Model-agnostic super-resolution in high dimensions

Xi Chen
Columbia University
xichen@cs.columbia.edu

Anindya De
University of Pennsylvania
anindyad@cis.upenn.edu

Yizhi Huang
Columbia University
yizhi@cs.columbia.edu

Shivam Nadimpalli
MIT
shivamn@mit.edu

Rocco A. Servedio
Columbia University
rocco@cs.columbia.edu

Tianqi Yang
Columbia University
tianqi@cs.columbia.edu

November 12, 2025

Abstract

The problem of *super-resolution*, roughly speaking, is to reconstruct an unknown signal to high accuracy, given (potentially noisy) information about its low-degree Fourier coefficients. Prior results on super-resolution have imposed strong modeling assumptions on the signal, typically requiring that it is a linear combination of spatially separated point sources.

In this work we analyze a very general version of the super-resolution problem, by considering completely general signals over the d-dimensional torus $[0,1)^d$; we do not assume any spatial separation between point sources, or even that the signal is a finite linear combination of point sources. The question naturally arises: what can be said about super-resolution in such a general setting? We give two sets of results for two different natural notions of signal reconstruction:

1. **Reconstruction in Wasserstein distance:** Our first set of results is for reconstructing signals with respect to the Wasserstein distance. We give essentially matching upper and lower bounds on the cutoff frequency T and the magnitude κ of the noise for which accurate reconstruction in Wasserstein distance is possible. Roughly speaking, our results here show that for d-dimensional signals, estimates of $\approx \exp(d)$ many Fourier coefficients are necessary and sufficient for accurate reconstruction under the Wasserstein distance.
2. **"Heavy hitter" reconstruction:** We go on to define a new notion of "heavy hitter" reconstruction for non-negative signals (which, after normalization, correspond to probability distributions); this notion essentially amounts to achieving high-accuracy reconstruction of all "sufficiently dense" regions of the distribution. For this notion too, we give essentially matching upper and lower bounds on the cutoff frequency T and the magnitude κ of the noise for which accurate reconstruction is possible. Our results show that—in sharp contrast with Wasserstein reconstruction—accurate estimates of only $\approx \exp(\sqrt{d})$ many Fourier coefficients are necessary and sufficient for heavy hitter reconstruction.

1 Introduction

This paper considers the fundamental signal processing problem of *super-resolution*—recovering fine-scale structure from coarse, noisy measurements—from the perspective of theoretical computer science. A convenient starting point to motivate the problem is the *diffraction limit* from optical physics: when illuminated by light waves of wavelength λ, an optical system cannot resolve features separated by distances much smaller than λ. This idea, dating back to the 1870s [Abb73, Ray79, Spa16], has recently been formalized and studied through the lens of theoretical computer science by Chen and Moitra [CM21], who established statistical and computational limits on diffraction-limited imaging. Since frequency and wavelength are inversely related, this line of work can be summarized as the following high-level idea:

> Resolving increasingly fine details from noisy measurements requires access to correspondingly higher frequencies: shorter wavelengths enable finer resolution.

However, in many settings, systems can only be observed at low frequencies, raising the fundamental question of *super-resolution*: Can we recover a signal even from *bandlimited* measurements? Put differently, if our measurements are limited to frequencies below a certain cutoff, to what extent can we still reconstruct the underlying signal?

The modern mathematical study of super-resolution originates in the seminal work of Donoho [Don92]. Although our focus will be on *high-dimensional* signals, it is instructive to begin with the *one-dimensional* setting introduced by Candès and Fernandez–Granda [CFG13]. Here, the domain is the one-dimensional torus $\mathbb{T}$, which we identify with the interval $[0, 1)$. A signal is modeled as

$$f(x) := \sum_{j=1}^{k} u_j \delta_{t_j}(x)\,, \tag{1}$$

where $\delta_{t_j}(x)$, where $\delta_{t_j}(\cdot)$ is a Dirac delta centered at the point $t_j \in \mathbb{T}$ and $u_j \in \mathbb{R}$ is the corresponding amplitude. In other words, the signal f is a superposition of k *point sources* where the j^{th} point source has location t_j and amplitude u_j. We will sometimes refer to a Dirac delta δ_t as a *spike* located at t.

Next, recall that the Fourier spectrum over the torus $\mathbb{T}$ is indexed by the integers $\mathbb{Z}$. In particular, for a signal f, the Fourier coefficient at frequency $\ell \in \mathbb{Z}$ is given by

$$\widehat{f}(\ell) = \int_{x\in\mathbb{T}} f(x) \cdot \exp\left(2\pi i \ell x\right) dx\,.$$

By definition of the Dirac delta δ_t, for any test function g we have $\int_x g(x)\delta_t(x) = g(t)$. In particular, applying this to a signal of the form (1), we obtain

$$\widehat{f}(\ell) = \sum_{j=1}^{k} u_j \exp(2\pi i \ell t_j)\,.$$

The main question posed by Candès and Fernandez–Granda in [CFG14] was whether bandlimited measurements suffice to recover an unknown signal f, under the promise that f is a superposition of k point sources (i.e., f is as in Equation (1)). They proposed a convex optimization framework for this inverse problem and showed, among other things, that if $\widehat{f}(\ell)$ is known *exactly* for all $\ell \leq k$ (in other words, the signal is observed up to bandlimit k), then their convex program perfectly reconstructs f.

It is worth noting that the ability to recover a k-sparse signal f from the *exact* Fourier coefficients $\{\widehat{f}(\ell)\}_{|\ell|\le k}$ is not new: Prony's method, dating back to 1795 [dP95], achieves the same guarantee. The novelty of [CFG14] lies in showing that such recovery can be accomplished through a convex optimization program. That said, the assumption of having perfect access to the *exact* Fourier data is quite restrictive; in practice, the measurements $\widehat{f}(\ell)$ are inevitably perturbed by noise. This raises the natural question of whether recovery is possible given noisy measurements.

Noisy measurements. Suppose the algorithm no longer has access to the true Fourier coefficients but instead receives $\{\widetilde{f}(\ell)\}_{|\ell|\le T}$ satisfying

$$|\widetilde{f}(\ell) - \widehat{f}(\ell)| \le \kappa \,.$$

In other words, the observed coefficients are perturbed by additive noise of magnitude at most κ. This modification substantially changes the nature of the problem. Indeed, it is easy to see that given $\{\widetilde{f}(\ell)\}_{|\ell|\le T}$, one cannot distinguish between

$$f_1(x) = \delta_0(x) \qquad \text{versus} \qquad f(x) = \frac{1}{2}\cdot\delta_0(x) + \frac{1}{2}\cdot\delta_\Delta(x) \tag{2}$$

for $\Delta \ll \min\{\kappa, T^{-1}\}$. In other words, even deciding whether f consists of one spike or two becomes impossible without imposing an additional *separation* condition between the spikes.

Thus, to recover a signal f from noisy estimates of $\widehat{f}(\ell)$, it appears necessary to impose a separation assumption on the support points of f (spoiler alert: this is precisely the assumption we reexamine in this paper). To formalize what "separation" means on the torus, we first recall the natural distance metric between two points $a, b \in \mathbb{T}$: the *toroidal distance* $\mathrm{d}_{\mathrm{tor}}(a,b)$, which is the shortest distance between a and b when identifying the interval $[0,1)$ with the unit circle (so that 0 and 1 coincide; see Section 2.2 for the formal definition.)

With this definition in hand, Candès and Fernandez–Granda [CFG13] studied the setting where any two supports points t_i, t_j of f satisfy $\mathrm{d}_{\mathrm{tor}}(t_i,t_j) \ge \Delta$. They proved that f can be approximately recovered even in the presence of noise, provided that the support points $t_1,\dots,t_k$ lie on a grid and the bandlimit T is at least $2/\Delta$. (We note that the error metric in [CFG13] is based on the Fejér kernel and is somewhat nonstandard.) Subsequent work [LF16, FG13] removed the grid assumption, although the resulting noise-tolerance guarantees were not directly comparable.

Before proceeding, it will be helpful to standardize some notation and terminology that will be used throughout our discussion:

1. We will denote by T the *bandlimit*; that is, the maximum frequency index (or degree of the Fourier coefficients) available to the algorithm.

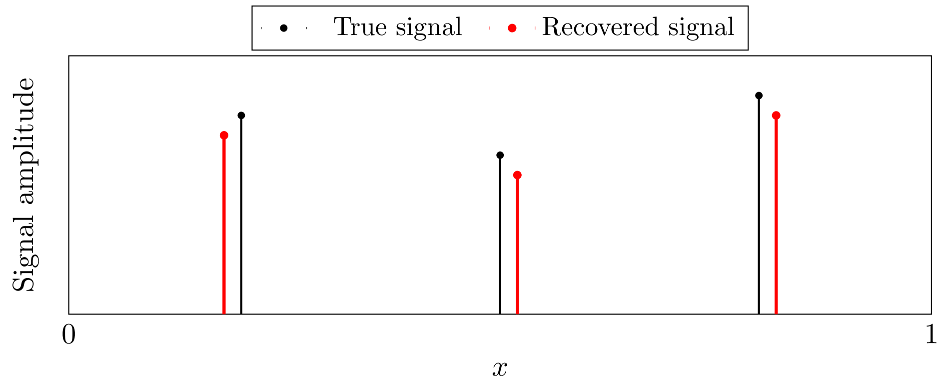
2. We will denote by κ the *noise level per Fourier coefficient*; that is, the true Fourier coefficient $\widehat{f}(\ell)$ and its noisy estimate $\widetilde{f}(\ell)$ satisfy

$$|\widehat{f}(\ell) - \widetilde{f}(\ell)| \le \kappa \,.$$

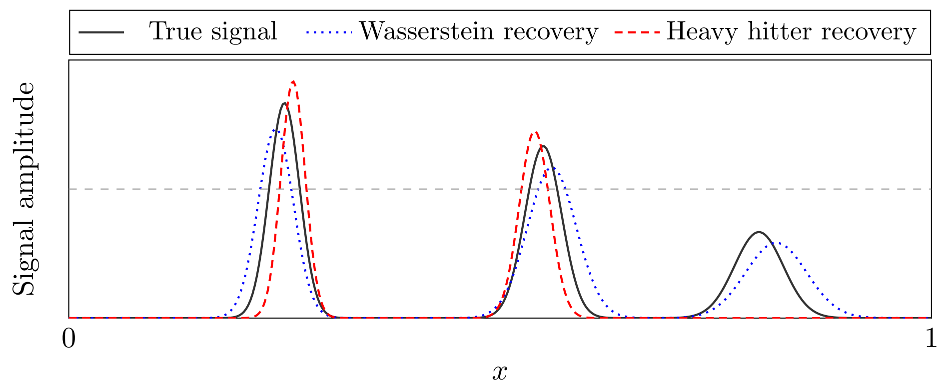
3. If the signal f is supported on a finite set of points, then we will write k for its *sparsity.* When these support points $\{t_i\}_{i\in[k]}$ are separated, we let

$$\Delta := \min_{i,j\in[k]} \mathrm{d}_{\mathrm{tor}}(t_i,t_j)$$

denote the *minimum separation* between them.



(a) An illustration of classical super-resolution as considered in prior works [CFG13, Moi15], which assumes that the underlying signal is a discrete sum of Dirac delta $\delta_t(\cdot)$ signals.



(b) An illustration of *model-agnostic* super-resolution considered in this work, where the input signal may be arbitrary (discrete or continuous). The gray line indicates the recovery threshold for "heavy hitters."

Figure 1: An illustration of the modeling assumptions and recovery objectives in this work compared to prior formulations.

The most relevant prior for in the one-dimensional setting is that of Moitra [Moi15], which was the first to study problems of this kind from the perspective of theoretical computer science. Moitra showed that for minimum separation Δ, if

$$T \geq \frac{1}{\Delta} + 1 \qquad \text{and} \qquad \kappa \leq \mathrm{poly}\left(\frac{1}{\Delta}, \varepsilon\right),$$

then we can efficiently recover a signal f (normalized so that $\sum_j |u_j| = 1$) to error ε; that is, each support location t_i and amplitude u_i can be recovered to within additive error ε. Note that on the one dimensional torus, a minimum separation of Δ implies that the sparsity k of f satisfies $k \leq 1/\Delta$.

This work: Recovery *without* sparsity or separation. The preceding work, including [Moi15], naturally raises the following question:

Can we still recover f *without* any separation or sparsity assumptions?

In particular, we eschew the assumption that f is a mixture of finitely (or even countably) many point sources. Instead, we consider $f : \mathbb{T} \to \mathbb{R}$ to be an arbitrary L^1-integrable function on $\mathbb{T}$.

(Strictly speaking, since we allow for Dirac deltas, f should be viewed as a *signed measure of finite variation* rather than a function. However, the reader will lose little intuition by simply thinking of L^1-integrable functions, and we will freely make this notational simplification; see Section 2.1 for formal definitions.) For normalization, we assume that

$$\int_{x\in\mathbb{T}} |f(x)|\, dx = 1\,.$$

To connect this the previous discussion, observe that when $f = \sum_{j=1}^{k} u_j \delta_{t_j}(x)$ as in (1), we have

$$\int_{x\in\mathbb{T}} |f(x)|\, dx = \sum_{j=1}^{k} |u_j|\,.$$

Of course, in such a general setting it is no longer meaningful to speak of recovering the support of f. Instead, we consider two notions of recovery: one based on the Wasserstein distance, and another that we introduce in this work, called *heavy-hitter recovery*. Additionally, while the discussion thus far has centered on the one-dimensional case, the central goal of this paper is to develop a modeling-agnostic theory of super-resolution in *high dimensions*. The one-dimensional formulation serves primarily as a warm-up: it highlights the key challenges and ideas, but our main results and techniques apply to the high-dimensional setting, where fundamentally new phenomena arise. Before turning to this setting and presenting our results, we briefly survey the most relevant related work.

Related work. While several earlier works have also studied super-resolution under the Wasserstein distance or related metrics [DDP17, SRR18, ETT+21, FL23, KR23], there are several key differences from our work. First, although these works do not assume a separation condition, they all assume a sparsity condition—that the signal is a linear combination of finitely many point sources—whereas we make no such assumption. Second, each of the above papers either restricts attention to the one-dimensional case (while our main results are high-dimensional) or adopts a different measurement model, such as approximate moments, instead of the approximate Fourier coefficients used in our work.

In this context we also mention the recent work of [MMRS25], which, like ours, does not assume sparsity but differs in two respects: it focuses on the one-dimensional setting and considers approximate moments as the measurement model rather than Fourier coefficients. Finally, as we discuss later, our techniques—particularly our use of extremal polynomials—have a substantially different flavor from the existing super-resolution literature.

The problem of super-resolution has also been studied in higher dimensions, for example in [HK15, PP19, LA24]. However, these works continue to assume both that the signal is sparsely supported and that there is a minimum separation between support points. In contrast, our goal is to study super-resolution in a *model-agnostic* manner: we consider completely general signals, without imposing either a sparsity or a separation condition.

1.1 Our results

The central goal of this paper is to develop techniques for super-resolution without assumptions in high dimensions. Our true domain of interest is the the d-dimensional torus $\mathbb{T}^d$ which we identify with $[0,1)^d$. As before, a signal f over $\mathbb{T}^d$ will be normalized to satisfy

$$\int_{x\in\mathbb{T}^d} |f(x)|\, dx = 1\,.$$

The Fourier coefficients of f are now indexed by integer tuples $\ell \in \mathbb{Z}^d$ and are defined as

$$\widehat{f}(\ell) := \int_{x \in \mathbb{T}^d} f(x) \cdot \exp\left(2\pi i \ell \cdot x\right) dx$$

where $\ell \cdot x$ denotes the usual inner product in $\mathbb{R}^d$. For two points $x, y \in \mathbb{T}^d$ the toroidal distance $\mathrm{d_{tor}}(x, y)$ now corresponds to

$$\mathrm{d_{tor}}(x, y) := \sqrt{\sum_{i=1}^{d} \mathrm{d}^2_{\mathrm{tor}}(x_i, y_i)}\,,$$

that is, the Euclidean distance between x and y where each coordinate distance is measured using the one-dimensional toroidal metric.

1.1.1 Wasserstein distance recovery

We begin by recalling the definition of the Wasserstein distance between signals. Given two signals f_1, f_2, their *Wasserstein distance*, denoted by $\mathrm{d_W}(f_1, f_2)$, is defined as

$$\mathrm{d_W}(f_1, f_2) := \sup_{g : \mathbb{T}^d \to \mathbb{R}} \int_x g(x) \cdot (f_1(x) - f_2(x)) dx,$$

where the supremum is taken over all 1-Lipschitz functions g on $\mathbb{T}^d$ (with respect to the toroidal distance) satisfying $\|g\|_\infty \leq 1$.

When f_1 and f_2 are probability distributions (i.e., non-negative signals), this coincides with the usual Wasserstein distance (also known as the *earthmover distance*) between probability distributions; see Section 2. For general signed measures, this extension has been studied previously in [Han99, PRT23].

Recovery in one dimension. As a warm-up, we first analyze the one-dimensional ($d = 1$) case of our Wasserstein distance recovery result, Theorem 10. We establish the following: For any $\varepsilon > 0$, given κ-noisy Fourier coefficients $\{\widetilde{f}(\ell)\}_{|\ell| \leq T}$ with bandlimit $T = \Theta(1/\varepsilon)$ and noise level $\kappa = O(\varepsilon/\log(1/\varepsilon))$, we show that f can be recovered in Wasserstein distance up to error ε. More precisely, we give an efficient algorithm (see Section 4) that outputs a function $g : \mathbb{T} \to \mathbb{R}$ satisfying $\mathrm{d_W}(f, g) \leq \varepsilon$.

Moreover, our guarantees are nearly optimal in a strong sense. Theorem 11 shows that for the Wasserstein reconstruction notion considered here, the dependence on the bandlimit $T = \Theta(1/\varepsilon)$ is optimal up to constant factors even in the noiseless case ($\kappa = 0$) where the Fourier coefficients are known exactly. Furthermore, a simple counterexample (Observation 16 demonstrates that if the noise level κ is allowed to be as large as $O(\varepsilon)$, then even access to *all* Fourier coefficients (i.e., infinite bandlimit) does not permit ε-accurate Wasserstein recovery.

It is instructive to see how the Wasserstein error guarantee in Theorem 10 circumvents the need for a separation condition. For example, returning to the signals from Equation (2), note that the two signals f_1 and f_2 there are $\Delta/2$-close in Wasserstein distance. Intuitively, the Wasserstein metric blurs the distinction between nearby points, allowing for meaningful approximate recovery even when spikes are arbitrarily close.

Of course, if the support points are well separated, then accurate recovery in Wasserstein distance automatically yields accurate estimates of the true support locations $\{t_j\}$ and amplitudes $\{u_j\}$. In this sense, the Wasserstein criterion provides a *soft interpolation* between rigid separation-based guarantees and meaningful recovery in the absence of any separation assumption.

High-dimensional Wasserstein recovery. We now turn to the high-dimensional setting. Our main result, Theorem 10, whose one-dimensional special case was discussed above, can be summarized as follows: For any $\varepsilon > 0$, given noisy Fourier coefficients $\{\widetilde{f}(\ell)\}_{\|\ell\|_\infty \le T}$ with

$$T = \Theta\left(\frac{\sqrt{d}}{\varepsilon}\right) \qquad \text{and} \qquad \kappa = \Theta\left(\frac{\varepsilon}{\sqrt{d}}\right)^d,$$

we can recover f to Wasserstein error ε. Moreover, we give an algorithm which, for constant dimension d, runs in time polynomial in the size of its input and outputs a function $g : \mathbb{T}^d \to \mathbb{R}$ satisfying $\mathrm{d_W}(f, g) \le \varepsilon$ (see Section 4).

This extends our one-dimensional result to arbitrary dimension d, but at a cost: the admissible noise level κ decays exponentially with d. Moreover, while the bandlimit T grows only as $\sqrt{d}/\varepsilon$, the number of Fourier coefficients $\widehat{f}(\ell)$ satisfying $\|\ell\|_\infty \le T$ (and hence the number of measurements required) increases exponentially in d.

Unfortunately, both of these dependencies turn out to unavoidable. In particular, Theorem 11 shows, via a simple construction, that recovering a signal f to Wasserstein error $\pm\varepsilon$ requires access to Fourier coefficients up to a bandlimit $T \approx \sqrt{d}/(2\varepsilon)$ even in the noiseless case ($\kappa = 0$). This naturally raises the question of whether one can trade off the bandlimit T against the noise level κ. However, our next lower bound, Theorem 13, rules out this possibility as well: even if *all* Fourier coefficients are available (i.e., $T = \infty$), achieving Wasserstein error still requires the noise parameter κ to roughly satisfy $\kappa \le \varepsilon^{d/4}$.

The takeaway from these results is as follows: On the positive side, we obtain nearly tight upper and lower bounds on the required bandlimit T and noise level κ for Wasserstein recovery in high dimensions. However, the noise level must necessarily grow as $\kappa = \varepsilon^{\Theta(d)}$, and the bandlimit (in ℓ_∞ norm) must scale roughly as $T = \Theta(\sqrt{d}/\varepsilon)$. Consequently, Wasserstein recovery in dimension requires $\approx (\sqrt{d}/\varepsilon)^d$ Fourier coefficients, each accurate to within $\kappa \approx \varepsilon^{\Theta(d)}$. In short, the curse of dimensionality—an exponential dependence on d—is unavoidable for recovery in Wasserstein distance.

1.1.2 Heavy hitter recovery

A closer examination of our lower-bound constructions (Theorems 11 and 13) for Wasserstein recovery reveals a common feature: in the hard instances, the mass of f is distributed rather evenly across the domain. In contrast, many practical scenarios involve signals whose mass is concentrated in a few localized regions—what we might informally call the "spikes" or high-density regions of f. This observation naturally leads to the following question: borrowing terminology from data streaming [Woo16], can we recover the "heavy hitters" of f—the dominant regions carrying most of its mass—while avoiding the curse of dimensionality?

Before proceeding, it is useful to note that from this point onward we restrict attention to arbitrary *nonnegative* signals. This restriction is both natural and essential given our goal of not requiring any separation assumption. Indeed, if negative values are allowed and there is no minimal separation assumption, then it is impossible to distinguish the absence of a spike at x_0 from the two-spike signed signal

$$\frac{1}{2} \cdot \delta_{x_0} - \frac{1}{2} \cdot \delta_{x_0 + \Delta},$$

(where Δ may be arbitrarily small) using only Fourier measurements, and this means that our heavy-hitter objective of "recovering the spikes" cannot be achieved. Indeed, since the intuitive goal of heavy-hitter recovery is to "find the spikes," i.e., identify regions where the signal's mass is concentrated, it is most natural to study this problem for nonnegative signals, where such regions correspond to high-density areas of a distribution.

For the remainder of this section, we thus restrict our attention to non-negative signals. In particular, we use D to denote non-negative signals, which we interpret as distributions (since $\int |D| = \int D = 1$), and reserve f for general, possibly signed, signals.

Since the *raison d'être* of this paper is to be model-agnostic, we continue to make no assumptions about the support of D; in particular, D need not have finite or even countable support. Henceforth, D will denote a general pobability measure over $\mathbb{T}^d$, and our goal will be to design recovery guarantees in terms of a new metric that captures similarity of "spike-like regions." With this motivation, we now introduce the *heavy-hitter distance* between distributions.

The heavy hitter distance. We will use the following terminology: for any point $x \in \mathbb{T}^d$, any $\tau \geq 0$, define

$$\mathrm{Ball}(x, \tau) := \{y \mid \mathrm{d}_{\mathrm{tor}}(x, y) \leq \tau\}.$$

For any distribution D over $\mathbb{T}^d$, we slightly abuse the notation and define

$$D(\mathrm{Ball}(x, \tau)) := \int_{\mathrm{d}_{\mathrm{tor}}(x,y) \leq \tau} D(y) dy,$$

i.e., $D(\mathrm{Ball}(x, \tau))$ is the probability that D puts in the toroidal-distance ball of radius τ around x.

The heavy hitter distance between signals is parameterized by a scale parameter $\varepsilon_{\mathrm{dist}} > 0$, and is defined as follows:

Definition 1 (Heavy hitter distance)**.** Given a scale parameter $0 < \varepsilon_{\mathrm{dist}} < 1$, the *heavy hitter distance at scale* $\varepsilon_{\mathrm{dist}}$ between two distributions D_1, D_2 over $\mathbb{T}^d$, written $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D_1, D_2)$, is the smallest value $\varepsilon \geq 0$ such that the following holds: For every $0 \leq \tau \leq \varepsilon_{\mathrm{dist}}$ and every $x \in \mathbb{T}^d$,

$$D_1(\mathrm{Ball}(x, \tau)) \leq D_2(\mathrm{Ball}(x, \tau + \varepsilon_{\mathrm{dist}})) + \varepsilon\,, \quad \text{and}$$
$$D_2(\mathrm{Ball}(x, \tau)) \leq D_1(\mathrm{Ball}(x, \tau + \varepsilon_{\mathrm{dist}})) + \varepsilon\,.$$

We remark that the term "heavy hitter distance" should not be interpreted literally as meaning that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(\cdot, \cdot)$ satisfies the triangle inequality; rather, as we discuss later, it is a way of measuring the error between two distributions vis-à-vis the "heavy hitter-like" regions of the distributions. Intuitively, having $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(f_1, f_2) \leq \varepsilon$ means that for any "small" ball, the "mass" that f_1 and f_2 put in the ball are essentially the same, provided that we are allowed to dilate the ball by an additive $\varepsilon_{\mathrm{dist}}$ and adjust the probability mass additively by ε.

Remark 2 (Heavy hitter distance versus Wasserstein distance)**.** It is not difficult to show that having small heavy hitter distance is essentially a weaker condition than having small Wasserstein distance — more precisely, having small Wasserstein distance between two distributions D_1, D_2 implies that the heavy hitter distance (at a suitable scale $\varepsilon_{\mathrm{dist}}$) must also be small. See Section A for a detailed statement and its simple proof.

Remark 3 (Heavy hitter distance versus Lévy-Prokhorov)**.** The reader may have noticed the similarity of this definition with the Lévy-Prokhorov metric (though in the Lévy-Prokhorov metric, the role that balls play in the heavy hitter distance is played by arbitrary measurable sets). This naturally suggests the following question: can we achieve the positive results that are presented below for $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}$ for the Lévy-Prokhorov metric? Unfortunately, the lower bounds for Wasserstein distance (Theorems 11, 13 and 25) that we mentioned earlier are easily seen to hold with the exact same parameters for the Lévy-Prokhorov metric as well.

Next, we discuss the relationship between the $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(\cdot,\cdot)$ distance notion and the presence of spikes in our distributions. Suppose that D has a spike of amplitude $\geq 2\varepsilon$ at x_0; in more detail, suppose that $D = D_1 + D'$ where $D_1 = u_0\delta_{x_0}$, with $u_0 \geq 2\varepsilon$ (think of ε as a small but positive constant), and suppose moreover that this spike is "isolated" in the sense that not only does D have no spikes within (toroidal) distance Δ of x_0, but moreover $D(x) \leq C$ for all $x \neq x_0$ such that $\mathrm{d}_{\mathrm{tor}}(x, x_0) \leq \Delta$ (think of Δ too as a small constant bounded below 1). This "isolation" condition implies that while the distribution D may be non-zero in the neighborhood of x_0, for $\tau \leq \Delta$ we have

$$\int_{\tau \geq \mathrm{d}_{\mathrm{tor}}(x,x_0) > 0} D(x)dx \leq C \cdot \tau^d \,,$$

which is much less than ε (because of the exponential decay in d). Consequently, for $\tau \leq \Delta$, we have that $D(\mathrm{Ball}(x_0, \tau)) \leq u_0 + C \cdot \tau^d$. It follows that if we can do recovery in the heavy hitter distance, i.e., if we can obtain another distribution D_2 such that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D, D_2) \leq \varepsilon$ for some $\varepsilon_{\mathrm{dist}} \leq \Delta/2$, it would imply that $D_2(\mathrm{Ball}(x_0, \varepsilon_{\mathrm{dist}})) \leq D(\mathrm{Ball}(x_0, 2\varepsilon_{\mathrm{dist}})) + \varepsilon \lesssim u_0 + \varepsilon$ and $D_2(\mathrm{Ball}(x_0, \varepsilon_{\mathrm{dist}})) \geq D(\mathrm{Ball}(x_0, 0)) - \varepsilon = u_0 - \varepsilon \geq \varepsilon$. Thus, given D_2, we can achieve the following goals:

- We can locate the position x_0 of the spike in D up to an additive (toroidal) distance of $\varepsilon_{\mathrm{dist}}$;
- We can compute the amplitude u_0 of the spike in D up to an additive error of roughly ε.

Note that these guarantees are meaningful only when the amplitude of the spike u_0 exceeds ε in absolute value. In the language of data streaming [Woo16], we may think of the spike at x_0 as being an "ε-heavy hitter."

Finally, we note that recovery under the heavy hitter error metric can be meaningful even without a true "spike" (δ-function) at x_0. In particular, instead of having $D_1 = u_0\delta_{x_0}$, if D_1 instead satisfies

$$\int_{\mathrm{d}_{\mathrm{tor}}(x,x_0) \leq \varepsilon_{\mathrm{dist}}/2} D_1(x)dx = u_0 \,,$$

that is, D_1 has mass u_0 in a ball of radius $\varepsilon_{\mathrm{dist}}/2$ around x_0 (an "approximate spike near x_0 at the scale of $\varepsilon_{\mathrm{dist}}/2$"), then our output guarantees stay meaningful.

Our results for heavy hitter recovery. Theorem 19 gives our main positive result for heavy-hitter distance recovery. In essence, the theorem states the following: Let D be an unknown distribution over $\mathbb{T}^d$ (so $D(x) \geq 0$ and $\int_{x \in \mathbb{T}^d} D(x))dx = 1$), and fix error parameters $\varepsilon, \varepsilon_{\mathrm{dist}} > 0$. Given noisy Fourier coefficients $\{\tilde{D}(\ell)\}_{\|\ell\|_1 \leq T}$ each corrupted by additive noise of magnitude at most κ, where

$$T = O\left(\sqrt{d} \cdot \log(1/\varepsilon) \cdot \varepsilon_{\mathrm{dist}}^{-1}\right) \qquad \text{and} \qquad \kappa = \exp\left(-\sqrt{d}\log d \cdot \log(1/\varepsilon) \cdot \varepsilon_{\mathrm{dist}}^{-1}\right)$$

we can recover a function f' such that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D, D') \leq \varepsilon$.

Several remarks are in order. First, for constant ε and $\varepsilon_{\mathrm{dist}}$, the noise level κ is (inverse) *subexponential* in the dimension d, scaling as $2^{-\tilde{O}(\sqrt{d})}$. Second—and crucially—the bandlimit is measured in terms of the ℓ_1 norm (rather than the ℓ_∞ norm, as in our Wasserstein results) of the coefficient vector ℓ. This is a significant improvement, since for constant ε and $\varepsilon_{\mathrm{dist}}$, only $d^{O(T)} \approx 2^{\widetilde{O}(\sqrt{d})}$ Fourier coefficients are required. Thus, both the number of Fourier measurements and the required precision are subexponential in the ambient dimension d. To our knowledge, these are the first results for super-resolution without any sparsity assumptions whose dependence on d is subexponential.

Finally, it is natural to ask whether these bounds can be quantitatively improved. Unfortunately, the answer is negative: in Theorem 24, we show that for constant $\varepsilon_{\text{dist}}$ and ε, our dependencies on both the bandlimit T and noise level κ are essentially optimal.

1.2 Techniques

We now turn to a technical overview of our results.

1.2.1 Wasserstein recovery positive result

The main idea underlying our positive result for Wasserstein recovery (Theorem 10) is straightforward. Restating the result, we wish to show that if the Wasserstein distance between two signals f_1 and f_2 is large, say $\mathrm{d_W}(f_1, f_2) \geq \varepsilon$, then there must be a Fourier coefficient ℓ (with $\|\ell\|_\infty \leq T$) such that $|\widehat{f}_1(\ell) - \widehat{f}_2(\ell)| > \kappa$. Establishing this immediately implies that if the Fourier coefficients $\{\widetilde{f}(\ell)\}_{\|\ell\|_\infty \leq T}$ are known up to additive error κ, one can recover f within Wasserstein distance ε.

To establish this, we follow a standard analytic approach and introduce a *mollifier* J. A mollifier is a distribution "tightly centered" around; in our setting, we quantify this by

$$\eta := \mathop{\mathbf{E}}_{\boldsymbol{x} \sim J}[\mathrm{d_{tor}}(\boldsymbol{x}, 0^d)]\,.$$

The key tradeoff for a mollifier is between its localization—how concentrated it is near 0^d (captured by η)—and its bandlimit, the smallest T for which $\widehat{J}(\ell) = 0$ whenever $\|\ell\|_\infty > T$. For a detailed account of mollifiers and their properties, see [Rud91].

Assuming for now a suitable mollifier, in the rest of the proof we define

$$f_1' := f_1 * J \qquad \text{and} \qquad f_2' := f_2 * J\,.$$

Since J is η-concentrated around 0^d (as described above), it is straightforward to verify that if $\eta \leq \varepsilon/3$, we will have $\mathrm{d_W}(f_1', f_2') \geq \varepsilon/3$. On the Fourier side, convolution corresponds to pointwise multiplication:

$$\widehat{f_1'}(\ell) = \widehat{f}_1(\ell) \cdot \widehat{J}(\ell) \qquad \text{and} \qquad \widehat{f_2'}(\ell) = \widehat{f}_2(\ell) \cdot \widehat{J}(\ell)$$

Our argument further uses three more simple but crucial ingredients from Fourier analysis:

- Plancherel's identity, which allows us to relate the inner product of two signals to the inner product of their Fourier transforms;
- The bandlimitedness of the mollifier J, which ensures that $\widehat{f_1'}(\ell) = \widehat{f_2'}(\ell) = 0$ for all $\|\ell\|_\infty > T$; and
- The decay of Fourier coefficients for Lipschitz functions, namely that $|\widehat{g}(\ell)| = O(\|\ell\|_2^{-1})$ for Lipschitz g (Fact 4).

Using the three ingredients above, one can show that a lower bound on $\mathrm{d_W}(f_1', f_2')$ necessarily implies the existence of a frequency ℓ with $\|\ell\|_\infty \leq T$ for which $|\widehat{f}_1(\ell) - \widehat{f}_2(\ell)|$ is large. The resulting quantitative bounds depend on the choice of mollifier J. For our analysis, we employ *Jackson's kernel* $J_{d,n}$ (in d dimensions, parameterized by the integer n; see Definition 5 for the precise definition). Substituting this kernel into the framework outlined above yields the quantitative guarantees stated in Theorem 10.

1.2.2 Recovery in heavy hitter distance: The positive result

For our positive result for heavy hitter distance recovery (Theorem 19), we will again construct a suitable mollifier p (also referred to as a "bump function"). Qualitatively, $p(\cdot)$ will have three properties:

- $p(x) \in [0, 1]$ for all $x \in \mathbb{T}^d$;
- $p(0^d) = 1$ and $p(x)$ remains close to 1 when $\mathrm{d_{tor}}(x, 0^d)$ is small; and
- $p(x)$ is close to 0 when $\mathrm{d_{tor}}(x, 0^d)$ is large.

We defer the precise technical construction to later sections, but note that we are now content with a weaker guarantee than that achieved by Jackson's kernel $J_{d,n}$ used in the proof of Theorem 10. In particular, if we view $p(\cdot)$ as a probability distribution (after suitable normalization), then for $\boldsymbol{x} \sim p$ it may be the case that $\mathbf{E}[\mathrm{d_{tor}}(\boldsymbol{x}, 0^d)]$ is large; hence such a p would not suffice for Wasserstein recovery. Nevertheless, this weaker mollifier turns out to be entirely adequate for recovery in heavy-hitter distance.

In Lemma 22, we construct such a mollifier p using tools from approximation theory and complexity-theoretic polynomial constructions that pointwise approximate symmetric Boolean functions over $\{0, 1\}^d$. The key improvement over Jackson's kernel $J_{d,n}$ (in fact, over any mollifier suitable for Wasserstein recovery) is that p is a trigonometric polynomial with *total* degree scaling only $\sqrt{d}$, where d is the ambient dimension. (Note that this is in sharp contrast to the mollifier $J_{d,n}$ used in Section 3 for Wasserstein distance, where the total degree scales at least linearly in d; recall that the bandwidth constraint for Wasserstein reconstruction is that each Fourier coefficient $\ell = (\ell_1, \ldots, \ell_d)$ that is provided has *individual* degree $\|\ell\|_\infty$ at most T.)

Our construction of $p(\cdot)$ proceeds in two stages:

1. In stage one, we construct a simple constant degree trigonometric polynomial which is 1 at 0^d but is at most ε^2/d once $\mathrm{d_{tor}}(x, 0^d) \geq \varepsilon$.
2. In stage two, we compose this with a suitable univariate polynomial q such that (a) $q(x) \in [0, 1]$ for $x \in [0, 1]$; (b) $q(0) = 1$ and $q(x)$ is close to 1 when $x > 0$ is close to 0; and (c) $q(x)$ is close to 0 when $x \in [0, 1]$ is large. Crucially, this can be achieved with a polynomial q of degree $\approx \sqrt{d}$.

The key step above is stage two. As discussed earlier, the square-root savings in degree arise from results in polynomial approximation theory—specifically, Paturi's theorem [Pat92] on low-degree polynomial approximations of symmetric Boolean functions. This improvement in degree is precisely what allows our algorithm in Theorem 19 to tolerate noise parameters as large as $\kappa = \exp(-\sqrt{d} \log d)$ while requiring Fourier measurements of total degree only $O(\sqrt{d})$ (for fixed $\varepsilon, \varepsilon_{\mathrm{dist}} > 0$). Finally, we note that although extremal polynomials with optimal dependence on degree have been widely used in theoretical computer science and machine learning theory [KS01, HMPW08, She20], we are not aware of prior work in the signal processing or super-resolution context using them.

1.2.3 Lower bounds

We now give a brief discussion of the lower bounds in this paper.

1. Theorem 11 shows that even given the exact Fourier coefficients, if the bandlimit T (on individual degree) satisfies $T \lesssim \sqrt{d}/(2\varepsilon)$, then ε-accurate recovery in Wasserstein distance is impossible. This lower bound holds even for non-negative signals, i.e. probability distributions;

it is achieved by considering two distributions which are uniform on grids of width $1/T$ where the two grids are $1/T$-shifts of each other (for $T \approx \sqrt{d}/2\varepsilon$). As the distributions are uniform on grids, calculating their Fourier coefficients and showing that they are identical up to bandlimit T is a straightforward exercise.

2. Theorem 13 shows that even given all (infinitely many) Fourier coefficients to error $\kappa = \varepsilon^{0.249d}$, the minimum error in Wasserstein recovery must be at least ε; like Theorem 11, this holds even for non-negative signals. (More precisely, it establishes the existence of two distributions D_1 and D_2 such that $|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)| \leq \kappa$ for all $z \in \mathbb{Z}^d$, yet $\mathrm{d_W}(D_1, D_2) > \varepsilon$.) The lower bound construction uses two distributions each of which is the convolution of a suitable Jackson's kernel with a discrete distribution that is uniform on a randomly chosen discrete point set of size $(\approx)(1/\varepsilon)^d$. The proof of correctness uses the probabilistic method in tandem with the fact that convolution with Jackson's kernel annihilates higher-order Fourier coefficients.

3. Our final lower bound for recovery in heavy hitter distance, Theorem 24, is arguably the most novel lower bound proof in this paper. The heavy lifting in the proof of Theorem 24 occurs in the proof of an intermediate result, Theorem 25, where we construct two distributions P_1 and P_2 over the hypercube $\{0,1\}^d$ with two special properties. To explain these properties, we recall that any function $g : \{0,1\}^d \to \mathbb{R}$ admits a "Fourier expansion" (sometimes called the Walsh-Fourier expansion), meaning that it can be expressed as

$$g(x) = \sum_{S \subseteq [d]} \widehat{g}(S)\chi_S(x), \quad \text{where} \quad \widehat{g}(S) := \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{0,1\}^d}[g(\boldsymbol{x})\chi_S(\boldsymbol{x})] \quad \text{and} \quad \chi_S(x) = \prod_{i \in S} x_i.$$

We refer the reader to [O'D14] for background on this rich theory and its many applications in theoretical computer science and mathematics. Turning back to the distributions P_1 and P_2 (which we view as non-negative functions over $\{0,1\}^d$ with $\sum_x P_i(x) = 1$), we show that they have the following special properties:

(a) P_1 and P_2 differ significantly in the amount of mass they put on the point 1^d; more precisely, $|P_1(1^d) - P_2(1^d)| \geq 2\varepsilon$.

(b) When we view P_1 and P_2 as functions over $\{0,1\}^d$, their low-degree Fourier coefficients are very close to each other; more precisely, for all $|S| \leq c\sqrt{d/\log(1/\varepsilon)}$, we have that $|\widehat{P_1}(S) - \widehat{P_2}(S)| \leq 2^{-d} \cdot 2^{\Omega(\sqrt{d \log(1/\varepsilon)})}$.

We then map these distributions P_1 and P_2 over $\{0,1\}^d$ to distributions D_1 and D_2 over $\mathbb{T}^d$ via a simple embedding that achieves the desiderata of Theorem 24.

As the reader can see, the crux of this argument is the construction of the distributions P_1 and P_2. The existence of such a pair of distributions is argued using constructions of an extremal polynomial [BEK99] which, roughly speaking, has bounded coefficients but also has a zero of very high multiplicity at 1. (Interestingly, the arguments in [BEK99] giving such a real polynomial make essential use of complex analysis.) We omit further details here, but we note that these polynomials and their behavior was crucial in the study of a different "inverse statistical" problem, namely the *population recovery problem*, which has been extensively studied over the past decade or so in theoretical computer science [LZ17, DRWY12, MS13, PSW17, DOS20, LZ17]. As we elaborate below, the conceptual link to the population recovery literature played a central role in the development of this work.

1.3 Connections with population recovery

As mentioned above, the problem of population recovery has been studied in detail in theoretical computer science in recent years. The two most relevant works for us are the independent and concurrent works [PSW17, DOS20] which studied the problem of *recovering heavy hitters* from samples on the hypercube that have been contaminated with "bit-flip noise." In this problem, there is an unknown distribution $\mathcal{D}$ over $\{0,1\}^n$. Given a parameter $\rho \in (0,1]$, we define the "noisy distribution" $\mathcal{D}_\rho$ as follows: to generate a sample from $\mathcal{D}_\rho$, we first sample $\boldsymbol{x} \sim \mathcal{D}$ and then flip every bit of $\boldsymbol{x}$ independently with probability $(1-\rho)/2$. The goal is to recover (in some sense) the unknown distribution $\mathcal{D}$ from the noisy samples.

Both [PSW17] and [DOS20] showed that for any constants $\rho < 1$ and $\varepsilon > 0$, there is an algorithm which can recover the ε-heavy hitters of $\mathcal{D}$ (i.e., the points $x \in \{0,1\}^n$ such that $\mathcal{D}(x) \geq \varepsilon$) with sample complexity $\approx \exp(n^{1/3})$. Furthermore, for such points, the algorithm also outputs $\mathcal{D}(x)$ up to $\pm\varepsilon/2$. In other words, the algorithm outputs the exact location and the approximate mass of each of the heavy hitters.

To connect this to the current work, note that similar to [PSW17, DOS20], we also seek to recover the location of the "heavy hitters" (i.e. the spikes or approximate spikes). Of course, given that our domain $\mathbb{T}^d$ is continuous, we can only recover the locations of the "heavy hitters" up to some "small ball" (this is reflected in our use of the $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(\cdot,\cdot)$ reconstruction criterion). A second point of divergence between the current work and [PSW17, DOS20] is that while we assume access to noisy *Fourier coefficients*, [PSW17, DOS20] assume access to noisy *samples*. These access models are related in that noisy Fourier coefficients can be computed from noisy samples: in particular, for an unknown distribution $\mathcal{D}$, its Fourier coefficient $\mathcal{D}(S)$ can be computed to within additive error $\pm 2^{-d} \cdot \delta$ using $\approx (1/\delta^2) \cdot (1/\rho)^{|S|}$ noisy samples from $\mathcal{D}_\rho$. So we can estimate Fourier coefficients from noisy samples (but the sample complexity deteriorates exponentially in the degree of the Fourier coefficient S being estimated). However, we cannot go in the other direction and simulate noisy samples given noisy Fourier coefficients; intuitively, this is why [PSW17, DOS20] are able to achieve an $\exp(d^{1/3})$ sample complexity while our complexity (for our Fourier coefficient based approach) must scale as $\exp(d^{1/2})$.

We close this discussion by remarking that while both the domains considered in [PSW17, DOS20] are different from our setting, as well as the form of access via which the algorithm gets access to the underlying distribution, the [PSW17, DOS20] results on subexponential-time recover of heavy hitters over the hypercube served as an initial inspiration for our high-dimensional heavy hitter distance results for super-resolution.

2 Preliminaries

Throughout this paper, we write $\mathbb{T} := [0,1)$ to denote the torus and $\mathbb{T}^d := [0,1)^d$ to denote the d-dimensional torus.

2.1 Signals

In this paper, the term *signal* will refer to a *signed measure of finite total variation* on $\mathbb{T}^d$, normalized so that its total variation is 1. Equivalently, by a mild (and rather standard) abuse of notation, we treat such signals as L^1-integrable functions $f : \mathbb{T}^d \to \mathbb{R}$ satisfying $\int_{x \in \mathbb{T}^d} |f(x)|\, dx = 1$, while allowing discrete components such as Dirac deltas. This convention lets us work uniformly with both continuous and discrete signals. Note that positive (non-negative) normalized signals correspond to densities of probability distributions over $\mathbb{T}^d$.

For formal background on finite signed measures and total variation, see, e.g., Chapter 6 of [Rud87] or Chapter 3 of [Fol99].

2.2 Fourier analysis over $\mathbb{T}^d$

Recall that over the domain $\mathbb{T}^d$, for $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{Z}^d$, the ℓ^{th} Fourier coefficient of a signal f is given by

$$\widehat{f}(\ell) = \int_{x\in\mathbb{T}^d} f(x) \cdot e^{2\pi i(\ell\cdot x)} dx.$$

For a "point mass" at $r \in \mathbb{T}^d$ (corresponding to the Dirac delta function $f(x) = \delta_r(x) = \delta(x-r)$), the Fourier transform is given by

$$\widehat{f}(\ell) = \int_{x\in\mathbb{T}^d} \delta(x-r) \cdot e^{2\pi i(\ell\cdot x)} dx = e^{2\pi i(\ell\cdot r)}.$$

We recall *Plancherel's theorem*, which says that if $f \in L^2$, then

$$\int_{x\in\mathbb{T}^d} |f(x)|^2 dx = \sum_{\ell\in\mathbb{Z}^d} |\widehat{f}(\ell)|^2.$$

We further recall that given signals f_1 and f_2,, their *convolution*, written $f_1 * f_2$, is

$$(f_1 * f_2)(x) = \int_{y\in\mathbb{T}^d} f_1(y) f_2(x-y) dy,$$

and that if f_1, f_2 are L^1-integrable then we have $\widehat{f_1 * f_2}(\ell) = \widehat{f_1}(\ell) \cdot \widehat{f_2}(\ell)$ for all $\ell \in \mathbb{Z}^d$.

2.3 Distance and Wasserstein distance over the torus

Recall that the "toroidal distance" $\mathrm{d}_{\mathrm{tor}}$ between two points a_1, b_1 in the one-dimensional torus $\mathbb{T}$ (also known as the Lee metric) is

$$\mathrm{d}_{\mathrm{tor}}(b_1, a_1) = \mathrm{d}_{\mathrm{tor}}(a_1, b_1) := \min\left\{b_1 - a_1, a_1 + (1-b_1)\right\} \quad \text{where } 0 \le a_1 < b_1,$$

and that the Euclidean toroidal distance between two points a, b in the d-dimensional torus $\mathbb{T}^d$ is

$$\mathrm{d}_{\mathrm{tor}}(a, b) = \sqrt{\mathrm{d}_{\mathrm{tor}}(a_1, b_1)^2 + \dots + \mathrm{d}_{\mathrm{tor}}(a_d, b_d)^2}.$$

We recall that if $f, g : \mathbb{T}^d \to \mathbb{R}_{\ge 0}$ are non-negative valued functions with $\int_{x\in\mathbb{T}^d} f(x)dx = \int_{x\in\mathbb{T}^d} g(x)dx = 1$, corresponding to the pdfs of two probability distributions, then the *Wasserstein distance* $\mathrm{d}_{\mathrm{W}}(f, g)$ between f and g is the minimum "transportation cost" of turning f into g, i.e. the minimum, over all couplings $(\boldsymbol{X}, \boldsymbol{Y})$ of random variables $\boldsymbol{X} \sim f$, $\boldsymbol{Y} \sim g$, of $\mathbf{E}[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{X}, \boldsymbol{Y})]$. As is well known and was mentioned earlier, by Kantorovich-Rubenstein duality an alternative definition is that

$$\mathrm{d}_{\mathrm{W}}(f, g) := \sup_{h:\mathbb{T}^d\to\mathbb{R}} \int_x h(x) \cdot (f(x) - g(x)) dx, \tag{3}$$

where h ranges over all 1-Lipschitz (vis-à-vis toroidal distance) functions over $\mathbb{T}^d$ satisfying $\|h\|_\infty \le \sqrt{d}$. (Note that if f and g are pdfs of probability distributions, then the condition $\|h\|_\infty \le \sqrt{d}$ is immaterial since Equation (3) is unchanged when $h(x)$ is replaced by $h(x) + c$ for any $c \in \mathbb{R}$.)

The above characterization extends in a natural way to L^1-integrable functions. In more detail, given two $L^1$1-integrable functions with $\int_{x\in\mathbb{T}^d}|f(x)|dx = \int_{x\in\mathbb{T}^d}|g(x)|dx = 1$, we define the Wasserstein distance between f and g using Equation (3) above, where we stress again that h ranges over all 1-Lipschitz functions over $\mathbb{T}^d$ satisfying $\|h\|_\infty \le \sqrt{d}$. (We remark that in this more general context in which f and g need not integrate to the same value, the condition that $\|h\|_\infty \le \sqrt{d}$ is essential for the definition to make sense.) This notion has been previously considered in the literature, see e.g. [Han99, PRT23].

We will use the following fact, which gives a bound on the Fourier coefficients of 1-Lipschitz functions (see, e.g., [HSSV21] for the simple argument establishing this):

Fact 4. Let f be any 1-Lipschitz function over $\mathbb{T}^d$ and let $0^d \ne \ell \in \mathbb{Z}^d$. Then it holds that

$$\left|\widehat{f}(\ell)\right| \le \frac{1}{\|\ell\|_2}.$$

2.4 Jackson's kernel

We now define the Jackson's kernel and present some of its basic properties. We will later use this kernel to prove our upper bound and lower bound under Wasserstein distance. The proofs are given in Section B.

Definition 5 (Jackson's kernel)**.** For any integer $n \ge 1$, the 1-dimensional *Jackson's kernel* over the torus $\mathbb{T}$ is defined by

$$J_n(x) := \alpha_n \cdot \frac{\sin^4(\pi n x)}{\sin^4(\pi x)} \qquad \text{(and naturally } J_n(0) := \alpha_n n^4\text{)},$$

where α_n is a constant given by

$$\alpha_n := \frac{3}{n(2n^2+1)}.$$

The d-dimensional Jackson's kernel over $\mathbb{T}^d$ is defined by

$$J_{d,n}(x_1,\ldots,x_d) := \prod_{i=1}^{d} J_n(x_i) = \alpha_n^d \cdot \prod_{i=1}^{d} \frac{\sin^4(\pi n x_i)}{\sin^4(\pi x_i)}.$$

Note that the function $J_n(x)$ viewed as a function over $\mathbb{R}$ is periodic.

The following proposition shows that the Jackson's kernel defined above integrates to 1, so the $J_{d,n}$ defines a distribution over $\mathbb{T}^d$.

Proposition 6.

$$\int_{x\in\mathbb{T}} J_n(x)\,dx = 1.$$

The next lemma bounds the expected ℓ_2-norm of a random $\boldsymbol{x}$ under the distribution defined by $J_{d,n}$.

Lemma 7. For any $d, n \ge 1$,

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim J_{d,n}}\left[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{x}, 0^d)\right] \le \frac{\sqrt{d}}{n}.$$

We also need the following lemma about the Fourier coefficients of the Jackson's kernel.

Lemma 8. For any $d, n \ge 1$ and $\ell \in \mathbb{Z}^d$, $\left|\widehat{J_{d,n}}(\ell)\right| \le 1$. Moreover, $\widehat{J_{d,n}}(\ell) = 0$ if $\|\ell\|_\infty \ge 2n$.

We also use the fact that the Jackson's kernel is Lipschitz.

Lemma 9. For any $d, n \ge 1$, $J_{d,n}$ is $(3\pi n^2(3n/2)^{d-1}\sqrt{d})$-Lipschitz.

3 Multi-dimensional Wasserstein reconstruction

We start with our d-dimensional upper bound.

3.1 A d-dimensional upper bound

In this subsection, we prove an upper bound establishing parameters for which Wasserstein-distance reconstruction can be achieved in the general d-dimensional case. We show that reconstruction is possible within Wasserstein distance ε, given Fourier coefficients up to a bandlimit of individual degree $T = O(\sqrt{d}/\varepsilon)$ as long as the additive error in each Fourier coefficient is bounded by $\kappa = O(\varepsilon^d/d^{d/2})$. We will further show that the reconstruction process can be implemented by a polynomial-time algorithm in Section 4.

Theorem 10. Let f_1, f_2 be two signals over the the d-dimensional torus $\mathbb{T}^d$. For any $0 < \varepsilon < 1/2$, and any $d \geq 1$,

$$\text{let } T = \left\lceil 6\sqrt{d}/\varepsilon \right\rceil, \quad \text{and let } \kappa = \begin{cases} 0.001\varepsilon/\log(1/\varepsilon) & \text{when } d = 1 \\ \left(0.01\varepsilon/\sqrt{d}\right)^d & \text{when } d \geq 2 \end{cases}.$$

If f_1, f_2 satisfy $\left|\widehat{f_1}(\ell) - \widehat{f_2}(\ell)\right| \leq \kappa$ for all integer vectors $\ell = (\ell_1, \ldots, \ell_d) \in \mathbb{Z}^d$ with $\|\ell\|_\infty \leq T$, then $\mathrm{d_W}(f_1, f_2) \leq \varepsilon$.

We note that this theorem holds even for generalized signals f_1, f_2 whose absolute values need not integrate to 1. This will be useful for the algorithmic result given in Section 4.

Proof. Towards a contradiction, we assume that f_1 and f_2 are two signals such that $\mathrm{d_W}(f_1, f_2) > \varepsilon$.

The idea is to perform a smoothing on each of the signals using the Jackson's kernel. Define $\widetilde{f_1} := f_1 * J_{d,n}$ and $\widetilde{f_2} := f_2 * J_{d,n}$, where $n = \lceil T/2 \rceil$, $J_{d,n}$ is the Jackson's kernel in Definition 5, and $*$ stands for convolution. By Lemma 7, $\mathrm{d_W}(f_i, \widetilde{f_i}) \leq \sqrt{d}/n \leq \varepsilon/3$ for $i = 1, 2$, so by the triangle inequality we have that $\mathrm{d_W}(\widetilde{f_1}, \widetilde{f_2}) \geq \varepsilon/3$. This means that there exists some 1-Lipschitz function $h : \mathbb{T}^d \to \mathbb{R}$ with $\|h\|_\infty \leq \sqrt{d}$ satisfying

$$\int_{\mathbb{T}^d} \left(\widetilde{f_1}(x) - \widetilde{f_2}(x)\right) \cdot h(x)\, dx \geq \varepsilon/3.$$

Recalling that each $\widetilde{f_i}$ was obtained from f_i (which may not be square-integrable) by convolving with the Jackson's kernel, we see that each $\widetilde{f_i}$ is L^2-integrable, so we may apply Plancherel's theorem to restate the above inequality as

$$\sum_{\ell \in \mathbb{Z}^d} \left(\widehat{\widetilde{f_1}}(\ell) - \widehat{\widetilde{f_2}}(\ell)\right) \cdot \overline{\widehat{h}(\ell)} \geq \varepsilon/3.$$

By the definition of $\widetilde{f_1}$ and $\widetilde{f_2}$ and the convolution theorem, we have that $\widehat{\widetilde{f_1}}(\ell) - \widehat{\widetilde{f_2}}(\ell) = \left(\widehat{f_1}(\ell) - \widehat{f_2}(\ell)\right) \cdot \widehat{J_{d,n}}(\ell)$. However, by Lemma 8, we know that $\widehat{J_{d,n}}(\ell) = 0$ whenever $\|\ell\|_\infty \geq 2n$. Also, when $\|\ell\|_\infty = 0$, we have $|\widehat{h}(0^d)| \leq \sqrt{d}$ from $\|h\|_\infty \leq \sqrt{d}$. Combining this with the assumption that $\left|\widehat{f_1}(0^d) - \widehat{f_2}(0^d)\right| \leq \kappa$ and the choice of κ, it is easy to see that $\left(\widehat{\widetilde{f_1}}(0^d) - \widehat{\widetilde{f_2}}(0^d)\right) \cdot \overline{\widehat{h}(0^d)} \leq \varepsilon/6$.

Hence,

$$\sum_{0<\|\ell\|_\infty\le 2n-1} \left(\widehat{f}_1(\ell)-\widehat{f}_2(\ell)\right)\cdot \widehat{J_{d,n}}(\ell)\cdot \overline{\widehat{h}(\ell)} \ge \varepsilon/6. \tag{4}$$

On the other hand, since h is 1-Lipschitz, for $0^d \neq \ell \in \mathbb{Z}^d$ by Fact 4 we have that $\left|\widehat{h}(\ell)\right| \le 1/\|\ell\|_2$. Lemma 8 tells us that $\left|\widehat{J_{d,n}}(\ell)\right| \le 1$. Since by assumption we have that $\left|\widehat{f}_1(\ell)-\widehat{f}_2(\ell)\right| \le \kappa$ for all ℓ with $\|\ell\|_\infty \le 2n-1 \le T$, we have

$$\begin{aligned}
&\sum_{0<\|\ell\|_\infty\le 2n-1} \left(\widehat{f}_1(\ell)-\widehat{f}_2(\ell)\right)\cdot \widehat{J_{d,n}}(\ell)\cdot \overline{\widehat{h}(\ell)} \\
\le &\sum_{0<\|\ell\|_\infty\le 2n-1} \left|\widehat{f}_1(\ell)-\widehat{f}_2(\ell)\right|\cdot \left|\widehat{J_{d,n}}(\ell)\right|\cdot \left|\overline{\widehat{h}(\ell)}\right| \\
\le &\kappa\cdot \sum_{0<\|\ell\|_\infty\le 2n-1} \frac{1}{\|\ell\|_2} \\
\le &\kappa\cdot \sum_{0<\|\ell\|_\infty\le 2n-1} \frac{1}{\|\ell\|_\infty} \\
\le &\kappa\cdot \sum_{i=1}^{T} \frac{Q_i}{i},
\end{aligned}$$

where $Q_i := \left|\{\ell \mid \|\ell\|_\infty = i\}\right|$ is the number of ℓ with ℓ_∞-distance exactly i. The last inequality holds since $T \ge 2n-1$.

Using the trivial bound

$$Q_i = (2i+1)^d - (2i-1)^d \le 2d\cdot (2i+1)^{d-1},$$

we have

$$\sum_{i=1}^{T} \frac{Q_i}{i} \le 6d\cdot \sum_{i=1}^{2T+1} i^{d-2}.$$

When $d=1$, this is upper bounded by $6\cdot(\ln(2T+1)+1) \le 30\log(1/\varepsilon)$. When $d\ge 2$, this is upper bounded by $6d\cdot\frac{(2T+2)^{d-1}}{d-1} \le 12\cdot\left(\frac{24\sqrt{d}}{\varepsilon}\right)^{d-1}$.

In either case, by our choice of κ, it is easy to verify that

$$\sum_{0<\|\ell\|_\infty\le 2n-1} \left(\widehat{f}_1(\ell)-\widehat{f}_2(\ell)\right)\cdot \widehat{J_{d,n}}(\ell)\cdot \overline{\widehat{h}(\ell)} \le \kappa\cdot\sum_{i=1}^{T}\frac{Q_i}{i} < \varepsilon/6,$$

which contradicts with Equation (4). □

3.2 Multi-dimensional Wasserstein reconstruction: $T = \frac{\sqrt{d}}{2\varepsilon}-1$, $\kappa = 0$ lower bound

We show in this section that even if we have the exact Fourier coefficients up to individual degree $\sqrt{d}/(2\varepsilon)-1$ for a signal over the d-dimensional torus $\mathbb{T}^d$, we cannot reconstruct the signal within an ε Wasserstein distance. In fact, we will prove that this holds even for non-negative signals, i.e., for probability distributions.

Theorem 11. For any $0 < \varepsilon \le 1/2$, there are two distributions D_1, D_2 over the d-dimensional torus $\mathbb{T}^d$, such that $\widehat{D_1}(\ell) = \widehat{D_2}(\ell)$ for any $\ell \in \mathbb{Z}^d$ with $\|\ell\|_\infty \le \sqrt{d}/(2\varepsilon) - 1$, while $\mathrm{d_W}(D_1, D_2) \ge \varepsilon$.

Proof. Let $T' := \lfloor \sqrt{d}/(2\varepsilon) \rfloor$. Let D_1 be the uniform distribution over points of the form $(j_1/T', \dots, j_d/T')$ where $j_1, \dots, j_d \in \{0, 1, \dots, T'-1\}$, and let D_2 be the uniform distribution over points of the form $(j_1/T' + 1/(2T'), \dots, j_d/T' + 1/(2T'))$ where $j_1, \dots, j_d \in \{0, 1, \dots, T'-1\}$.

We first argue that $\mathrm{d_W}(D_1, D_2) \ge \varepsilon$. This is true because for any $x \in D_1, y \in D_2$,

$$\mathrm{d_{tor}}(x, y) = \sqrt{\mathrm{d_{tor}}(x_1, y_1)^2 + \cdots + \mathrm{d_{tor}}(x_d, y_d)^2} \ge \sqrt{d(1/(2T'))^2} \ge \varepsilon.$$

Below we fix any $\ell \in \mathbb{Z}^d$ such that $\|\ell\|_\infty \le \sqrt{d}/(2\varepsilon) - 1$ and prove that $\widehat{D_1}(\ell) = \widehat{D_2}(\ell)$. Note that $\|\ell\|_\infty$ is an integer, so $\|\ell\|_\infty \le T' - 1$.

When $\ell = 0^d$, $\widehat{D_1}(\ell) = 1 = \widehat{D_2}(\ell)$ since D_1 and D_2 are distributions.

When $\ell \ne 0^d$, without loss of generality, suppose $\ell_1 \ne 0$. Since $|\ell_1| \le \|\ell\|_\infty \le T' - 1$, we have

$$\sum_{j_1=0}^{T'-1} e^{2\pi i \ell_1 j_1/T'} = 0.$$

Thus,

$$\widehat{D_1}(\ell) = \sum_{j \in \{0,1,\dots,T'-1\}^d} \frac{1}{T'^d} e^{2\pi i(\ell_1(j_1/T') + \cdots + \ell_d(j_d/T'))} = \frac{1}{T'^d} \left(\sum_{j_1=0}^{T'-1} e^{2\pi i \ell_1 j_1/T'} \right) \cdots \left(\sum_{j_d=0}^{T'-1} e^{2\pi i \ell_d j_d/T'} \right) = 0,$$

and similarly,

$$\begin{aligned}\widehat{D_2}(\ell) &= \sum_{j \in \{0,1,\dots,T'-1\}^d} \frac{1}{T'^d} e^{2\pi i(\ell_1(j_1/T' + 1/(2T')) + \cdots + \ell_d(j_d/T' + 1/(2T')))} \\ &= \frac{e^{\pi i(\ell_1 + \cdots + \ell_d)}}{T'^d} \left(\sum_{j_1=0}^{T'-1} e^{2\pi i \ell_1 j_1/T'} \right) \cdots \left(\sum_{j_d=0}^{T'-1} e^{2\pi i \ell_d j_d/T'} \right) = 0.\end{aligned}$$

Therefore, for any $\ell \in \mathbb{Z}^d$ such that $\ell_\infty \le \sqrt{d}/(2\varepsilon) - 1$, $\widehat{D_1}(\ell) = \widehat{D_2}(\ell)$. $\square$

Remark 12. It is easy to see that the Lévy-Prokhorov distance between the distributions D_1 and D_2 constructed in the above proof is also at least ε. Thus, we cannot reconstruct a signal over $\mathbb{T}^d$ within an ε Lévy-Prokorov distance even if we have the exact Fourier coefficients up to individual degree $\sqrt{d}/(2\varepsilon) - 1$.

3.3 Multi-dimensional Wasserstein reconstruction: $T = \infty$, $\kappa = \varepsilon^{0.249d}$ lower bound

We show in this section that for a signal over the d-dimensional torus $\mathbb{T}^d$, even if we are given all (infinitely many) Fourier coefficients each to within $\pm\varepsilon^{0.249d}$ additive error, we cannot reconstruct the signal within an ε Wasserstein distance. In fact, we also prove that this holds for non-negative signals, i.e., probability distributions.

Theorem 13. There exists a universal constant $\varepsilon_0 > 0$, such that for any $0 < \varepsilon \le \varepsilon_0$, there are two distributions D_1, D_2 over the d-dimensional torus $\mathbb{T}^d$, such that $|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)| \le \varepsilon^{0.249d}$ for any $\ell \in \mathbb{Z}^d$, while $\mathrm{d_W}(D_1, D_2) \ge \varepsilon$.

Proof. Let $\kappa = \varepsilon^{0.249d}$, $M = \frac{1}{2}(8\varepsilon)^{-0.5d}$, and $n = 4\varepsilon^{-1}\sqrt{d}$. Let $x_1, \ldots, x_M, y_1, \ldots, y_M \in \mathbb{T}^d$ be $2M$ points that satisfy:

(i) For all $1 \le j, k \le M$, $\mathrm{d_{tor}}(x_j, y_k) > 4\varepsilon$.

(ii) For all $\ell \in \mathbb{Z}^d$ such that $0 < \|\ell\|_\infty \le 2n$, we have $\left|\frac{1}{M}\sum_{j=1}^M e^{2\pi i(\ell\cdot x_j)}\right| < \frac{\kappa}{2}$ and $\left|\frac{1}{M}\sum_{k=1}^M e^{2\pi i(\ell\cdot y_k)}\right| < \frac{\kappa}{2}$.

We show the existence of such $x_1, \ldots, x_M, y_1, \ldots, y_M$ in Claim 14.

Now let D_1' be the uniform distribution over $x_1, \ldots, x_M$, and let $D_1 := D_1' * J_{d,n}$ (so D_1' is the uniform mixture of M Jackson's kernel distributions centered at the points $x_1, \ldots, x_M$). Similarly, let D_2' be the uniform distribution over $y_1, \ldots, y_M$, and let $D_2 := D_2' * J_{d,n}$.

We first argue that $\mathrm{d_W}(D_1, D_2) \ge \varepsilon$. By Lemma 7, $\mathbf{E}_{\boldsymbol{x}\sim J_{d,n}}[\mathrm{d_{tor}}(\boldsymbol{x}, 0^d)] \le \sqrt{d}/n$, so by Markov's inequality,

$$\Pr_{\boldsymbol{x}\sim J_{d,n}}\left[\mathrm{d_{tor}}(\boldsymbol{x}, 0^d) \le \varepsilon\right] \ge 1 - \frac{\sqrt{d}}{\varepsilon n} \ge 0.75.$$

Recall that we defined D_1' to be the uniform distribution over $x_1, \ldots, x_M$ and $D_1 = D_1' * J_{d,n}$, so

$$\Pr_{\boldsymbol{x}\sim D_1}\left[\exists 1 \le j \le M \text{ s.t. } \mathrm{d_{tor}}(\boldsymbol{x}, x_j) \le \varepsilon\right] \ge 0.75.$$

Similarly,

$$\Pr_{\boldsymbol{y}\sim D_2}\left[\exists 1 \le k \le M \text{ s.t. } \mathrm{d_{tor}}(\boldsymbol{y}, y_k) \le \varepsilon\right] \ge 0.75.$$

Let C be an arbitrary coupling of D_1 and D_2. By a union bound,

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\sim C}\left[\exists 1 \le j, k \le M \text{ s.t. } \mathrm{d_{tor}}(\boldsymbol{x}, x_j) \le \varepsilon \text{ and } \mathrm{d_{tor}}(\boldsymbol{y}, y_k) \le \varepsilon\right] \ge 0.5.$$

Recall that $\mathrm{d_{tor}}(x_j, y_k) > 4\varepsilon$ for all $1 \le j, k \le M$, so $\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})\sim C}[\mathrm{d_{tor}}(\boldsymbol{x}, \boldsymbol{y}) > 2\varepsilon] \ge 0.5$. This implies

$$\mathrm{d_W}(D_1, D_2) \ge \varepsilon$$

since C is an arbitrary coupling of D_1, D_2.

Now we show that for any $\ell \in \mathbb{Z}^d$, $\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| \le \kappa$. When $\|\ell\|_\infty \ge 2n$, by Lemma 8, $\widehat{J_{d,n}}(\ell) = 0$, so $\widehat{D_1}(\ell) = 0 = \widehat{D_2}(\ell)$. When $\ell = 0^d$, $\widehat{D_1}(\ell) = 1 = \widehat{D_2}(\ell)$ since D_1 and D_2 are distributions. When $0 < \|\ell\|_\infty < 2n$, by Lemma 8, $\left|\widehat{J_{d,n}}(\ell)\right| \le 1$, so

$$\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| = \left|\widehat{J_{d,n}}(\ell)\right| \cdot \left|\widehat{D_1'}(\ell) - \widehat{D_2'}(\ell)\right| \le \left|\frac{1}{M}\sum_{j=1}^M e^{2\pi i(\ell\cdot x_j)} - \frac{1}{M}\sum_{k=1}^M e^{2\pi i(\ell\cdot y_k)}\right| < \kappa,$$

where the final uses (ii) and the triangle inequality. ☐

Claim 14. There exists $x_1, \ldots, x_M, y_1, \ldots, y_M \in \mathbb{T}^d$ that satisfy:

(i) For all $1 \le j, k \le M$, $\mathrm{d_{tor}}(x_j, y_k) > 4\varepsilon$.

(ii) For all $\ell \in \mathbb{Z}^d$ such that $0 < \|\ell\|_\infty < 2n$, $\frac{1}{M}\left|\sum_{j=1}^M e^{2\pi i(\ell\cdot x_j)}\right| < \frac{\kappa}{2}$ and $\frac{1}{M}\left|\sum_{k=1}^M e^{2\pi i(\ell\cdot y_k)}\right| < \frac{\kappa}{2}$.

Proof. The proof is by the probabilistic method: we uniformly sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_M$ from $\mathbb{T}^d$ and show that the outcome has the desired properties with positive probability.

For any $1 \le j, k \le M$, we have

$$\mathbf{Pr}[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{x}_j, \boldsymbol{y}_k) \le 4\varepsilon] \le (8\varepsilon)^d,$$

so by union bound over all $1 \le j, k \le M$,

$$\mathbf{Pr}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_M \text{ satisfy (i)}] \ge 1 - M^2(8\varepsilon)^d \ge 0.75.$$

For any $\ell \in \mathbb{Z}^d \setminus \{0^d\}$, without loss of generality assume $\ell_1 \neq 0$, then for any $1 \le j \le M$,

$$\mathbf{E}[e^{2\pi i(\ell \cdot \boldsymbol{x}_j)}] = \int_{\mathbb{T}^d} e^{2\pi i(\ell \cdot x)} dx = \int_{\mathbb{T}^d} e^{2\pi i \ell_1 x_1} e^{2\pi i(\ell_2 x_2 + \cdots + \ell_d x_d)} dx = 0$$

since $\int_{\mathbb{T}} e^{2\pi i \ell_1 x_1} dx_1 = 0$. Moreover, $|e^{2\pi i(\ell \cdot \boldsymbol{x}_j)}| = 1$, so by Hoeffding's inequality,

$$\mathbf{Pr}\left[\left|\sum_{j=1}^{M} \mathrm{Re}(e^{2\pi i(\ell \cdot \boldsymbol{x}_j)})\right| \ge \frac{\kappa M}{2\sqrt{2}}\right] \le 2e^{-\kappa^2 M/16}$$
$$\mathbf{Pr}\left[\left|\sum_{j=1}^{M} \mathrm{Im}(e^{2\pi i(\ell \cdot \boldsymbol{x}_j)})\right| \ge \frac{\kappa M}{2\sqrt{2}}\right] \le 2e^{-\kappa^2 M/16}$$

and thus

$$\mathbf{Pr}\left[\left|\sum_{j=1}^{M} e^{2\pi i(\ell \cdot \boldsymbol{x}_j)}\right| \ge \frac{\kappa M}{2}\right] \le 4e^{-\kappa^2 M/16}.$$

Similarly,

$$\mathbf{Pr}\left[\left|\sum_{j=1}^{M} e^{2\pi i(\ell \cdot \boldsymbol{y}_k)}\right| \ge \frac{\kappa M}{2}\right] \le 4e^{-\kappa^2 M/16}.$$

Therefore, by union bound over all $\ell \in \mathbb{Z}^d$ such that $0 < \|\ell\|_\infty < 2n$,

$$\mathbf{Pr}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_M \text{ satisfy (ii)}] \ge 1 - 8(4n-1)^d e^{-\kappa^2 M/16} \ge 0.75$$

where the last inequality holds because $0 < \varepsilon \le \varepsilon_0$ where ε_0 is sufficiently small. It follows that

$$\mathbf{Pr}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_M \text{ satisfy (i) and (ii)}] \ge 0.5,$$

and the claim is proved. $\square$

Remark 15. By an argument similar to that for Theorem 13 (consider the union of ε-radius balls centered at x_j), the Lévy-Prokhorov distance between D_1 and D_2 is at least ε. Thus, we cannot reconstruct a signal over $\mathbb{T}^d$ within an ε Lévy-Prokhorov distance even if we have all Fourier coefficients each with $\varepsilon^{0.249d}$ error.

3.4 A sharper one-dimensional lower bound

For the one-dimensional case, we can give a stronger lower bound than the one obtained by taking $d = 1$ in Section 3.3. This stronger lower bound implies that the upper bound obtained by taking $d = 1$ in Section 3.1 is essentially the best possible. In more detail, we show that ε-accurate Wasserstein reconstruction is impossible even if we are given $\pm 4\varepsilon$-accurate estimates of every Fourier coefficient, and even if the signal is promised to be non-negative (i.e. it is some distribution D):

Observation 16. There are two distributions D_1, D_2 over the 1-dimensional torus $\mathbb{T}$ with the following properties:

- $\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| \le 4\varepsilon$ for all $\ell \in \mathbb{Z}$, but
- $\mathrm{d_W}(D_1, D_2) \ge \varepsilon$.

Proof. The construction is extremely simple: we take $D_1 := \delta_0$, i.e. the distribution D_1 is a point mass at 0, and we take $D_2 = (1 - 2\varepsilon) \cdot \delta_0 + 2\varepsilon \cdot \delta_{1/2}$, i.e. a draw from D_1 outputs 0 with probability $1 - 2\varepsilon$ and outputs $1/2$ with the remaining 2ε probability.

It is clear that $\mathrm{d_W}(D_1, D_2) = \varepsilon$, and we have

$$\widehat{D_1}(\ell) = \int_{z \in [0,1)} \delta_0(z) e^{2\pi i \ell z} dz = e^{2\pi i \ell \cdot 0} = 1 \qquad \text{for all } \ell \in \mathbb{Z},$$

$$\begin{aligned} \widehat{D_2}(\ell) &= \int_{z \in [0,1)} \Big((1 - 2\varepsilon) \cdot \delta_0(z) + 2\varepsilon \cdot \delta_{1/2}(z)\Big) e^{2\pi i \ell z} dz \\ &= (1 - 2\varepsilon) + (2\varepsilon) e^{\pi i \ell} \\ &= \begin{cases} 1 & \text{if } \ell \text{ is even} \\ 1 - 4\varepsilon & \text{if } \ell \text{ is odd} \end{cases}, \end{aligned}$$

so $\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| \le 4\varepsilon$ for all $\ell \in \mathbb{Z}$. $\square$

4 An algorithm for Wasserstein reconstruction

In this section we show how the upper bounds for Wasserstein reconstruction that were established in Section 3 can be used to design an algorithm which, given the (noisy) low-frequency Fourier coefficients of some signal f, outputs a signal that is close to f in Wasserstein distance. The idea is to discretize the space and use linear programming to find a hypothesis signal that has Fourier coefficients which are close to those of f; using the upper bounds established earlier, such a hypothesis signal must be Wasserstein-close to f.

We first formally state the theorem:

Theorem 17. There is an algorithm that satisfies the following: Fix any $d \ge 1$ and $0 < \varepsilon < 1/8$, and let T and κ be as in Theorem 10, i.e.,

$$\text{let } T = \left\lceil 6\sqrt{d}/\varepsilon \right\rceil, \quad \text{and let } \kappa = \begin{cases} 0.001\varepsilon/\log(1/\varepsilon) & \text{when } d = 1 \\ \left(0.01\varepsilon/\sqrt{d}\right)^d & \text{when } d \ge 2 \end{cases}.$$

The algorithm takes as input $(2T+1)^d$ complex numbers $\left\{u_\ell \;\middle|\; \ell \in \mathbb{Z}^d, \|\ell\|_\infty \le T\right\}$ and outputs a signal g such that for any signal f satisfying $\left|\widehat{f}(\ell) - u_\ell\right| \le \kappa/8$ for all $\|\ell\|_\infty \le T$, we have $\mathrm{d_W}(f,g) \le 4\varepsilon$. The algorithm runs in in time $(d/\varepsilon)^{O(d^2)}$ (so for constant d, its running time is polynomial in the length of its input).

Regarding the representation of the hypothesis signal, the hypothesis is a Dirac comb signal supported on $(d/\varepsilon)^{O(d^2)}$ discrete points; the algorithm outputs the location of the points of the Dirac comb and their associated (positive or negative) masses.

For convenience, throughout this section we write $[N]$ to denote $\{0,1,\dots,N-1\}$.

Proof of Theorem 17. Let $K := \lceil 100d{\cdot}(2n)^{d+1}/\kappa\rceil$. We discretize $\mathbb{T}^d$ into K^d points $(j_1/K, j_2/K, \dots, j_d/K)$ where j ranges over $[K]^d$.

The following notation will be useful: we use $a \in \mathbb{R}^{K^d}$ to represent a sequence of real numbers (which can be positive or negative), $a = \left\{a_j \;\middle|\; j = (j_1,\dots,j_d) \text{ where } 0 \le j_1,\dots,j_d < K\right\}$. We define the (possibly generalized) signal

$$\mathrm{Comb}_a(x) := \sum_{j\in[K]^d} a_j \cdot \delta_{(j_1/K, j_2/K, \dots, j_d/K)}(x).$$

When $\sum_j |a_j| = 1$, this is a bona fide (normalized) signal with mass a_j at the point $(j_1/K,\dots,j_d/K)$. Let $\mathcal{U}_{[0,1/K)^d}$ denote the uniform distribution on $[0,1/K)^d$, and let

$$\widetilde{\mathrm{Comb}}_a(x) := \mathrm{Comb}_a(x) \;*\; \mathcal{U}_{[0,1/K)^d}$$

where $*$ stands for convolution. In words, when $\sum_j |a_j| = 1$, then $\widetilde{\mathrm{Comb}}_a$ is the signal with $\widetilde{\mathrm{Comb}}_a(x) = K^d \cdot a_j$ in the rectangle $\left[\frac{j_1}{K}, \frac{j_1+1}{K}\right) \times \left[\frac{j_2}{K}, \frac{j_2+1}{K}\right) \times \cdots \times \left[\frac{j_d}{K}, \frac{j_d+1}{K}\right)$.

Given the input $\left\{u_\ell \;\middle|\; \ell \in \mathbb{Z}^d \text{ s.t. } \|\ell\|_\infty \le T\right\}$, let f be the target signal with $\left|\widehat{f}(\ell) - u_\ell\right| \le \kappa$ for all $\|\ell\|_\infty \le T$. Our algorithm has three steps which we describe and analyze below (the algorithm is also summarized in Algorithm 1).

Step 1: Jackson smoothing. The first step is to apply a smoothing on u and f using the Jackson's kernel (Definition 5). Let $n := \left\lceil \sqrt{d}/\varepsilon \right\rceil$, and compute $\widetilde{u}_\ell := u_\ell \cdot \widehat{J_{d,n}}(\ell)$ for each $\ell \in \mathbb{Z}^d$ with $\|\ell\|_\infty \le T$.

Consider the function $\widetilde{f} := f * J_{d,n}$. This function has three properties that will be useful later:

1. By Lemma 7, $\mathrm{d_W}(f, \widetilde{f}) \le \sqrt{d}/n \le \varepsilon$.
2. By Lemma 8, $\left|\widehat{J_{d,n}}(\ell)\right| \le 1$. Combining with the assumption that $\left|\widehat{f}(\ell) - u_\ell\right| \le \kappa/8$ for all $\|\ell\|_\infty \le T$, we also have $\left|\widehat{\widetilde{f}}(\ell) - \widetilde{u}_\ell\right| \le \kappa/8$ for all $\|\ell\|_\infty \le T$.
3. By Lemma 9 and the fact that $\int_{\mathbb{T}^d} |f(x)| dx = 1$, $\widetilde{f}$ is $(3\pi n^2 (3n/2)^{d-1}\sqrt{d})$-Lipschitz. This uses the standard fact that for any signal f and Lipschitz function g, the function $f * g$ inherits the Lipschitz constant of g [Hen11].

Step 2: Linear programming. The second step is to use linear programming to find some $a \in \mathbb{R}^{K^d}$ such that $\widetilde{\mathrm{Comb}}_a$ is a signal that has Fourier coefficients close to $\widetilde{u}$. By Theorem 10, this means that the Wasserstein distance between $\widetilde{\mathrm{Comb}}_a$ and $\widetilde{f}$ is small.

There is an obvious obstacle in the way of a naive formalization of the constraints on such an a using linear programming: this is that the condition that $\widetilde{\mathrm{Comb}}_a$ is a normalized signal, i.e., $\sum_{j \in [K]^d} |a_j| = 1$, is not linear. Instead, we will find some vector a whose coordinates have total magnitude *at most* 1. We will deal with the possibility $\sum_{j \in [K]^d} |a_j| < 1$ later, in Step 3.

For any $j \in [K]^d$, let

$$c_{j,\ell} := \int_{\left[\frac{j_1}{K}, \frac{j_1+1}{K}\right) \times \left[\frac{j_2}{K}, \frac{j_2+1}{K}\right) \times \cdots \times \left[\frac{j_d}{K}, \frac{j_d+1}{K}\right)} e^{2\pi i (\ell \cdot x)} \, dx.$$

For any complex number $z \in \mathbb{C}$, denote the real part and the imaginary part of z by $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, respectively. Consider the following linear program whose indeterminates are the coordinates a_j of a:

$$\begin{aligned}
&\text{Find } \{a_j\}_{j \in [K]^d} && (5)\\
&\text{Subject to } \sum_{j \in [K]^d} |a_j| \le 1, &&\\
&\left| K^d \sum_{j \in [K]^d} \mathrm{Re}(c_{j,\ell})\, a_j - \mathrm{Re}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4}, && \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T,\\
&\left| K^d \sum_{j \in [K]^d} \mathrm{Im}(c_{j,\ell})\, a_j - \mathrm{Im}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4}, && \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T.
\end{aligned}$$

In words, the second and third set of inequalities essentially say that the real parts and imaginary parts of $\widetilde{\widetilde{\mathrm{Comb}}_a}(\ell)$ are close to those of $\widetilde{u}_\ell$, respectively.

We first show that the linear program (5) is satisfiable. Consider the following candidate solution $b \in \mathbb{R}^{K^d}$:

$$\begin{aligned}
b_j := K^{-d} \cdot \widetilde{f}(x) \text{ where } x \text{ minimizes } \left|\widetilde{f}(x)\right| \text{ within the rectangle}&\\
\left[\frac{j_1}{K}, \frac{j_1+1}{K}\right) \times \left[\frac{j_2}{K}, \frac{j_2+1}{K}\right) \times \cdots \times \left[\frac{j_d}{K}, \frac{j_d+1}{K}\right).& \qquad (6)
\end{aligned}$$

By definition,

$$\sum_{j \in [K]^d} |b_j| = \int_{\mathbb{T}^d} \widetilde{\mathrm{Comb}}_b(x)\, dx \le \int_{\mathbb{T}^d} \left|\widetilde{f}(x)\right| \, dx = 1,$$

satisfying the first constraint.

We know by Property 3 that $\widetilde{f}$ is $(3\pi n^2 (3n/2)^{d-1}\sqrt{d})$-Lipschitz, so we also have

$$\left\|\widetilde{\mathrm{Comb}}_b - \widetilde{f}\right\|_\infty \le \frac{\sqrt{d}}{K} \cdot (3\pi n^2 (3n/2)^{d-1}\sqrt{d}) \le \kappa/8\,,$$

where we use that, on every rectangle (cf. Equation (6)), $\widetilde{\mathrm{Comb}}_b$ agrees with $\widetilde{f}$ at some point; the Lipschitzness of $\widetilde{f}$ and the bound K on the rectangle's diameter then yields the desired uniform bound. Since for any (possibly generalized) signal g and any $\ell \in \mathbb{Z}^d$ we have $|\widehat{g}(\ell)| \le \|g\|_\infty$, for any $\ell \in \mathbb{Z}^d$ with $\|\ell\|_\infty \le T$ we have

$$\left|\widehat{\widetilde{\mathrm{Comb}}_b}(\ell) - \widehat{\widetilde{f}}(\ell)\right| \le \kappa/8.$$

We also know from Property 2 that

$$\left|\widehat{\widetilde{f}}(\ell) - \widetilde{u}_\ell\right| \le \kappa/8,$$

so by the triangle inequality we have

$$\left|\widehat{\widetilde{\mathrm{Comb}}_b}(\ell) - \widetilde{u}_\ell\right| \le \kappa/4.$$

Therefore, b also satisfies the second and third inequalities of (5), and hence the linear program in Equation (5) is indeed feasible.

Therefore, by solving Equation (5) with, say, the interior point method with error $\delta := \kappa/8$, in $\mathrm{poly}(K^d, \log(1/\delta))$ time we can find some $a^* \in \mathbb{R}^{K^d}$ satisfying the following linear program:

$$\begin{aligned}
&\text{Find } \{a^*_j\}_{j \in [K]^d} \\
&\text{Subject to } \sum_{j \in [K]^d} |a^*_j| \le 1 + \delta, \\
&\qquad \left| K^d \sum_{j \in [K]^d} \mathrm{Re}(c_{j,\ell})\, a^*_j - \mathrm{Re}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4} + \delta, \qquad \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T, \\
&\qquad \left| K^d \sum_{j \in [K]^d} \mathrm{Im}(c_{j,\ell})\, a^*_j - \mathrm{Im}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4} + \delta, \qquad \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T.
\end{aligned} \tag{7}$$

By the second and third sets of inequalities, for any $\ell \in \mathbb{Z}^d$ with $\|\ell\|_\infty \le T$, we have

$$\left|\widehat{\widetilde{\mathrm{Comb}}_{a^*}}(\ell) - \widetilde{u}_\ell\right| \le \kappa/2 + 2\delta \le 3\kappa/4.$$

Using Property 2 of $\widetilde{f}$ and the triangle inequality, we have

$$\left|\widehat{\widetilde{\mathrm{Comb}}_{a^*}}(\ell) - \widehat{\widetilde{f}}(\ell)\right| \le 3\kappa/4 + \kappa/8 < \kappa.$$

By Theorem 10, this means that $\mathrm{d_W}(\widetilde{\mathrm{Comb}}_{a^*}, \widetilde{f}) \le \varepsilon$.

Step 3: Re-normalizing a^*. Before re-normalizing, we first move to a Dirac comb function. Consider the function Comb_{a^*}; by the definition of $\widetilde{\mathrm{Comb}}_{a^*}$, we have

$$\mathrm{d_W}(\mathrm{Comb}_{a^*}, \widetilde{\mathrm{Comb}}_{a^*}) \le \frac{\sqrt{d}}{K} \le \varepsilon.$$

The next step is to re-normalize a^* so that the absolute values sum up to 1, so that the Dirac comb function becomes a signal. There are two cases here depending on $\gamma := \sum_{j \in [K]^d} |a^*_j|$.

Algorithm 1: Wasserstein reconstruction given noisy low-frequency Fourier coefficients

Input : $\{u_\ell : \ell \in \mathbb{Z}^d \text{ such that } \|\ell\|_\infty \le T\}$

Output : A signal g s.t. $\mathrm{d_W}(f,g) \le 4\varepsilon$ for any f with $\left|\widehat{f}(\ell) - u_\ell\right| \le \kappa/8$ for all $\|\ell\|_\infty \le T$

Let $K \leftarrow \lceil 100d \cdot (2n)^{d+1}/\kappa \rceil$, $n \leftarrow \lceil \sqrt{d}/\varepsilon \rceil$, $\delta \leftarrow \kappa/8$;
Let $\widetilde{u}_\ell \leftarrow u_\ell \cdot \widehat{J_{d,n}}(\ell)$;
Solve the LP in Equation (5) to obtain some $a^* \in \mathbb{R}^{K^d}$ satisfying Equation (7);
if $\gamma := \sum_{j \in [K]^d} |a_j^*| < 1$ **then**
  Output $g(x) := \mathrm{Comb}_{a^*}(x) + \frac{1-\gamma}{2}\delta_{(1/(3K),0,\ldots,0)}(x) - \frac{1-\gamma}{2}\delta_{(2/(3K),0,\ldots,0)}(x)$.
else
  Let $\breve{a}_i^* \leftarrow a_i^*/\gamma$;
  Output $g(x) := \mathrm{Comb}_{\breve{a}^*}(x)$
end

The first case is when $\gamma < 1$. Let x_0 and x_1 be two arbitrary points on $\mathbb{T}^d$ with Euclidean distance at most ε that are not on the grid $\{(j_1/K, j_2/K, \ldots, j_d/K) \mid j \in [K]^d\}$ (for concreteness, take $x_0 := (1/(3K), 0, \ldots, 0)$ and $x_1 := (2/(3K), 0, \ldots, 0)$). Then $\|x_0 - x_1\|_2 = 1/(3K) \le \varepsilon$. Then define

$$g(x) := \mathrm{Comb}_{a^*}(x) + \frac{1-\gamma}{2}\delta_{x_0}(x) - \frac{1-\gamma}{2}\delta_{x_1}(x).$$

By definition, $\int_{\mathbb{T}^d} g(x)dx = 1$, and $\mathrm{d_W}(g, \mathrm{Comb}_{a^*}) \le (1-\gamma) \cdot \|x_0 - x_1\|_2 \le \varepsilon$.

The second case is when $\gamma \ge 1$. By the first inequality of Equation (7), we have $\gamma \le 1 + \delta = 1 + \kappa/8$. Define

$$\breve{a}_j^* := \frac{1}{\gamma} \cdot a_j^*$$

and consider the signal.

$$g(x) := \mathrm{Comb}_{\breve{a}^*}(x)$$

By the definition of $\breve{a}^*$, it is easy to see that $\int_{\mathbb{T}^d} g(x)dx = 1$. Moreover,

$$\mathrm{d_W}(g, \mathrm{Comb}_{a^*}) \le \sqrt{d} \cdot \sum_{j \in [K]^d} |a_j^* - \breve{a}_j^*| = \sqrt{d} \cdot (\gamma - 1) \le \varepsilon.$$

Putting everything together. In both of the above two cases, we get a signal g with $\mathrm{d_W}(g, \mathrm{Comb}_{a^*}) \le \varepsilon$. Putting everything together, we have

$$\mathrm{d_W}(f,g) \le \mathrm{d_W}(f, \widetilde{f}) + \mathrm{d_W}(\widetilde{f}, \widetilde{\mathrm{Comb}}_{a^*}) + \mathrm{d_W}(\widetilde{\mathrm{Comb}}_{a^*}, \mathrm{Comb}_{a^*}) + \mathrm{d_W}(\mathrm{Comb}_{a^*}, g) \le 4\varepsilon.$$

The final algorithm is shown in Algorithm 1. □

Remark 18. We note that when f is a distribution instead of a signal, our algorithm can be modified to also output a distribution g. In this case, the linear program in Equation (5) should be replaced by

Find $\{a_j\}_{j \in [K]^d}$

$$\begin{aligned}
\text{Subject to } & \sum_{j\in[K]^d} a_j = 1, & \\
& a_j \ge 0, & \forall\, j \in [K]^d, \\
& \left| K^d \sum_{j\in[K]^d} \mathrm{Re}(c_{j,\ell})\, a_j - \mathrm{Re}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4}, & \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T, \\
& \left| K^d \sum_{j\in[K]^d} \mathrm{Im}(c_{j,\ell})\, a_j - \mathrm{Im}(\widetilde{u}_\ell) \right| \le \frac{\kappa}{4}, & \forall\, \ell \in \mathbb{Z}^d : \|\ell\|_\infty \le T.
\end{aligned}$$

We can no longer proceed as we did formerly in the $\gamma < 1$ case. Fortunately, we can write down the constraint that $\sum_{j\in[K]^d} a_j = 1$ in a linear program, and a standard normalization suffices. The LP solution a^* might have some negative coordinates, but we can simply clip all the negative a_j^*'s before doing the normalization.

5 A d-dimensional upper bound for heavy hitter reconstruction

We now turn to heavy hitter reconstruction of probability distributions. Our goal in this section is to prove the following:

Theorem 19. Given $0 < \varepsilon \le 1/2, 0 < \varepsilon_{\mathrm{dist}} < 1$, let D_1, D_2 be two distributions over $[0,1)^d$ such that

$$\left| \widehat{D_1}(\ell) - \widehat{D_2}(\ell) \right| \le \kappa := \frac{\varepsilon}{9}\left(\frac{1}{d}\right)^{O(T)} \text{ for all } \ell \in \mathbb{Z}^d \text{ with } \|\ell\|_1 \le T\,, \tag{8}$$

for some $T = O\left(\frac{\sqrt{d}\log(1/\varepsilon)}{\varepsilon_{\mathrm{dist}}}\right)$. Then $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D_1, D_2) \le \varepsilon$.

Note that in Theorem 19 the degree bound is in the sense of *total degree* rather than *individual degree* as was the case for our Wasserstein results. As mentioned earlier, this is important because it means that the total number of Fourier coefficients required for reconstruction scales (as a function of d) as $2^{\tilde{O}(\sqrt{d})}$ rather than as $\exp(d)$.

5.1 A low-degree trigonometric polynomial "bump function"

An essential ingredient of our approach is a (relatively) low-degree trigonometric polynomial which serves as a "bump function" over the torus $[0,1)^d$; the main result of this subsection, Lemma 22, establishes the existence of such a trigonometric polynomial.

5.1.1 A useful univariate polynomial

To build the desired multivariate trigonometric polynomial bump function, we begin with a construction of a univariate polynomial with the following behavior over the interval $[0,1]$: it takes values in $[0,1]$ on inputs in $[0,1]$, takes values close to 1 on inputs that are within a certain prescribed distance of 0, and takes values close to 0 on inputs that are "even a little bit beyond" that prescribed distance. The following result, giving such a univariate polynomial, is the main result of this subsubsection:

Lemma 20. Let $0 < \varepsilon, \varepsilon_{\text{dist}} < 1$, let $0 < b < A$ where $A, b = \Theta(\varepsilon_{\text{dist}}^2)$, and let d be an asymptotic parameter. There is a real polynomial $q(x)$ of degree

$$\Psi(\varepsilon, A, b, d) = O\left(\frac{\sqrt{d} \cdot \log(1/\varepsilon)}{\varepsilon_{\text{dist}}}\right) \tag{9}$$

with the following properties:

(1) For $x \in [0,1]$ we have $q(x) \in [0,1]$;

(2) $q(x) \in [1 - \varepsilon/2, 1]$ for $x \in [0, A/d]$;

(3) $q(x) \in [0, \varepsilon/8]$ for $x \in [(A+b)/d, 1]$.

We remark that in [She20] Sherstov gives sharp upper and lower bounds for a related problem, by constructing optimal (in terms of degree) polynomials which appproximately achieve a prescribed sequence of 0/1 values at the discrete points $x = 0, \frac{1}{d}, \frac{2}{d}, \ldots, \frac{d-1}{d}, 1$ of the real line. In our setting, though, we require high-accuracy approximators over the *continuous* intervals $[0, A/d]$ and $[(A+b)/d, 1]$, not just at discrete values.

To prove Lemma 20 we will use the following result, which is implicit in the proof of Theorem 3 (the main upper bound) of [Pat92]:

Lemma 21. Fix any $\ell \in [0, m/2)$ and let $g : [0, m] \to [0,1]$ be the piecewise linear continuous function defined by

$$g(x) = \begin{cases} 1 & \text{if } x \in [0, \ell] \\ 1 - (x - \ell) & \text{if } x \in (\ell, \ell+1) \\ 0 & \text{if } x \in [\ell+1, m] \end{cases}.$$

Then there is a polynomial p of degree at most $O(\sqrt{m\ell})$ such that $|p(x) - g(x)| \le 1/3$ for all $x \in [0, m]$. (So taking $r(x) = \frac{3}{5}(p(x) + \frac{1}{3})$, we get that $r(x) \in [0,1]$ for all $x \in [0, m]$; $r(x) \in [\frac{3}{5}, 1]$ for all $x \in [0, \ell]$; and $r(x) \in [0, \frac{2}{5}]$ for all $x \in [\ell+1, m]$.)

We will combine Lemma 21 with a standard construction of "amplifying polynomials" (see e.g. Claim 3.8 of [DGJ+10]). In more detail, let

$$a_k(t) := \mathbf{Pr}[k \text{ tosses of a coin with heads probability } t \text{ yield at least } k/2 \text{ heads}]. \tag{10}$$

It is immediate that a_k is a degree-k polynomial, since

$$a_k(t) = \sum_{j=\lceil k/2 \rceil}^{k} \binom{k}{j} t^j (1-t)^{k-j}.$$

Moreover, a standard additive Hoeffding bound gives that for $k = O(\log(1/\tau))$,

$$a_k(t) \begin{cases} \ge 1 - \tau & \text{for } t \in [3/5, 1] \\ \le \tau & \text{for } t \in [0, 2/5] \end{cases}.$$

Putting the pieces together, we see that given $\varepsilon, \varepsilon_{\text{dist}}, A, b$ as in the statement of Lemma 20, taking Lemma 21's m parameter to be $d/\Theta(\varepsilon_{\text{dist}}^2)$, its ℓ parameter to be $\Theta(1)$, the k parameter of Equation (10) to be $O(\log(1/\tau))$ and τ to be $\varepsilon/8$, we get that the polynomial

$$q(x) := a_k(r(x)), \quad \text{for the polynomial } r \text{ given in Lemma 21},$$

satisfies properties (1), (2) and (3) of Lemma 20, and Lemma 20 is proved. $\square$

5.1.2 The multivariate trigonometric polynomial "bump function"

With Lemma 20 in hand, we build the desired multivariate "bump function" $p(x_1,\dots,x_d)$ over the torus $[0,1)^d$ straightforwardly as follows:

$$p(x_1,\dots,x_d) := q\left(\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}\right). \tag{11}$$

Lemma 22. The polynomial $p(x_1,\dots,x_d)$ is a trigonometric polynomial of total degree at most $2\Psi(\varepsilon,A,b,d)$ (where Ψ is as in Lemma 20) with the following properties: for $\varepsilon,\varepsilon_{\mathrm{dist}},A,b$ as in Lemma 20,

(i) For $x\in[0,1)^d$ we have $p(x)\in[0,1]$;

(ii) If $\mathrm{d_{tor}}(x,0^d)\le\frac{\sqrt{A}}{\pi}$, then $p(x)\in[1-\varepsilon/2,1]$;

(iii) If $x\in[0,1)^d$ has $\mathrm{d_{tor}}(x,0^d)\ge\frac{\sqrt{A+b}}{2}$, then $p(x)\in[0,\varepsilon/8]$.

Proof. The degree bound is immediate from the fact that the polynomial q from Lemma 20 has degree $\Psi(\varepsilon,A,b,d)$. Item (i) follows immediately from item (1) of Lemma 20, since the argument $\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}$ to q in Equation (11) is always between 0 and 1.

For (ii), fix any x such that $\mathrm{d_{tor}}(x,0^d)\le\frac{\sqrt{A}}{\pi}$, i.e.

$$\mathrm{d_{tor}}(x_1,0)^2+\cdots+\mathrm{d_{tor}}(x_d,0)^2\le\frac{A}{\pi^2}.$$

For each $x_i\in[0,1)$ we have that $\sin^2(\pi x_i)=\sin^2(\pi\mathrm{d_{tor}}(x_i,0))\le\pi^2\mathrm{d_{tor}}(x_i,0)^2$, so

$$\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}\le\pi^2\cdot\frac{\mathrm{d_{tor}}(x_1,0)^2+\cdots+\mathrm{d_{tor}}(x_d,0)^2}{d}\le\frac{A}{d},\quad\text{and hence}$$

$$p(x)=q\left(\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}\right)\in[1-\varepsilon/2,1]$$

by part (2) of Lemma 20, as desired.

For item (iii) we use the following claim:

Claim 23. Let x be a point in the n-dimensional torus $[0,1)^d$ such that $\mathrm{d_{tor}}(x,0^d)\ge\gamma$. Then

$$\frac{4\gamma^2}{d}\le\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}\le 1.$$

Proof of Claim 23. The upper bound is immediate. For the lower bound, fix any x such that $\mathrm{d_{tor}}(x,0^d)\ge\gamma$, i.e.

$$\mathrm{d_{tor}}(x_1,0)^2+\cdots+\mathrm{d_{tor}}(x_d,0)^2\ge\gamma^2.$$

For any $x_i\in[0,1)$ we have that $\sin^2(\pi x_i)\ge 4\mathrm{d_{tor}}(x_i,0)^2$, so

$$\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{4}\ge\mathrm{d_{tor}}(x_1,0)^2+\cdots+\mathrm{d_{tor}}(x_d,0)^2\ge\gamma^2$$

which directly gives the lower bound of Claim 23. □

Taking $\gamma=\frac{\sqrt{A+b}}{2}$, Claim 23 gives that the argument $\frac{\sin^2(\pi x_1)+\cdots+\sin^2(\pi x_d)}{d}$ to p in Equation (11) is at least $\frac{A+b}{d}$, and item (iii) follows from item (3) of Lemma 20. This concludes the proof of Lemma 22. □

5.2 Proof of Theorem 19

We prove the theorem via its contrapositive. Fix two distributions D_1, D_2 over the torus $[0,1)^d$ and suppose that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D_1, D_2) \geq \varepsilon$. By Definition 1 (interchanging the roles of D_1 and D_2 if necessary), this means that there exist some $z \in [0,1)^d$ and some $0 \leq \tau \leq \varepsilon_{\mathrm{dist}}$ such that

$$D_1(\mathrm{Ball}(z,\tau)) > D_2(\mathrm{Ball}(z, \tau + \varepsilon_{\mathrm{dist}})) + \varepsilon. \tag{12}$$

We may suppose that $z = 0^d$ (this is without loss of generality by shifting the argument to p by z in what follows). We fix the values of A and b as a function of $\varepsilon_{\mathrm{dist}}$ depending on the value of τ:

- If $0.6\varepsilon_{\mathrm{dist}} \leq \tau \leq \varepsilon_{\mathrm{dist}}$, then we set A and b as follows:
$$A := (\pi\varepsilon_{\mathrm{dist}})^2 \quad \text{and} \quad b := \big((3.2)^2 - \pi^2\big)\varepsilon_{\mathrm{dist}}^2\,.$$
- If $0 \leq \tau < 0.6\varepsilon_{\mathrm{dist}}$, then we set A and b as follows:
$$A := (0.6\pi\varepsilon_{\mathrm{dist}})^2 \quad \text{and} \quad b := \big(4 - (0.6\pi)^2\big)\varepsilon_{\mathrm{dist}}^2\,.$$

It is readily checked that in both cases, we have $0 < b < A$ where $A, b = \Theta(\varepsilon_{\mathrm{dist}}^2)$, and moreover

$$\tau \leq \frac{\sqrt{A}}{\pi} \qquad \text{and} \qquad \tau + \varepsilon_{\mathrm{dist}} \geq \frac{\sqrt{A+b}}{2}\,. \tag{13}$$

In order to establish Theorem 19, we will show that $|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)| > \kappa$ for some $\ell = (\ell_1, \dots, \ell_d)$ with $\|\ell\|_1 \leq 2T$, where

$$\kappa = \frac{\varepsilon}{3}\left(\frac{1}{d}\right)^{O(T)} \qquad \text{and} \qquad T = \Psi(\varepsilon, A, b, d)\,,$$

for $\Psi(\varepsilon, A, b, d)$ from Lemma 20.

Suppose to the contrary that $|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)| \leq \kappa$ for every $\ell \in \mathbb{Z}^d$ with $\|\ell\|_1 \leq 2T$. Let $p(x_1, \dots, x_d)$ be the polynomial from Lemma 22. We then have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x}\sim D_1}\big[p(\boldsymbol{x})\big] &= \int_{x\in[0,1)^d} D_1(x)p(x)\,dx \\
&\geq \int_{x\in \mathrm{Ball}(0^d,\tau)} D_1(x)p(x)\,dx \qquad\qquad (\text{since } D_1, p \geq 0 \text{ on } [0,1)^d) \\
&\geq (1-\varepsilon/2)\cdot D_1\Big(\mathrm{Ball}(0^d,\tau)\Big)\,, \qquad\qquad (14)
\end{aligned}$$

where Equation (14) uses $\tau \leq \frac{\sqrt{A}}{\pi}$ (see Equation (13)) and item (ii) of Lemma 22.

We further have that

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x}\sim D_2}\big[p(\boldsymbol{x})\big] &= \int_{x\in[0,1)^d} D_2(x)p(x)\,dx \\
&= \int_{x\in[0,1)^d} D_2(x)\left(\sum_{\|\ell\|_1\leq 2T} \widehat{p}(\ell)e^{-2\pi i\ell x}\right)dx \qquad (\text{by the total degree bound on } p) \\
&= \sum_{\|\ell\|_1\leq 2T} \widehat{p}(\ell)\left(\int_{[0,1)^d} D_2(x)e^{-2\pi i\ell x}\,dx\right) \qquad\qquad (\text{Fubini})
\end{aligned}$$

$$= \sum_{\|\ell\|_1 \le 2T} \widehat{p}(\ell)\widehat{D_2}(-\ell)$$

$$= \sum_{\|\ell\|_1 \le 2T} \widehat{p}(\ell)\overline{\widehat{D_2}(\ell)} = \sum_{\|\ell\|_1 \le 2T} \overline{\widehat{p}(\ell)}\widehat{D_2}(\ell)\,, \tag{15}$$

where the final line relies on the fact that as D_2 is real valued (since it is a probability density), we have

$$\widehat{D_2}(-\ell) = \overline{\widehat{D_2}(\ell)}\,.$$

Continuing, we thus have

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim D_2}\big[p(\boldsymbol{x})\big] = \sum_{\|\ell\|_1 \le 2T} \overline{\widehat{p}(\ell)}\widehat{D_2}(\ell) \tag{16}$$

$$= \sum_{\|\ell\|_1 \le 2T} \overline{\widehat{p}(\ell)}\widehat{D_1}(\ell) + \sum_{\|\ell\|_1 \le 2T} \overline{\widehat{p}(\ell)}\Big(\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\Big)$$

$$= \mathop{\mathbf{E}}_{\boldsymbol{x}\sim \mathcal{D}_1}\big[p(\boldsymbol{x})\big] + \sum_{\|\ell\|_1 \le 2T} \overline{\widehat{p}(\ell)}\Big(\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\Big)\,, \tag{17}$$

where Equation (17) relies on repeating the steps giving Equation (15) with D_1 instead of D_2. Next, note that $|\overline{\widehat{p}(\ell)}| \le 1$ for all ℓ as $p \in [0,1]$ and additionally we assumed $|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)| \le \kappa$ for all $|\ell|_1 \le 2T$. Since there are at most $d^{O(T)}$ vectors $\ell \in \mathbb{Z}^d$ with $\|\ell\|_1 \le 2T$, we can continue from Equation (17) as follows:

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim D_2}\big[p(\boldsymbol{x})\big] \ge \mathop{\mathbf{E}}_{\boldsymbol{x}\sim \mathcal{D}_1}\big[p(\boldsymbol{x})\big] - \kappa \cdot d^{O(T)} \tag{18}$$

$$\ge (1-\varepsilon/2)\cdot D_1\Big(\mathrm{Ball}(0^d,\tau)\Big) - \kappa \cdot d^{O(T)} \qquad \text{(using Equation (14))}$$

$$\ge (1-\varepsilon/2)\cdot \Big(D_2\Big(\mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})\Big) + \varepsilon\Big) - \kappa d^{O(T)} \tag{19}$$

$$\ge D_2\Big(\mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})\Big) + \varepsilon/2 - \varepsilon^2/2 - \kappa d^{O(T)} \tag{20}$$

$$\ge D_2\Big(\mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})\Big) + \varepsilon/4 - \kappa d^{O(T)}\,. \tag{21}$$

where Equation (19) uses Equation (12), Equation (20) uses that $D_2\Big(\mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})\Big) \le 1$, and Equation (21) uses that $\varepsilon < 1/2$.

On the other hand, we also have that

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim D_2}\big[p(\boldsymbol{x})\big] = \int_{x\in \mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})} D_2(x)p(x)\,dx + \int_{x\notin \mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})} D_2(x)p(x)\,dx$$

$$\le D_2(\mathrm{Ball}(0^d,\tau+\varepsilon_{\mathrm{dist}})) + \varepsilon/8, \tag{22}$$

where Equation (22) uses item (iii) of Lemma 22. (Note that we indeed have that $\tau + \varepsilon_{\mathrm{dist}} \ge \frac{\sqrt{A+b}}{2}$ as required for our application of item (iii) of Lemma 22 thanks to Equation (13).) Since $\kappa = (\varepsilon/9)/d^{O(T)}$, Equation (21) and Equation (22) together give a contradiction. This concludes the proof of Theorem 19. □

6 A d-dimensional lower bound for heavy hitter reconstruction

Our goal in this section is to prove the following:

Theorem 24 (Heavy hitter distance lower bound)**.** There is a suitable absolute constant $c > 0$ such that the following holds. Fix $\varepsilon_{\text{dist}} = 0.49$ and let $\varepsilon : 2^{-d/3} < \varepsilon < 1/170$. There are two probability distributions D_1 and D_2 over $\mathbb{T}^d$ with the following properties:

(a) For every $\ell \in \mathbb{Z}^d$ with $\|\ell\|_1 \le c\sqrt{\frac{d}{\log(1/\varepsilon)}}$, we have $\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| \le 2^{-\Omega\left(\sqrt{d\log(1/\varepsilon)}\right)}$.

(b) $\mathrm{d}_{\text{HH}}^{(\varepsilon_{\text{dist}})}(D_1, D_2) > \varepsilon$.

In words, Theorem 24 shows that the upper bound given in Theorem 19 is nearly best possible: there exist two distributions whose low-degree (almost up to the degree bound of Theorem 19, off by only a $\log(1/\varepsilon)$ factor) Fourier coefficients "nearly match each other" (where the closeness is again quantitatively quite similar to that which Theorem 19 says would be sufficient for small heavy hitter distance), yet the heavy hitter distance between these two distributions is greater than ε.

6.1 A lower bound over the discrete cube $\{\pm 1\}^d$

Most of the work in establishing Theorem 24 is in proving Theorem 25, which shows the existence of two specially crafted probability distributions P_1, P_2 over the discrete cube $\{\pm 1\}^d$. As we will see, the existence of these distributions follows from the existence of certain univariate polynomials with many repeated roots at 1 and a large-magnitude constant term, established by Erdélyi [Erd16].

Before stating the theorem, we introduce some basic notation for distributions over $\{\pm 1\}^d$. For a distribution P over $\{\pm 1\}^d$, we write $P(x)$ to denote the amount of probability weight that P puts on the point $x \in \{\pm 1\}^d$, i.e. we view $P(x)$ as a non-negative function over $\{\pm 1\}^d$ such that $\sum_x P(x) = 1$. For $S \subseteq \{1, \dots, d\}$ we write $\widehat{P}(S)$ to denote the "S-th Fourier coefficient" of P, namely

$$\widehat{P}(S) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{\pm 1\}^d}\left[P(\boldsymbol{x})\chi_S(\boldsymbol{x})\right]$$

where $\chi_S : \{\pm 1\}^d \to \{\pm 1\}$ is $\chi_S(x) = \prod_{i \in S} x_i$. To get a sense of the right scaling for the Fourier coefficients of a distribution over $\{\pm 1\}^d$, note that for any distribution P over $\{\pm 1\}^d$ we have

$$\sum_{S \subseteq \{1,\dots,d\}} \widehat{P}(S)^2 = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \{\pm 1\}^d}\left[P(\boldsymbol{x})^2\right] \in \left[2^{-2d}, 2^{-d}\right],$$

so every Fourier coefficient $\widehat{P}(S)$ satisfies $|\widehat{P}(S)| \le 2^{-d/2}$.

The following theorem says that there are two distributions P_1, P_2 that put significantly different amounts of weight on the point 1^d yet have close-to-matching low-degree Fourier coefficients:

Theorem 25. There is a suitable absolute constant $c > 0$ such that the following holds. Fix an $\varepsilon : 2^{-d/3} < \varepsilon < 1/170$. There are two distributions P_1, P_2 over $\{\pm 1\}^d$ with the following properties:

(1) For every $S \subset \{1, \dots, d\}$ with $|S| \le c\sqrt{\frac{d}{\log(1/\varepsilon)}}$, we have $\left|\widehat{P_1}(S) - \widehat{P_2}(S)\right| \le 2^{-d} \cdot 2^{-\Omega\left(\sqrt{d\log(1/\varepsilon)}\right)}$.

(2) $\left|P_1(1^d) - P_2(1^d)\right| \ge 2\varepsilon$.

We prove Theorem 25 in the rest of Section 6.1, and use it to prove Theorem 24 in Section 6.3.

6.1.1 Setup for the proof of Theorem 25

Given $t \geq 0$, let Prod_t denote the i.i.d. product distribution over $\{\pm 1\}^d$ defined as follows: for each $i \in \{1, \ldots, d\}$, we have $\mathbf{E}_{\boldsymbol{x} \sim \mathrm{Prod}_t}[\boldsymbol{x}_i] = e^{-t}$. For any $z \in \{\pm 1\}^d$ that has ℓ coordinates that are 1 and $d - \ell$ coordinates that are -1, the probability that Prod_t puts on z is

$$\mathrm{Prod}_t(z) = 2^{-d}(1 + e^{-t})^{\ell}(1 - e^{-t})^{d-\ell}. \tag{23}$$

Moreover, for any $S \subseteq \{1, \ldots, d\}$, we have that the S-th Fourier coefficient of $\mathrm{Prod}_t : \{\pm 1\}^d \to \mathbb{R}$ is

$$\widehat{\mathrm{Prod}_t}(S) = 2^{-d} \sum_{x \in \{\pm 1\}^d} \mathrm{Prod}_t(x) \chi_S(x) = 2^{-d} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Prod}_t} \left[\chi_S(\boldsymbol{x})\right] = 2^{-d} \prod_{j \in S} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Prod}_t} [\boldsymbol{x}_j] = 2^{-d} e^{-t|S|}. \tag{24}$$

Given a vector of non-negative probabilities $\bar{p} = (p_0, p_1, \ldots, p_d)$ that sum to 1, and a vector of non-negative values $\bar{t} = (t_0, t_1, \ldots, t_d)$, we write $\mathrm{Mix}_{\bar{p},\bar{t}}$ to denote the mixture distribution that puts weight p_i on component distribution Prod_{t_i} for each $i = 0, 1, \ldots, d$; so $\mathrm{Mix}_{\bar{p},\bar{t}}$ is a distribution over $\{\pm 1\}^d$. By Equation (24) and the linearity of the Fourier transform, we have that

$$\widehat{\mathrm{Mix}_{\bar{p},\bar{t}}}(S) = 2^{-d} \sum_{j=0}^{d} p_j e^{-t_j |S|}. \tag{25}$$

We further observe that the probability that the $\mathrm{Mix}_{\bar{p},\bar{t}}$ distribution puts on the all-1's string is

$$\mathrm{Mix}_{\bar{p},\bar{t}}(1^d) = 2^{-d} \sum_{j=0}^{d} p_j (1 + e^{-t_j})^d. \tag{26}$$

We now partially define the two distributions P_1 and P_2 in the statement of Theorem 25: Set

$$\gamma := \frac{8 \log(1/\varepsilon)}{d} \tag{27}$$

and set

$$P_1 := \mathrm{Mix}_{\bar{\mu},\bar{t}}, \quad \text{where } \bar{\mu} = (\mu_0, \mu_1, \ldots, \mu_d),\ \bar{t} = (t_0, t_1, \ldots, t_d) \text{ where } t_j = j\gamma; \tag{28}$$

$$P_2 := \mathrm{Mix}_{\bar{\nu},\bar{t}}, \quad \text{where } \bar{\nu} = (\nu_0, \nu_1, \ldots, \nu_d) \text{ and as above } \bar{t} = (t_0, t_1, \ldots, t_d) \text{ where } t_j = j\gamma, \tag{29}$$

where the mixing weights $\mu_0, \mu_1, \ldots, \mu_d$ and $\nu_0, \nu_1, \ldots, \nu_d$ will be determined in the next subsection, thus completing the definition of P_1 and P_2.

6.1.2 Proof of Theorem 25

To finalize γ and the mixing weights in $\bar{\mu}$ and $\bar{\nu}$, let's understand what we need from them so that P_1 and P_2 satisfy conditions listed in Theorem 25.

By Equation (25) and Equations (28) and (29) we have for each $S \subseteq \{1, \ldots, d\}$ that

$$\widehat{P_1}(S) = 2^{-d} \sum_{j=0}^{d} \mu_j e^{-j|S|\gamma} \quad \text{and} \quad \widehat{P_2}(S) = 2^{-d} \sum_{j=0}^{d} \nu_j e^{-j|S|\gamma}. \tag{30}$$

Desideratum (1) from the statement of Theorem 25 translates into requiring that

$$\left| \sum_{j=0}^{d} (\mu_j - \nu_j) \left(e^{-|S|\gamma}\right)^j \right| \leq 2^{-\Omega\left(\sqrt{d \log(1/\varepsilon)}\right)}, \tag{31}$$

for all $|S| = 0, 1, \ldots, c\sqrt{d/\log(1/\varepsilon)}$ for some constant $c > 0$. Let $A(x)$ denote the real polynomial

$$A(x) = \sum_{j=0}^{d} a_j x^j, \quad \text{where each } a_j = \mu_j - \nu_j.$$

Given that

$$e^{-|S|\gamma} \geq 1 - |S|\gamma = 1 - 8c\sqrt{\frac{\log(1/\varepsilon)}{d}},$$

the following condition implies desideratum (1):

$$\big|A(x)\big| \leq 2^{\Omega\left(\sqrt{d\log(1/\varepsilon)}\right)}, \quad \text{for all } x \in \left[1 - 8c\sqrt{\frac{\log(1/\varepsilon)}{d}}, 1\right]. \tag{32}$$

To obtain a condition that implies desideratum (2), we will use the following claim:

Claim 26. For any two vectors of mixing weights $\overline{\mu}$ and $\overline{\nu}$ (which define distributions P_1 and P_2, recalling Equations (28) and (29)), we have that $P_1(1^d) - P_2(1^d) = \mu_0 - \nu_0 \pm \varepsilon$.

Proof. By Equation (26) and Equations (28) and (29), we have

$$P_1(1^d) - P_2(1^d) = \sum_{j=0}^{d} 2^{-d} \cdot (\mu_j - \nu_j) \cdot (1 + e^{-j\gamma})^d.$$

When $j = 0$, the summand is simply $\mu_0 - \nu_0$. To bound the rest of the sum, we claim that

$$2^{-d} \sum_{j=1}^{d} (1 + e^{-j\gamma})^d \leq \frac{\varepsilon}{2} \tag{33}$$

(we will prove this below). Given this, since $\sum_{j=1}^{d} |\mu_j - \nu_j| \leq \sum_{j=1}^{d} \mu_j + \nu_j \leq 2$, we get that

$$\left| \sum_{j=1}^{d} 2^{-d} \cdot (\mu_j - \nu_j) \cdot (1 + e^{-j\gamma})^d \right| \leq 2 \cdot \frac{\varepsilon}{2} = \varepsilon,$$

which gives Claim 26.

To establish Equation (33) we break the sum into two parts. For the first part, we have

$$2^{-d} \sum_{j=1}^{1/\gamma} (1 + e^{-j\gamma})^d \leq 2^{-d} \sum_{j=1}^{1/\gamma} \left(2 - \frac{j\gamma}{2}\right)^d < \sum_{j=1}^{1/\gamma} \left(1 - \frac{j\gamma}{4}\right)^d \leq \sum_{j=1}^{1/\gamma} e^{-j\gamma d/4} = \sum_{j=1}^{1/\gamma} \varepsilon^{2j} < \frac{\varepsilon}{4},$$

where we used the choice of γ from Equation (27) and the fact that $\varepsilon < 1/170$. For the rest, we have

$$2^{-d} \sum_{j>1/\gamma}^{d} (1 + e^{-j\gamma})^d \leq 2^{-d} \cdot d \cdot (1 + e^{-1})^d < \frac{\varepsilon}{4},$$

using $\varepsilon > 2^{-d/3}$. Combining these two parts gives Equation (33) as desired. $\square$

We use the following claim to finalize the mixing weights in $\bar{\mu}$ and $\bar{\nu}$:

Claim 27. There exists a real polynomial $A(x) = \sum_{j=0}^{d} a_j x^j$ such that

(i) $\sum_{j\in[0:d]\,:\,a_j\geq 0} a_j = \sum_{j\in[0:d]\,:\,a_j<0} -a_j = 1$;

(ii) Equation (32) holds (for some sufficiently small constant $c > 0$); and

(iii) $|a_0| \geq 3\varepsilon$.

Before proving Claim 27 we explain why it completes the proof of Theorem 25. Given such a polynomial $A(x)$, for each $j \in \{0, 1, \ldots, d\}$, we set the mixing weights as follows:

$$\mu_j = a_j \text{ and } \nu_j = 0 \text{ if } a_j \geq 0; \quad \mu_j = 0 \text{ and } \nu_j = -a_j \text{ if } a_j < 0.$$

These μ_j, ν_j values are non-negative, satisfy $a_j = \mu_j - \nu_j$ for each j, and satisfy $\sum_j \mu_j = \sum_j \nu_j = 1$. As shown above (ii) (i.e. Equation (32)) implies desideratum (1) of Theorem 25. Finally, by Claim 26, if $|\mu_0 - \nu_0| \geq 3\varepsilon$, then $|P_1(1^d) - P_2(1^d)| \geq 2\varepsilon$ giving desideratum (2) of Theorem 25.

6.2 Proof of Claim 27

We use the polynomial that is given by the following result of Erdélyi:

Lemma 28 (Lemma 3.3 of [Erd16])**.** For any $L \in (0, 1/17)$ and any $n \in \mathbb{N}$, there exists a real-coefficient polynomial $A(x) = \sum_{j=0}^{n} a_j x^j$ with $|a_0| \geq L \cdot \sum_{j=1}^{n} |a_j|$ such that A has at least

$$\min\left\{\frac{2}{7}\sqrt{n \cdot (-\ln L)}, n\right\}$$

repeated roots at 1.

We apply Lemma 28 with $n = d$ and $L = 10\varepsilon$ (recall that $\varepsilon < 1/170$) to obtain the polynomial $A(x) = \sum_{j=0}^{d} a_j x^j$ for Claim 27. Note that by scaling we may further assume that $\sum_{j=0}^{d} |a_j| = 2$. With these choices we have that $|a_0| \geq 10\varepsilon(2 - |a_0|)$, hence $|a_0| \geq 3\varepsilon$ (so property (iii) of Claim 27 holds). On the other hand, since A has a root at 1, we have $\sum_{j=0}^{d} a_j = 0$, which together with $\sum_{j=0}^{d} |a_j| = 2$ gives us property (i) of Claim 27.

For property (ii) of Claim 27, we will use the following result of Borwein, Erdélyi and Kós, and the fact that $A(x)$ has at least $\Omega(\sqrt{d \log(1/\varepsilon)})$ repeated roots at 1 (where we used $\varepsilon > 2^{-d/3}$)):

Lemma 29 (Lemma 5.4 of [BEK99])**.** Let $B : \mathbb{C} \to \mathbb{C}$ be defined as $B(x) = \sum_{j=0}^{n} b_j x^j$ satisfying $|b_j| \leq 1$ for all $0 \leq j \leq n$. If B have k repeated roots at 1, then we have

$$\sup_{x\in I} \big|B(x)\big| \leq (n+1)\left(\frac{e}{9}\right)^k, \quad \text{where } I = \left[1 - \frac{k}{9n}, 1\right].$$

Taking $B(x) = A(x)/2$ we have that $|b_j| \leq 1$ for all j, and B has

$$k \geq \frac{2}{7}\sqrt{d \ln\left(\frac{1}{10\varepsilon}\right)}$$

repeated roots at 1. For a suitably small positive choice of the absolute constant c, we have that

$$8c\sqrt{\frac{\log(1/\varepsilon)}{d}} \leq \frac{k}{9d},$$

so we may apply Lemma 29 and we get that for any x in the interval in Equation (32), we have

$$\big|A(x)\big| = 2 \cdot \big|B(x)\big| \leq 2^{-\Theta(k)} = 2^{-\Omega\left(\sqrt{d \log(1/\varepsilon)}\right)}.$$

This gives property (ii) and concludes the proof of Claim 27.

6.3 From the discrete cube to the $\mathbb{T}^d$ torus

We now use the pair of distributions P_1, P_2 over $\{\pm 1\}^d$ from Section 6.1 to give a pair of distributions D_1, D_2 over the torus $\mathbb{T}^d$ such that all low-degree Fourier coefficients are close to matching, yet the heavy hitter distance between D_1 and D_2 is large, thus proving Theorem 24. For ease of reference we restate Theorem 24:

Theorem 24 (Heavy hitter distance lower bound)**.** There is a suitable absolute constant $c > 0$ such that the following holds. Fix $\varepsilon_{\text{dist}} = 0.49$ and let $\varepsilon : 2^{-d/3} < \varepsilon < 1/170$. There are two probability distributions D_1 and D_2 over $\mathbb{T}^d$ with the following properties:

(a) For every $\ell \in \mathbb{Z}^d$ with $\|\ell\|_1 \le c\sqrt{\frac{d}{\log(1/\varepsilon)}}$, we have $\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| \le 2^{-\Omega\left(\sqrt{d\log(1/\varepsilon)}\right)}$.

(b) $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\text{dist}})}(D_1, D_2) > \varepsilon$.

Proof. Let us define the following simple embedding of $\{\pm 1\}^d$ into $\mathbb{T}^d$: For each $x \in \{\pm 1\}^d$,

$$\mathrm{Emb}(x) = \left(\frac{1 - x_1}{4}, \ldots, \frac{1 - x_d}{4}\right),$$

so $\mathrm{Emb}(x)$ ranges over $\{0, 1/2\}^d \subset \mathbb{T}^d$ and in particular we have $\mathrm{Emb}(1^d) = (0, \ldots, 0)$.

We also define the following mapping from $\mathbb{Z}^d$ to subsets of $\{1, \ldots, d\}$ which will be useful for translating between Fourier coefficients over $\mathbb{T}^d$ and Fourier coefficients over $\{\pm 1\}^d$: For each $\ell \in \mathbb{Z}^d$,

$$\mathrm{Par}(\ell) = \left\{j \in \{1, \ldots, d\} : (-1)^{\ell_j} = -1\right\} \quad \text{or equivalently,} \quad \left\{j \in \{1, \ldots, d\} : \ell_j \text{ is odd}\right\}.$$

For $b \in \{1, 2\}$ and $z \in \{\pm 1\}^d$, recall that $P_b(z)$ is the amount of mass that the distribution P_b puts on z. We may write the distribution P_b as

$$P_b(x) = \sum_{z \in \{\pm 1\}^d} P_b(z)\delta_z(x).$$

The distribution D_b is defined as

$$D_b(y) := \sum_{z \in \{\pm 1\}^d} P_b(z)\delta_{\mathrm{Emb}(z)}(y), \tag{34}$$

so the support of D_b (a distribution over $\mathbb{T}^d$) is contained in $\{0, 1/2\}^d$, and for $y \in \{0, 1/2\}^d$, the amount of weight that D_b puts on y is equal to the amount of weight that P_b puts on the unique string $z = \mathrm{Emb}^{-1}(y) \in \{\pm 1\}^d$. From Equation (34) we have that for $\ell \in \mathbb{Z}^d$, the Fourier coefficient

$$\begin{aligned}
\widehat{D_b}(\ell) &= \sum_{z \in \{\pm 1\}^d} P_b(z) e^{2\pi i(\ell \cdot \mathrm{Emb}(z))} \\
&= \sum_{z \in \{\pm 1\}^d} P_b(z)\chi_{\mathrm{Par}(\ell)}(z) && (35) \\
&= 2^d \cdot \mathop{\mathbf{E}}_{\boldsymbol{z} \sim \{\pm 1\}^d}\left[P_b(\boldsymbol{z})\chi_{\mathrm{Par}(\ell)}(\boldsymbol{z})\right] = 2^d \cdot \widehat{P_b}(\mathrm{Par}(\ell)), && (36)
\end{aligned}$$

where Equation (35) holds because if ℓ_j is even then $e^{2\pi i \ell_j(1 - z_j)/4} = 1$ for both $z_j \in \{\pm 1\}$, and if ℓ_j is odd then $e^{2\pi i \ell_j(1 - z_j)/4} = z_j$ for $z_j \in \{\pm 1\}$.

Fix any $\ell \in \mathbb{Z}^d$ such that

$$\|\ell_1\|_1 \le c\sqrt{\frac{d}{\log(1/\varepsilon)}},$$

where $c > 0$ is the absolute constant in Theorem 25. Since $|\mathrm{Par}(\ell)| \le \|\ell\|_1$, by Equation (36) and property (1) of Theorem 25, we have that

$$\left|\widehat{D_1}(\ell) - \widehat{D_2}(\ell)\right| = 2^d \cdot \left|\widehat{P_1}(\mathrm{Par}(\ell)) - \widehat{P_2}(\mathrm{Par}(\ell))\right| \le 2^{-\Omega\left(\sqrt{d\log(1/\varepsilon)}\right)},$$

establishing property (a) of Theorem 24.

It remains to establish property (b) of Theorem 24, i.e. to show that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D_1, D_2) > \varepsilon$. This is simple: let x be the point $0^d \in \mathbb{T}^d$ and let $\tau = 0$. We have that $D_1(x) - D_2(x) = P_1(1^d) - P_2(1^d)$, which is at least 2ε by part (2) of Theorem 25. Moreover, $D_1(\mathrm{Ball}(x,\tau)) = D_1(x) = P_1(1^d)$, and $D_2(\mathrm{Ball}(x, \tau + \varepsilon_{\mathrm{dist}})) = D_2(\mathrm{Ball}(x, \varepsilon_{\mathrm{dist}})) = D_2(x)$ (since x is the only point in the support of D_2 that lies in $\mathrm{Ball}(x, \tau + \varepsilon_{\mathrm{dist}})$, recalling that $\varepsilon_{\mathrm{dist}} < 1/2$), which equals $P_2(1^d)$. So $D_1(\mathrm{Ball}(x,\tau)) - D_2(\mathrm{Ball}(x, \varepsilon_{\mathrm{dist}})) \ge 2\varepsilon$ and hence indeed $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\mathrm{dist}})}(D_1, D_2) > \varepsilon$. This concludes the proof of Theorem 24. $\square$

Acknowledgements

X.C. is supported by NSF grants CCF-2106429 and CCF-2107187. A.D. is supported by NSF grant CCF 2045128. Y.H. is supported by NSF grants CCF-2211238, CCF-2106429, and CCF-2238221 R.A.S. is supported by NSF grants CCF-2211238 and CCF-2106429. T.Y. is supported by NSF grants CCF-2211238, CCF-2106429, and AF-Medium 2212136. T.Y. and Y.H. are also supported by an Amazon Research Award, Google CyberNYC award, and NSF grant CCF-2312242.

The authors would like to thank Ankur Moitra for helpful discussions.

References

[Abb73] Ernst Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9:413–468, 1873. 1

[BEK99] Peter Borwein, Tamás Erdélyi, and Géza Kós. Littlewood-type problems on $[0, 1]$. *Proc. London Math. Soc.*, 79(1):22–46, 1999. 11, 33

[CFG13] E. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013. 1, 2, 3

[CFG14] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-Resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014. 1, 2

[CM21] Sitan Chen and Ankur Moitra. Algorithmic Foundations for the Diffraction Limit. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, page 308–321, New York, NY, USA, 2021. Association for Computing Machinery. 1

[DDP17] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017. 4

[DGJ+10] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, and E. Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010. 26

[Don92] David L Donoho. Superresolution via sparsity constraints. *SIAM journal on mathematical analysis*, 23(5):1309–1331, 1992. 1

[DOS20] Anindya De, Ryan O'Donnell, and Rocco A. Servedio. Sharp Bounds for Population Recovery. *Theory of Computing*, 16(6):1–20, 2020. 11, 12

[dP95] Gaspard Riche de Prony. Essai expérimental et analytique sur la puissance motrice du feu. *Journal de l'École Polytechnique*, 1:24–76, 1795. 2

[DRWY12] Zeev Dvir, Anup Rao, Avi Wigderson, and Amir Yehudayoff. On Learning Boolean Functions from Noisy Examples. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1097–1106. ACM, 2012. 11

[Erd16] Tamás Erdélyi. Coppersmith–Rivlin type inequalities and the order of vanishing of polynomials at 1. *Acta Arithmetica*, 172:271–284, 2016. 30, 33

[ETT+21] Armin Eftekhari, Jared Tanner, Andrew Thompson, Bogdan Toader, and Hemant Tyagi. Sparse non-negative super-resolution—simplified and stabilised. *Applied and Computational Harmonic Analysis*, 50:216–280, 2021. 4

[FG13] Carlos Fernandez-Granda. Support detection in super-resolution. *International Conference on Sampling Theory and Appl.*, pages 145–148, 2013. 2

[FL23] Zhiyuan Fan and Jian Li. Efficient algorithms for sparse moment problems without separation. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3510–3565. PMLR, 2023. 4

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999. 13

[Gia24] George Giapitzakis. Computing the normalization of Jackson Kernel? Mathematics Stack Exchange, 2024. URL: https://web.archive.org/web/20251030203946/https://math.stackexchange.com/questions/4879787/computing-the-normalization-of-jackson-kernel#4879808. 39

[Han99] Leonid G. Hanin. An extension of the Kantorovich norm. *Contemporary Mathematics*, 226:113–130, 1999. 5, 14

[Hen11] Henry. "preservation of lipschitz constant by convolutions", Feb 2011. Answered Feb 17, 2011; accessed Nov 3, 2025. 21

[HK15] Qingqing Huang and Sham M Kakade. Super-resolution off the grid. *Advances in Neural Information Processing Systems*, 28, 2015. 4

[HMPW08] Thomas Holenstein, Michael Mitzenmacher, Rina Panigrahy, and Udi Wieder. Trace reconstruction with constant deletion probability and related results. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 389–398, 2008. 10

[HSSV21] Daniel Hsu, Clayton H. Sanford, Rocco A. Servedio, and Emmanouil Vasileios Vlatakis-Gkaragkounis. On the Approximation Power of Two-Layer Networks of Random ReLUs. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pages 2423–2461. PMLR, 2021. 14

[KR23] Bakytzhan Kurmanbek and Elina Robeva. Multivariate super-resolution without separation. *Information and Inference: A Journal of the IMA*, 12(3):1938–1963, 2023. 4

[KS01] Adam R Klivans and Rocco Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 258–265, 2001. 10

[LA24] Ping Liu and Habib Ammari. A mathematical theory of super-resolution and two-point resolution. *Forum of Mathematics, Sigma*, 12, 2024. 4

[LF16] Wenjing Liao and Albert Fannjiang. MUSIC for Single-Snapshot Spectral Estimation: Stability and Super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):1–32, 2016. 2

[LZ17] Shachar Lovett and Jiapeng Zhang. Noisy Population Recovery from Unknown Noise. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1417–1431, 2017. 11

[MMRS25] Cameron Musco, Christopher Musco, Lucas Rosenblatt, and Apoorv Vikram Singh. Sharper Bounds for Chebyshev Moment Matching, with Applications. In *The Thirty Eighth Annual Conference on Learning Theory (COLT)*, pages 4309–4358, 2025. 4

[Moi15] Ankur Moitra. Super-resolution, extremal functions and the condition number of Vandermonde matrices. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, pages 821–830, 2015. 3

[MS13] Ankur Moitra and Michael E. Saks. A polynomial time algorithm for lossy population recovery. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 110–116. IEEE Computer Society, 2013. 11

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 11

[Pat92] Ramamohan Paturi. On the Degree of Polynomials that Approximate Symmetric Boolean Functions (Preliminary Version). In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC)*, pages 468–474. ACM, 1992. 10, 26

[PP19] Clarice Poon and Gabriel Peyré. Multidimensional sparse super-resolution. *SIAM Journal on Mathematical Analysis*, 51(1):1–44, 2019. 4

[PRT23] B. Piccoli, F. Rossi, and M. Tournus. A Wasserstein norm for signed measures, with application to non-local transport equation with source term. *Commun. Math. Sci.*, 21:1279–1301, 2023. 5, 14

[PSW17] Yury Polyanskiy, Ananda Theertha Suresh, and Yihong Wu. Sample Complexity of Population Recovery. In *Proceedings of the 30th Conference on Learning Theory*

(COLT), volume 65 of *Proceedings of Machine Learning Research*, pages 1589–1618. PMLR, 2017. 11, 12

[Ray79] Lord Rayleigh. Investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274, 1879. 1

[Rud87] Walter Rudin. *Real and complex analysis*. McGraw-Hill, Inc., 1987. 13

[Rud91] Walter Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991. 9

[Sha05] Alexei Shadrin. Lecture 9, lecture notes for approximation theory. University of Cambridge, 2005. URL: https://web.archive.org/web/20050513194146/http://damtp.cam.ac.uk/user/na/PartIIIat/b09.pdf. 39

[She20] Alexander A. Sherstov. Algorithmic polynomials. *SIAM J. Comput.*, 49(6):1173–1231, 2020. 10, 26

[Spa16] C. M. Sparrow. On spectroscopic resolving power. *The Astrophysical Journal*, 44:76, 1916. 1

[SRR18] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. *Information and Inference: A Journal of the IMA*, 7(1):1–30, 2018. 4

[Woo16] David P Woodruff. New algorithms for heavy hitters in data streams. In *International Conference on Database Theory*. Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, 2016. 6, 8

A Small Wasserstein distance implies small heavy hitter distance

Proposition 30. Fix any scale parameter $0 < \varepsilon_{\text{dist}} \leq 1$, and let D_1, D_2 be two distributions over $\mathbb{T}^d$ (so $\int_{x\in\mathbb{T}^d} |D_1(x)|dx = \int_{x\in\mathbb{T}^d} |D_2(x)|dx = 1$) satisfying $\mathrm{d_W}(D_1, D_2) \leq \varepsilon \cdot \varepsilon_{\text{dist}}$. Then $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\text{dist}})}(D_1, D_2) \leq \varepsilon$.

Proof. We show the contrapositive; so suppose that $\mathrm{d}_{\mathrm{HH}}^{(\varepsilon_{\text{dist}})}(D_1, D_2) > \varepsilon$. This means that there exists some $x \in \mathbb{T}^d$ and some $\tau \in [0, \varepsilon_{\text{dist}}]$ such that $D(\mathrm{Ball}(x, \tau)) > D'(\mathrm{Ball}(x, \tau + \varepsilon_{\text{dist}})) + \varepsilon$, where either $D = D_1, D' = D_2$ or vice versa. Hence $\mathrm{d_W}(D, D') \geq \varepsilon \cdot \varepsilon_{\text{dist}}$, since at least an ε amount of mass under D must be moved a distance of at least $\varepsilon_{\text{dist}}$ (to take it from lying inside the radius-τ ball around x to lying outside the radius-$(\tau + \varepsilon_{\text{dist}})$ ball around x) in order to transform D into D'. □

B Proofs for properties of Jackson's kernel

Here we provide missing proofs for Section 2.4. Recall that $\alpha_n = \frac{3}{n(2n^2+1)}$.

Proposition 6.

$$\int_{x\in\mathbb{T}} J_n(x)\,dx = 1.$$

Proof. The proof here is adapted from [Gia24]. We have

$$\frac{\sin(\pi n x)}{\sin(\pi x)} = \frac{e^{\pi i n x} - e^{-\pi i n x}}{e^{\pi i x} - e^{-\pi i x}} = e^{-\pi i (n-1)x} \cdot \frac{e^{2\pi i n x} - 1}{e^{2\pi i x} - 1} = e^{-\pi i (n-1)x} \sum_{j=0}^{n-1} e^{2\pi i j x}, \tag{37}$$

so

$$\frac{\sin^4(\pi n x)}{\sin^4(\pi x)} = e^{-4\pi i (n-1)x} \sum_{j_1,j_2,j_3,j_4 \in \{0,\ldots,n-1\}} e^{2\pi i (j_1+j_2+j_3+j_4)x}.$$

Note that for $t \in \mathbb{Z}$ we have $\int_{x\in\mathbb{T}} e^{2\pi i t x} dx = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases}$, so we thus have

$$\int_{\mathbb{T}} \frac{\sin^4(\pi n x)}{\sin^4(\pi x)} dx = \left| \{(j_1,j_2,j_3,j_4) \in \{0,\ldots,n-1\}^4 \mid j_1+j_2+j_3+j_4 = 2n-2\} \right|.$$

The size of the set above equals the number of $(j_1,j_2,j_3,j_4) \in \mathbb{Z}^4_{\geq 0}$ such that $j_1+j_2+j_3+j_4 = 2n-2$, minus the number of $(j_1,j_2,j_3,j_4) \in \mathbb{Z}^4_{\geq 0}$ such that $j_1+j_2+j_3+j_4 = 2n-2$ and one of j_1,j_2,j_3,j_4 is at least n (there cannot be more than one of them that is at least n since that would make the sum at least $2n$). Thus

$$\int_{\mathbb{T}} \frac{\sin^4(\pi n x)}{\sin^4(\pi x)} = \binom{2n+1}{3} - 4\binom{n+1}{3} = \frac{2n^3+n}{3} = \frac{1}{\alpha_n}. \qquad \square$$

Lemma 7. For any $d, n \geq 1$,

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim J_{d,n}} \left[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{x}, 0^d) \right] \leq \frac{\sqrt{d}}{n}.$$

Proof. The proof here uses ideas from [Sha05, Lemma 9.6].

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x}\sim J_{d,n}} \left[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{x}, 0^d) \right] &\leq \sqrt{\mathop{\mathbf{E}}_{\boldsymbol{x}\sim J_{d,n}} \left[\mathrm{d}_{\mathrm{tor}}(\boldsymbol{x}, 0^d)^2 \right]} \\
&= \sqrt{2^d \int_{[0,1/2]^d} \|x\|_2^2 J_{d,n}(x) dx} && \text{(by symmetry)} \\
&= \sqrt{2^d \int_{[0,1/2]^d} (x_1^2 + \cdots + x_d^2) J_n(x_1) \cdots J_n(x_d) dx_1 \cdots dx_d} \\
&= \sqrt{2^d d \int_{[0,1/2]^d} x_1^2 J_n(x_1) \cdots J_n(x_d) dx_1 \cdots dx_d} && \text{(by symmetry)} \\
&= \sqrt{2d \int_0^{1/2} x_1^2 J_n(x_1) dx_1} && \text{(by Proposition 6)} \\
&\leq \sqrt{\frac{d\alpha_n}{8} \int_0^{1/2} x_1^2 \left(\frac{\sin(\pi n x_1)}{x_1} \right)^4 dx_1} && (\sin(\pi x_1) \geq 2x_1 \text{ for } 0 \leq x_1 \leq 1/2) \\
&\leq \sqrt{\frac{\pi n d \alpha_n}{8} \int_0^{\pi n/2} u^2 \left(\frac{\sin u}{u} \right)^4 du} && \text{(substitute } u := \pi n x_1) \\
&\leq \sqrt{\frac{\pi n d \alpha_n}{8} \left(\int_0^1 u^2 du + \int_1^{+\infty} u^{-2} du \right)} && (\sin u \leq \min\{u, 1\} \text{ for } u \geq 0)
\end{aligned}$$

$$\leq \frac{\sqrt{\pi}}{2} \cdot \frac{\sqrt{d}}{n}. \qquad (\alpha_n \leq 3/(2n^3))$$

□

Lemma 8. For any $d, n \geq 1$ and $\ell \in \mathbb{Z}^d$, $\left|\widehat{J_{d,n}}(\ell)\right| \leq 1$. Moreover, $\widehat{J_{d,n}}(\ell) = 0$ if $\|\ell\|_\infty \geq 2n$.

Proof. Note that

$$\widehat{J_{d,n}}(\ell) = \int_{x \in \mathbb{T}^d} e^{2\pi i(\ell \cdot x)} J_{d,n}(x) dx = \prod_{k=1}^{d} \left(\int_{x_k \in \mathbb{T}} e^{2\pi i \ell_k x_k} J_n(x_k) dx_k \right) = \prod_{k=1}^{d} \widehat{J_{1,n}}(\ell_k),$$

so we only need to consider the case where $d = 1$. When $d = 1$,

$$\begin{aligned}
\widehat{J_{1,n}}(\ell) &= \int_{\mathbb{T}} e^{2\pi i \ell x} \alpha_n \cdot \frac{\sin^4(\pi n x)}{\sin^4(\pi x)} dx \\
&= \alpha_n \int_{\mathbb{T}} e^{2\pi i(\ell - 2n + 2)x} \cdot \left(\frac{e^{2\pi n x} - 1}{e^{2\pi x} - 1} \right)^4 dx && \text{(recalling Equation (37))} \\
&= \alpha_n \int_{\mathbb{T}} e^{2\pi i(\ell - 2n + 2)x} \cdot \left(\sum_{j=0}^{n-1} e^{2\pi j x} \right)^4 dx \\
&= \alpha_n \int_{\mathbb{T}} \sum_{j_1, j_2, j_3, j_4 \in \{0, \ldots, n-1\}} e^{2\pi i(\ell - 2n + 2 + j_1 + j_2 + j_3 + j_4)x} dx \\
&= \alpha_n \cdot \left| \{(j_1, j_2, j_3, j_4) \in \{0, \ldots, n-1\}^4 \mid j_1 + j_2 + j_3 + j_4 = 2n - 2 - \ell\} \right|.
\end{aligned}$$

The size of the set above can be calculated by the number of $(j_1, j_2, j_3, j_4) \in \mathbb{Z}^4_{\geq 0}$ such that $j_1 + j_2 + j_3 + j_4 = 2n - 2 - \ell$, minus the number of $(j_1, j_2, j_3, j_4) \in \mathbb{Z}_{\geq 0}$ such that $j_1 + j_2 + j_3 + j_4 = 2n - 2 - \ell$ and one of j_1, j_2, j_3, j_4 is at least n (there cannot be more than one of them that is at least n since that would make the sum at least $2n$). Thus

$$\widehat{J_{1,n}}(\ell) = \begin{cases} \alpha_n \left(\binom{2n+1-\ell}{3} - 4\binom{n+1-\ell}{3} \right) & 0 \leq \ell \leq n-2 \\ \alpha_n \binom{2n+1-\ell}{3} & n-1 \leq \ell \leq 2n-2 \\ 0 & \ell \geq 2n-1 \end{cases},$$

which can be straightforwardly verified to be at most 1 since $\alpha_n = \frac{3}{n(2n^2+1)}$. □

Lemma 9. For any $d, n \geq 1$, $J_{d,n}$ is $(3\pi n^2 (3n/2)^{d-1} \sqrt{d})$-Lipschitz.

Proof. We first observe that for $0 < x < 1$,

$$\left| \frac{\sin \pi n x}{\sin \pi x} \right| = \left| \frac{e^{\pi i n x} - e^{-\pi i n x}}{e^{\pi i x} - e^{-\pi i x}} \right| = \left| e^{\pi i(n-1)x} + e^{\pi i(n-3)x} + \cdots + e^{\pi i(-n+1)x} \right| \leq n. \tag{38}$$

Next, we show that $0 \leq J_n(x) \leq 3n/2$ for all $x \in \mathbb{T}$. The lower bound is trivial. For the upper bound, when $x = 0$, $J_n(x) = \alpha_n n^4 \leq 3n/2$; when $0 < x < 1$, thanks to Equation (38) we also have that $J_n(x) \leq \alpha_n n^4 \leq 3n/2$.

We now show that $|J_n'(x)|$ is bounded by $3\pi n^2$ for every $x \in \mathbb{T}$. Note that the Taylor series of $\sin t$ at $t = 0$ is $t + O(t^3)$, so we have

$$J_n'(0) = \lim_{x\to 0} \frac{1}{x} \cdot \left(\alpha_n n^4 - \frac{\alpha_n \sin^4(\pi n x)}{\sin^4(\pi x)} \right) = \alpha_n \lim_{x\to 0} \frac{n^4(\pi^4 x^4 + O(x^6)) - (\pi^4 n^4 x^4 + O(x^6))}{x(\pi^4 x^4 + O(x^6))} = 0.$$

For $0 < x < 1$, we have

$$\left| \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)' \right| = \left| \left(\sum_{j=0}^{n-1} e^{\pi i(2j-n+1)x} \right)' \right| = \left| \sum_{j=0}^{n-1} \pi i(2j-n+1) e^{\pi i(2j-n+1)x} \right| \leq \pi \sum_{j=0}^{n-1} |2j-n+1| \leq \frac{\pi n^2}{2},$$

so

$$\left|J_n'(x)\right| = \left| 4\alpha_n \cdot \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)' \cdot \left(\frac{\sin(\pi n x)}{\sin(\pi x)} \right)^3 \right| \leq \frac{6}{n^3} \cdot \frac{\pi n^2}{2} \cdot n^3 = 3\pi n^2.$$

It follows that

$$\left| \frac{\partial J_{d,n}(y)}{\partial y_k} \right| = \left| J_n'(y_k) \cdot \prod_{1\leq j\leq d, j\neq k} J_n(y_j) \right| \leq 3\pi n^2 \cdot \left(\frac{3n}{2} \right)^{d-1},$$

so by Cauchy-Schwarz, the function $J_{d,n}$ is $(3\pi n^2 (3n/2)^{d-1}\sqrt{d})$-Lipschitz. □