdot \frac{L}{d} \cdot \|\boldsymbol{\mu}\|_2^2$$

Thus, if $\boldsymbol{\mu} = \mathbf{0}$, $\tilde{\boldsymbol{\mu}} = \mathbf{0}$ and otherwise $\|\boldsymbol{\mu}\|_2 \geq \varepsilon$ and with probability at least $8/25$, $\|\tilde{\boldsymbol{\mu}}\|_2^2 \geq \frac{L\varepsilon^2}{100d}$. The users and referee then use the private-coin protocol of Theorem 8 with $\tilde{\boldsymbol{X}}^{(k)}$ to distinguish between the two cases with

$$O\left(\frac{(L)^{3/2}}{\ell\left(\frac{L}{d}\varepsilon^2\right)}\right) = O\left(\frac{d\sqrt{L}}{\ell\varepsilon^2}\right)$$

users, (almost) proving the theorem. It remains to amplify the success probability.

Thus, with $O\left(\frac{d\sqrt{L}}{\ell\varepsilon^2}\right)$ users, if $\boldsymbol{\mu} = \mathbf{0}$, the referee accepts with probability at least $99/100$ (which is the failure probability from Theorem 8). But when $\|\boldsymbol{\mu}\|_2 \geq \varepsilon$, all we can say is that, by a union bound, the referee rejects with probability at least $1 - (\frac{17}{25} + \frac{8}{25} \cdot \frac{1}{100}) > 3/10$, as now the failure probability comes both from the choice of R and that of Theorem 8. Repeating the protocol 7 independent times, each time with new R, and accepting only if all repetitions output accept reduces the error probability below $1/10$. This increases the number of users only by a constant factor. Since $L = \max(d/2^{\lfloor s/28 \rfloor}, \ell)$, the theorem follows. ◀

With $s = O(\log(d/\ell))$, Theorem 1 yields sample complexity $O(d/\sqrt{\ell}\varepsilon^2)$, which matches the public-coin sample complexity in [3] when users must share an infinite amount of randomness.

▶ **Corollary 15** (Randomness-efficient optimal public-coin protocol). *There exists a protocol for distributed Gaussian mean testing under ℓ-bit communication constraints with $O(d/(\sqrt{\ell}\varepsilon^2))$ users, each holding one sample, and sharing $O(\log(d/\ell))$ bits of randomness.*

5 Heterogeneous samples

In this section, we prove Theorem 2. The core idea, which we refer to as *Signal Aggregation*, is to generalize the "naive strategy" previously discussed (see Footnote 2) where all users have the same number m of samples and aggregate them into a single "aggregated sample":

$$\boldsymbol{X}^{(k,1)}, \boldsymbol{X}^{(k,2)}, \cdots, \boldsymbol{X}^{(k,m_k)} \sim \mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d) \quad \rightsquigarrow \quad \tilde{\boldsymbol{X}}^{(k)} := \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \boldsymbol{X}^{(k,i)} \sim \mathcal{G}(\sqrt{m}\boldsymbol{\mu}, \mathbb{I}_d)$$

Doing so effectively amplifies the signal, which means that one can then use a protocol designed for *one* sample per user, but with better distance parameter $\varepsilon' := \sqrt{m}\varepsilon$.

In the heterogeneous sample case, one can attempt to do the same thing, but now (since user k has m_k samples, while user k' has a possible different number of samples $m_{k'}$) the resulting aggregated samples no longer have the same distribution in the no-case (when $\boldsymbol{\mu} \neq \mathbf{0}$). Dealing with the fact that samples are no longer i.i.d. is the key technical difficulty in proving Theorem 2.

▶ **Theorem 16** (Theorem 2, restated). *For $\ell \in [r, rd]$, $s = \Omega(\log(d/\ell))$, and $\boldsymbol{m} = (m_1, \ldots, m_n) \in \mathbb{Z}_+^n$ satisfying $m_k \in [r, r\varepsilon^{-2}]$ for all $k \in [n]$, there exists a protocol (Algorithm 2) for distributed Gaussian mean testing under ℓ-bit communication constraint with n users provided that*

$$\frac{\sum_{1 \leq k_1 \neq k_2 \leq n} \sqrt{m_{k_1} m_{k_2}}}{n} \geq C \cdot \frac{d}{\sqrt{\ell}\varepsilon^2}$$

where $C, r > 0$ are absolute constants.

Proof. We first describe the protocol in Algorithm 2 and then analyze its correctness.

Algorithm 2 ℓ-bit, $\boldsymbol{m}$-sample distributed Gaussian mean testing.

User k, for every $k \in [n]$:

Require: $\forall j \in [m_k]$, $\boldsymbol{X}^{(k,j)} \sim \mathcal{G}(\boldsymbol{\mu}, \mathbb{I}_d)$; communication budget $\ell \in [7, 7d]$; $\forall t \in [7]$, R_t: $(d, 7d/\ell)$-BRHT

for $t \in [7]$ **do**
  Samples $\boldsymbol{X}^{(k,t)} \leftarrow \frac{1}{\sqrt{\lfloor m_k/7 \rfloor}} \sum_{j=(t-1)\lfloor m_k/7 \rfloor + 1}^{t\lfloor m_k/7 \rfloor} \boldsymbol{X}^{(k,j)}$
  $\tilde{\boldsymbol{X}}^{(k,t)} \leftarrow R_t \boldsymbol{X}^{(k,t)}$
  $\boldsymbol{y}^{(k,t)} \leftarrow \mathbf{1}\left(\tilde{\boldsymbol{X}}^{(k,t)}_{[1:\lfloor \ell/7 \rfloor]} > 0\right)$
Send $\boldsymbol{Y}^{(k)} = (\boldsymbol{y}^{(k,1)}, \ldots, \boldsymbol{y}^{(k,7)})$

Referee:

Require: Messages $\forall k \in [n]$, $\boldsymbol{Y}^{(k)} \in \{0,1\}^\ell$, threshold ε

$N \leftarrow \sum_{k_1 \neq k_2} \sqrt{\lfloor m_{k_1}/7 \rfloor \lfloor m_{k_2}/7 \rfloor}$
for $t \in [7]$ **do**
  Run Algorithm 1 on samples $\boldsymbol{y}^{(1,t)}, \ldots, \boldsymbol{y}^{(n,t)}$ with
    threshold $\varepsilon' = \frac{\varepsilon}{80}\sqrt{\frac{\ell N}{7dn(n-1)}}$ ▷ $\tau \leftarrow (\varepsilon')^2/2$
  if output is reject **then**
    return reject
return accept

The protocol performs 7 independent repetitions for probability amplification. We analyze a single repetition and omit the superscript (t) throughout. For simplicity, we also ignore flooring operations, as they affect the bounds only up to constant factors. Fix one repetition, and let R denote the corresponding $(d, 7d/\ell)$-BRHT. Define

$$\tilde{\boldsymbol{\mu}} := (R\boldsymbol{\mu})_{[1:\ell/7]}.$$

For every user k,

$$\boldsymbol{X}^{(k)} = \frac{1}{\sqrt{m_k/7}} \sum_{j=1}^{m_k/7} \boldsymbol{X}^{(k,j)} \sim \mathcal{G}\left(\sqrt{m_k/7}\,\boldsymbol{\mu}, \mathbb{I}_d\right).$$

Since R is orthogonal,

$$\tilde{\boldsymbol{X}}^{(k)} := (R\boldsymbol{X}^{(k)})_{[1:\ell/7]} \sim \mathcal{G}\left(\sqrt{m_k/7}\,\tilde{\boldsymbol{\mu}}, \mathbb{I}_{\ell/7}\right).$$

Define

$$\boldsymbol{y}^{(k)} = \mathbf{1}(\tilde{\boldsymbol{X}}^{(k)} > 0) \in \{0,1\}^{\ell/7}.$$

The vectors $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(n)}$ are independent, though generally not identically distributed. For every coordinate $i \in [\ell/7]$, define

$$p_i^{(k)} := \Pr[y_i^{(k)} = 1], \qquad \tilde{p}_i^{(k)} := p_i^{(k)} - \frac{1}{2}.$$

The referee computes

$$T_i := \sum_{k_1 \neq k_2} (y_i^{(k_1)} - 1/2)(y_i^{(k_2)} - 1/2).$$

Since for all $k_1 \neq k_2$, $\tilde{p}_i^{(k_1)} \tilde{p}_i^{(k_2)} \geq 0$, applying Lemma 20 coordinate-wise yields

$$\mathbb{E}[T_i] = \sum_{k_1 \neq k_2} \tilde{p}_i^{(k_1)} \tilde{p}_i^{(k_2)},$$

and

$$\mathrm{Var}(T_i) \leq \frac{n(n-1)}{8} + (n-2)\mathbb{E}[T_i].$$

Define

$$T := \frac{1}{n(n-1)} \sum_{i=1}^{\ell/7} T_i.$$

Then

$$\begin{aligned}
\mathbb{E}[T] &= \frac{1}{n(n-1)} \sum_{i=1}^{\ell/7} \sum_{k_1 \neq k_2} \tilde{p}_i^{(k_1)} \tilde{p}_i^{(k_2)}, \\
\mathrm{Var}(T) &= \frac{1}{n^2(n-1)^2} \sum_{i=1}^{\ell/7} \mathrm{Var}(T_i) \\
&\leq \frac{\ell}{56\, n(n-1)} + \frac{(n-2)\mathbb{E}[T]}{n(n-1)}.
\end{aligned}$$

We now analyze the two cases:

Case 1: $\mu = 0$. In this case, $p_i^{(k)} = 1/2$ for all i, k, and hence $\tilde{p}_i^{(k)} = 0$. Therefore,

$$\mathbb{E}[T] = 0.$$

Applying Chebyshev's inequality,

$$\Pr[T \geq \tau] \leq \frac{\mathrm{Var}(T)}{\tau^2} \leq \frac{\ell}{56\, n(n-1)\tau^2}.$$

Using the threshold

$$\tau = \frac{1}{2} \cdot \frac{1}{6400} \cdot \frac{N}{n(n-1)} \cdot \frac{\ell\varepsilon^2}{7d},$$

we obtain

$$\Pr[T \geq \tau] = O\left(\frac{d^2 n(n-1)}{N^2 \ell \varepsilon^4}\right) < \frac{1}{100},$$

whenever

$$\frac{N}{n} > C_1 \cdot \frac{d}{\sqrt{\ell}\varepsilon^2}$$

for a sufficiently large absolute constant $C_1 > 0$.

Case 2: $\|\mu\|_2 \geq \varepsilon$. Using standard properties of Erfc,

$$|\tilde{p}_i^{(k)}| \geq \frac{1}{8}\sqrt{\frac{m_k}{7}}\,|\tilde{\mu}_i|,$$

Therefore,

$$\begin{aligned}
\mathbb{E}[T] &= \frac{1}{n(n-1)} \sum_{i=1}^{\ell/7} \sum_{k_1 \neq k_2} \tilde{p}_i^{(k_1)} \tilde{p}_i^{(k_2)} \\
&\geq \frac{1}{64\, n(n-1)} \sum_{i=1}^{\ell/7} \sum_{k_1 \neq k_2} \sqrt{\frac{m_{k_1}}{7}}\sqrt{\frac{m_{k_2}}{7}}\,\tilde{\mu}_i^2 \\
&= \frac{N}{64\, n(n-1)} \|\tilde{\boldsymbol{\mu}}\|_2^2.
\end{aligned}$$

By Lemma 11, with probability at least $8/25$,

$$\|\tilde{\boldsymbol{\mu}}\|_2^2 \geq \frac{1}{100} \cdot \frac{\ell}{7d} \cdot \varepsilon^2.$$

Hence, with probability at least $8/25$,

$$\mathbb{E}[T] \geq \frac{1}{6400} \cdot \frac{N}{n(n-1)} \cdot \frac{\ell\varepsilon^2}{7d}.$$

Conditioned on this event, applying Chebyshev's inequality,

$$\begin{aligned}
\Pr[T \leq \tau] &= \Pr\left(\mathbb{E}[T] - T \geq \frac{\mathbb{E}[T]}{2}\right) \\
&\leq \frac{4\,\mathrm{Var}(T)}{\mathbb{E}[T]^2} \\
&= O\left(\frac{d^2 n^2}{N^2 \ell \varepsilon^4} + \frac{nd}{N\ell\varepsilon^2}\right).
\end{aligned}$$

Thus,

$$\Pr[T \le \tau] < \frac{1}{100},$$

provided

$$\frac{N}{n} > C_2 \cdot \frac{d}{\sqrt{\ell}\varepsilon^2}$$

for a sufficiently large absolute constant $C_2 > 0$.

Therefore, a single repetition accepts with probability at least 99/100, if $\|\boldsymbol{\mu}\|_2 = 0$ and rejects with probability at least $1 - \left(\frac{17}{25} + \frac{8}{25} \cdot \frac{1}{100}\right) > \frac{3}{10}$, if $\|\boldsymbol{\mu}\|_2 \ge \varepsilon$. Repeating independently 7 times and accepting only if all repetitions output accept reduces the error probability below 1/10. Finally, by Remark 12, the 7 independent BRHTs require only $O(\log(d/\ell))$ shared random bits. ◀

▶ Remark 17. Note that for a fixed sampling budget $m = \sum_{k\in[n]} m_k$, $\sum_{k_1\neq k_2\in[n]} \sqrt{m_{k_1}}\sqrt{m_{k_2}}$ is maximized when $m_{k_1} = m_{k_2} = m/n$ for all $k_1, k_2 \in [n]$. Then, we need $m = O(d/\sqrt{\ell}\varepsilon^2)$.

6 Heterogeneous communication

Having addressed limited (public) randomness and heterogeneous sample sizes in the previous two sections, we now turn to the last generalization considered in this work, heterogeneous communication budgets. Specifically, we will establish Theorem 3, restated below:

▶ **Theorem 18** (Theorem 3, restated). *For $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n) \in \mathbb{Z}_+^n$ and $s = \Omega(\log(d/\|\boldsymbol{\ell}\|_\infty))$, there exists a protocol for distributed Gaussian mean testing under $\boldsymbol{\ell}$-bit communication constraint with n users, where user k has one sample and can send ℓ_k bits to the referee, provided that*

$$\frac{\|\boldsymbol{\ell}\|_1}{\sqrt{\|\boldsymbol{\ell}\|_\infty}} \ge C \cdot \frac{d}{\varepsilon^2}$$

where $C > 0$ is an absolute constant.

Proof. In view of the previous protocols, the idea is relatively straightforward: the users use $s/7$ bits of shared randomness to sample a (d, d_s)-BRHT R, for

$$d_s := d/2^{\lfloor s/28 \rfloor}.$$

Let $L := d_s \vee \max_{1\le k\le n} \ell_k = \Theta(\max_{1\le k\le n} \ell_k)$ (the last equality by our assumption on s) and $\tilde{\boldsymbol{\mu}} = (R\boldsymbol{\mu})_{[1:L]}$. As in the proof of Theorem 1, each user applies R to its sample to get $\tilde{\boldsymbol{X}}^{(k)} = (R\boldsymbol{X}^{(k)})_{[1:L]}$, so that $\tilde{\boldsymbol{X}}^{(k)} \sim \mathcal{G}(\tilde{\boldsymbol{\mu}}, \mathbb{I}_L)$. Using Lemma 11, with probability at least 8/25,

$$\|\tilde{\boldsymbol{\mu}}\|_2^2 \ge \frac{1}{100} \cdot \frac{L}{d} \cdot \|\boldsymbol{\mu}\|_2^2$$

so that in one case $\tilde{\boldsymbol{\mu}} = \mathbf{0}$ and in the other $\|\tilde{\boldsymbol{\mu}}\|_2^2 = \Omega(\frac{L\varepsilon^2}{d})$ with constant probability. This is where we depart from Theorem 1. The users and referee will now use the same idea of distributed simulation, but in a slightly different fashion: recall that all parameters are known to all parties, and that by choice of L we have $\ell_k \le L$ for all k. Then, user 1 sends the first ℓ_1 bits of their resulting L-dimensional sample; user 2 sends the following ℓ_2 bits,

wrapping around (i.e., the bits $\ell_1+1, \ell_1+2, \ldots, \ell_1+\ell_2 \bmod L$), and so on. Overall, the $\ell_1+\cdots+\ell_n = \|\boldsymbol{\ell}\|_1$ bits sent allow the referee to reconstruct exactly $n' := \lfloor \|\boldsymbol{\ell}\|_1/L \rfloor$ binary samples $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n'}$ as in the proof of Theorem 8, which leads to a successful protocol whenever

$$n' \geq C \cdot \frac{\sqrt{d_s}}{\varepsilon^2(d_s/d)}$$

i.e.,

$$\frac{\|\boldsymbol{\ell}\|_1}{\|\boldsymbol{\ell}\|_\infty} \geq C \cdot \frac{d}{\varepsilon^2\sqrt{\|\boldsymbol{\ell}\|_\infty}}$$

from our settings of L and d_s (after, as before, including the probability amplification over 7 independent repetitions). ◀

7 Putting it all together: Theorem 4

With protocols for the three settings, we conclude by providing a protocol for the general setting where shared randomness, sample counts and communication budgets are all free parameters. The protocol is relatively straightforward as it combines the protocols for the three settings.

▶ **Theorem 19** (Theorem 4, restated). *For $s \geq 0$, $\boldsymbol{m} = (m_1, \ldots, m_n) \in \mathbb{Z}_+^n$, and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n) \in \mathbb{Z}_+^n$, there exists a protocol for distributed Gaussian mean testing with n users sharing s bits of public randomness, where user k has m_k samples and can send ℓ_k bits to the referee, provided that there exists a partition $\mathcal{P} = \{P_1, \ldots, P_K\}$ of the n users such that*

$$\frac{\sum_{j_1 \neq j_2 \in [K]} \sqrt{\min_{i_1 \in P_{j_1}}(m_{i_1}) \cdot \min_{i_2 \in P_{j_2}}(m_{i_2})}}{K} \geq C \cdot \frac{d}{\varepsilon^2\sqrt{\max\left(\frac{d}{2^{\Theta(s)}}, \|\boldsymbol{\ell}\|_\infty\right)}}$$

and, for every $P \in \mathcal{P}$, $\sum_{i \in P} \ell_i \geq r \cdot \max\left(\frac{d}{2^{\Theta(s)}}, \|\boldsymbol{\ell}\|_\infty\right)$ and $\min_{i \in P}(m_i) \in [r, r\varepsilon^{-2}]$; where $C, r > 0$ are absolute constants.

Proof. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a partition of users (for all j_1, j_2 $P_{j_1} \cap P_{j_2} = \emptyset$, $\bigcup_j P_j = [n]$) such that for all j, $\sum_{i \in P_j} \ell_i \geq 7 \cdot \max\left(\frac{d}{2^{\lfloor s/28 \rfloor}}, \|\boldsymbol{\ell}\|_\infty\right)$. Define

$$L := \max\left(\frac{d}{2^{\lfloor s/28 \rfloor}}, \|\boldsymbol{\ell}\|_\infty\right)$$

Using the distributed simulation idea in Theorem 18, users in each $P_j \in \mathcal{P}$ can simulate *one* L-dimensional binary sample. But, users in P_j hold $(m_i)_{i \in P_j}$ samples. Thus, the setting can be reduced to $K = |\mathcal{P}|$ users holding $(\min_{i \in P_1}(m_i), \ldots, \min_{i \in P_K}(m_i))$ samples under L-bit communication constraints, for which we invoke the protocol of Algorithm 2. ◀

References

1 Jayadev Acharya, Clément L. Canonne, Yanjun Han, Ziteng Sun, and Himanshu Tyagi. Domain compression and its application to randomness-optimal distributed goodness-of-fit. In *COLT*, Proceedings of Machine Learning Research, pages 3–40. PMLR, 2020. URL: http://proceedings.mlr.press/v125/acharya20a.html.

2 Jayadev Acharya, Clément L. Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. In *NeurIPS*, 2023.

3 Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Distributed signal detection under communication constraints. In *COLT*, Proceedings of Machine Learning Research, pages 41–63. PMLR, 2020. URL: http://proceedings.mlr.press/v125/acharya20b.html.

4 Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints II: communication constraints and shared randomness. *arXiv preprint*, 66(12):7856–7877, 2020. arXiv:1905.08302.

5 Nir Ailon and Bernard Chazelle. The fast johnsonlindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39:302–322, May 2009. doi:10.1137/060673096.

6 Clément L. Canonne. Topics and techniques in distribution testing: A biased but representative sample. *Found. Trends Commun. Inf. Theory*, 19(6):1032–1198, 2022. doi:10.1561/0100000114.

7 Clement L. Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '21, pages 321–336, USA, 2021. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611976465.21.

8 Clément L. Canonne, Abigail Gentle, and Vikrant Singhal. Uniformity testing under user-level local privacy. In *ITCS*, LIPIcs, pages 33:1–33:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi:10.4230/LIPIcs.ITCS.2026.33.

9 Clément L. Canonne, Themis Gouleakis, Yuhao Wang, and Joy Qiping Yang. Gaussian mean testing under truncation. In *AISTATS*, Proceedings of Machine Learning Research, pages 4879–4887. PMLR, 2025. URL: https://proceedings.mlr.press/v258/canonne25a.html.

10 Ilias Diakonikolas, Daniel M. Kane, and Ankit Pensia. Gaussian mean testing made simple. In *SOSA*, pages 348–352. SIAM, 2023. doi:10.1137/1.9781611977585.CH31.

11 Botond Szabó, Lasse Vuursteen, and Harry van Zanten. Optimal distributed composite testing in high-dimensional gaussian models with 1-bit communication. *IEEE Trans. Inf. Theory*, 68(6):4070–4084, 2022. doi:10.1109/TIT.2022.3150599.

12 Botond Szabó, Lasse Vuursteen, and Harry van Zanten. Optimal high-dimensional and nonparametric distributed testing under communication constraints. *Ann. Statist.*, 51(3):909–934, 2023. doi:10.1214/23-aos2269.

13 Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. doi:10.1561/0400000010.

A Deferred Proofs

▶ **Lemma 20.** *Let $Z_1, \dots, Z_n$ be independent random variables such that $Z_k \in \{-1/2, 1/2\}$ and $\mathbb{E}[Z_k] = p_k$. Assume furthermore $p_{k_1} p_{k_2} \geq 0$ for all $k_1 \neq k_2$ and define*

$$T := \sum_{k_1 \neq k_2} Z_{k_1} Z_{k_2}.$$

Then the following holds:

1. $\mathbb{E}[T] = \sum_{k_1 \neq k_2} p_{k_1} p_{k_2}$
2. $\mathrm{Var}(T) \leq \frac{n(n-1)}{8} + (n-2)\mathbb{E}[T]$

Proof. Define

$$S := \sum_{k \in [n]} Z_k.$$

Since $Z_k^2 = 1/4$, we have

$$T = \sum_{k_1 \neq k_2} Z_{k_1} Z_{k_2} = \left(\sum_{k \in [n]} Z_k \right)^2 - \sum_{k \in [n]} Z_k^2 = S^2 - \frac{n}{4}.$$

1. $\mathbb{E}[\boldsymbol{T}]$**.** Using independence,

$$\mathbb{E}[S] = \sum_{k \in [n]} p_k$$

and

$$\mathrm{Var}(S) = \sum_{k \in [n]} \left(\frac{1}{4} - p_k^2 \right).$$

Hence,

$$\begin{aligned} \mathbb{E}[S^2] &= \mathrm{Var}(S) + \mathbb{E}[S]^2 \\ &= \sum_{k \in [n]} \left(\frac{1}{4} - p_k^2 \right) + \left(\sum_{k \in [n]} p_k \right)^2 \\ &= \frac{n}{4} + \sum_{k_1 \neq k_2} p_{k_1} p_{k_2}. \end{aligned}$$

Therefore,

$$\mathbb{E}[T] = \mathbb{E}[S^2] - \frac{n}{4} = \sum_{k_1 \neq k_2} p_{k_1} p_{k_2}.$$

2. $\mathbf{Var}[\boldsymbol{T}]$**.** Since $T = S^2 - n/4$,

$$\mathrm{Var}(T) = \mathrm{Var}(S^2) = \mathbb{E}[S^4] - \mathbb{E}[S^2]^2.$$

Expanding S^4 and using independence together with

$$\mathbb{E}[Z_k] = p_k, \qquad Z_k^2 = \frac{1}{4}, \qquad \mathbb{E}[Z_k^3] = \frac{p_k}{4}, \qquad Z_k^4 = \frac{1}{16},$$

we obtain

$$\begin{aligned} \mathbb{E}[S^4] = \frac{n}{16} &+ \frac{3n(n-1)}{16} \\ &+ \sum_{k_1 \neq k_2} p_{k_1} p_{k_2} + \frac{3(n-2)}{2} \sum_{k_1 \neq k_2} p_{k_1} p_{k_2} \\ &+ \sum_{k_1 \neq k_2 \neq k_3 \neq k_4} p_{k_1} p_{k_2} p_{k_3} p_{k_4}. \end{aligned}$$

On the other hand,

$$\mathbb{E}[S^2]^2 = \frac{n^2}{16} + \frac{n}{2} \sum_{k_1 \neq k_2} p_{k_1} p_{k_2} + \left(\sum_{k_1 \neq k_2} p_{k_1} p_{k_2} \right)^2.$$

Subtracting,

$$\mathrm{Var}(T) = \mathbb{E}[S^4] - \mathbb{E}[S^2]^2 = \frac{n(n-1)}{8} + (n-2)\sum_{k_1 \neq k_2} p_{k_1} p_{k_2} - 4\sum_{k_1 \neq k_2 \neq k_3} p_{k_1}^2 p_{k_2} p_{k_3} - 2\sum_{k_1 \neq k_2} p_{k_1}^2 p_{k_2}^2.$$

Since $p_{k_1} p_{k_2} \geq 0$ for all $k_1 \neq k_2$ and

$$\mathbb{E}[T] = \sum_{k_1 \neq k_2} p_{k_1} p_{k_2},$$

the result follows. ◀

▶ **Theorem 21.** *Algorithm 1 achieves sample complexity $\Theta(\sqrt{d}/\varepsilon^2)$ for testing whether $\boldsymbol{p} = \boldsymbol{u}_d$ versus $\|\boldsymbol{p} - \boldsymbol{u}_d\|_2 \geq \varepsilon$ for a d-dimensional binary product distribution with mean vector $\boldsymbol{p}$.*

Proof. Let $\tilde{\boldsymbol{p}} := \boldsymbol{p} - \boldsymbol{u}_d$. For each coordinate $i \in [d]$, define

$$T_i := \sum_{k_1 \neq k_2 \in [n]} \left(X_i^{(k_1)} - \frac{1}{2}\right)\left(X_i^{(k_2)} - \frac{1}{2}\right),$$

and

$$T := \frac{1}{n(n-1)} \sum_{i=1}^{d} T_i.$$

Fix $i \in [d]$, and define

$$Z_k := X_i^{(k)} - \frac{1}{2}.$$

Then $Z_1, \ldots, Z_n$ are independent random variables taking values in $\{-1/2, 1/2\}$, with $\mathbb{E}[Z_k] = \tilde{p}_i$. Applying Lemma 20,

$$\mathbb{E}[T_i] = n(n-1)\tilde{p}_i^2$$

and

$$\mathrm{Var}(T_i) \leq \frac{n(n-1)}{8} + n(n-1)(n-2)\tilde{p}_i^2.$$

Summing over coordinates,

$$\mathbb{E}[T] = \|\tilde{\boldsymbol{p}}\|_2^2 = \|\boldsymbol{p} - \boldsymbol{u}_d\|_2^2,$$

and

$$\begin{aligned}\mathrm{Var}(T) &= \frac{1}{n^2(n-1)^2} \sum_{i=1}^{d} \mathrm{Var}(T_i) \\ &\leq \frac{d}{8n(n-1)} + \frac{n-2}{n(n-1)} \|\tilde{\boldsymbol{p}}\|_2^2.\end{aligned}$$

We now analyze the two cases:

Case 1: $p = u_d$. In this case, $\mathbb{E}[T] = 0$ and therefore

$$\mathrm{Var}(T) \leq \frac{d}{8n(n-1)}.$$

By Chebyshev's inequality,

$$\Pr\left(T > \frac{\varepsilon^2}{2}\right) \leq \frac{4\,\mathrm{Var}(T)}{\varepsilon^4} \leq O\left(\frac{d}{n^2\varepsilon^4}\right).$$

Case 2: $\|p - u_d\|_2 \geq \varepsilon$. In this case,

$$\mathbb{E}[T] = \|\boldsymbol{p} - \boldsymbol{u}_d\|_2^2 \geq \varepsilon^2.$$

Again by Chebyshev's inequality,

$$\begin{aligned}\Pr\left(T \leq \frac{\varepsilon^2}{2}\right) &= \Pr\left(\mathbb{E}[T] - T \geq \frac{\mathbb{E}[T]}{2}\right)\\ &\leq \frac{4\,\mathrm{Var}(T)}{\mathbb{E}[T]^2}\\ &\leq O\left(\frac{d}{n^2\varepsilon^4} + \frac{1}{n\varepsilon^2}\right).\end{aligned}$$

Thus, choosing

$$n > C \cdot (\sqrt{d}/\varepsilon^2)$$

for some absolute constant $C > 0$, makes both error probabilities smaller than $1/100$. ◀