Oblivious near-optimal sampling for multidimensional signals with Fourier constraints

Xingyu Xu
Tsinghua University

Yuantao Gu
Tsinghua University

Abstract

We study the problem of reconstructing a continuous multidimensional signal from a small number of samples under Fourier constraints assuming that the Fourier power spectrum of the signal has some desirable properties, e.g. being compactly supported, being sparse. We further assume that the Fourier constraint can be expressed as a prior distribution on the Fourier power spectrum, which subsumes the aforementioned examples. The study of sampling and reconstructing in this vein has attracted much attention with a long history. In this paper, we are interested in finding oblivious sampling strategies, that is, sampling without knowing what specific constraint is put on the Fourier power spectrum. We show that it is possible to obliviously sample a Fourier-constrained multidimensional signal with a near-optimal (up to a logarithmic factor) number of samples that guarantee successful reconstruction, partially answering an open question in Avron et al. (2019) which considered the 1-dimensional case. Our approach highlights a phenomenon that is unique for dimension $d \geq 2$ that the sampling strategy should depend on the geometry of the region on which the signal is to be reconstructed, unlike the case $d = 1$ where all regions are of the form $[a, b]$ which are all geometrically equivalent. Our proof, using tools from convex geometry, also illuminates an idea obscured in $d = 1$, that to reconstruct a signal in a given region, it can be helpful to take some samples outside that region.

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

1 INTRODUCTION

A fundamental problem in science and engineering is to reconstruct a continuous signal, say $f(x)$, from a few of its samples $f(x_1), \ldots, f(x_k)$. The signal can be multidimensional[1], i.e. x can take values in $\mathbb{R}^d$. It is obvious that such reconstruction is possible only if one assumes sufficient "regularity" of f with some proper notion of "regularity". Apart from the most straight-forward smoothness assumptions, an important class of regularity assumptions are *Fourier constraints*, which require the Fourier spectrum f to have some desirable properties. In this work we consider a special form of Fourier constraints, following Ramani et al. (2005); Eldar and Unser (2006); Avron et al. (2019), which is simple yet powerful enough to express most commonly-used Fourier constraints. Specifically, we model the assumptions on the Fourier power spectrum as a *prior distribution* μ, thereby expressing f as

$$f(x) = \int \alpha(\xi) e^{i\langle \xi, x \rangle} d\mu(\xi). \tag{1}$$

This model subsumes common Fourier constraints like bandlimited constraints, spectral decay constraints and Fourier-sparse constraints, which are absolutely fundamental and have been extensively studied in signal processing, communication, image processing, etc. (Rasmussen, 2003; Ripley, 2005; Mishali and Eldar, 2010). Furthermore, it has a clear meaning within Bayesian approaches to signal reconstruction. The latter point links our model to the dominantly useful "Gaussian process regression" and "kriging" methods for fitting continuous signals in diverse scientific disciplines including geostatistics and economics (Ripley, 2005; Ramani et al., 2006); see Avron et al. (2019) for a precise formulation of this link.

In this paper we investigate oblivious sampling strategies under model (1), which in our settings means we will sample the signal without knowing what μ is. Of course, we will nevertheless have to use information about μ when reconstructing the signal using the samples, but oblivious sampling remains to be of interest since it allows to apply

[1] Please be aware that, following conventions in signal processing, this means the *domain* of f rather than the *range* of f is multidimensional.

the same sampling strategy, which in practice may correspond to a fixed hardware design, to different tasks. It was first shown by Avron et al. (2019) that near-optimal oblivious sampling is possible for one-dimensional signal. The more general multidimensional case is proposed as an open problem therein. Multidimensional signals are often encountered in practice, e.g. in array signal processing, medical imaging (Friston et al., 1994; Mishali and Eldar, 2010). Our main result provides a partial answer to this problem by establishing a near-optimal oblivious sampling strategy for a wide class of regions which covers most practical purposes. More concretely, we show that for multidimensional signals,

- There is an oblivious sampling strategy whose sample complexity is at most a polynomial of the optimal sample complexity, regardless of the prior distribution μ, and achieves accurate reconstruction.
- If the reconstruction region is a polyhedron (or a simplicial complex) with not too many vertices, the above strategy is near-optimal.
- If we allow sampling outside the reconstruction region, there is similarly an oblivious sampling strategy that achieves near-optimal sample complexity and accurate reconstruction under very mild assumption on the reconstruction region (see Section 4.3).

We remark that even the first point is already non-trivial: to the best of our knowledge no comparable result has been obtained in the multidimensional setting before, considering that it is already difficult (Slepian, 1964) to obtain satisfactory result for very specific choices of the reconstruction region and the Fourier prior μ.

Arguably, in practice the most important cases are $d = 1, 2, 3, 4$, e.g. in array signal processing, geostatistics and image processing. For this reason, in this work we focus on *multidimensional* signals rather than *high-dimensional* signals. Technically speaking, this means we will not try to optimize the dependence on dimension d. Instead, we are free to think of "constants that depend only on d" as absolute constants, though we will always make it clear when we are doing so.

1.1 Related Works

The study of optimal sampling and reconstruction strategy for Fourier-constrained signals has a long history, dating back to Nyquist (1928); Kotelnikov (1933); Shannon (1949); it is probably futile to try to make a comprehensive list of related works here. We only take some samples of them that are most representative and most closely related to our work. Classical works such as Nyquist (1928); Kotelnikov (1933); Shannon (1949) mainly considered bandlimited unidimensional signals, corresponding to μ being the normalized Lebesgue measure on a fixed interval. Subsequent works in this setting have managed to achieve optimal sample complexity using techniques stemming from the so-called prolate spheroidal wave functions (Slepian and Pollak, 1961; Landau and Pollak, 1961; Xiao et al., 2001; Osipov and Rokhlin, 2014). These works focus on the bandlimited signals and are very difficult to generalize; even generalizing to multiband signals can take a substantial amount of efforts (Zhu and Wakin, 2017).

It was Avron et al. (2019) who provide the first oblivious near-optimal sampling strategy in dimension one. It is left as open problem there whether their results can be extended to higher dimensions, and our work provides a partial affirmative answer to this problem. As we will see soon, the multidimensional setting is much more challenging due to the complicated geometry of multidimensional regions, in sharp contrast to the unidimensional setting where all regions are geometrically equivalent to the unit interval. It takes novel ideas and substantial efforts to overcome such challenges. More discussions on the technical and methodological differences between our work and Avron et al. (2019) can be found at the end of this subsection.

Sampling multidimensional signals has been studied for some specific regions in Slepian (1964) for bandlimited signals, though with a much less clear and less definitive answer than in dimension one. Modern works on this problem also concentrate on specific forms of μ and are hard to generalize (Pesquet-Popescu and Véhel, 2002; Ramani et al., 2006).

Another type of Fourier-constrained signals that have attracted significant attention in recent years are Fourier-sparse signals (Donoho, 2006; Eldar and Kutyniok, 2012). There is a major difference between our work and these works: sparse recovery is interested in recovering the signal with μ sparse but unknown, while we do not assume μ to be sparse but do need μ to be known for reconstruction (though the knowledge of μ is not required in sampling). It is an interesting future direction to see if these two techniques can be combined to obtain better sampling and reconstruction schemes.

Comparison with Avron et al. (2019). On a high-level, our proof is greatly inspired by Avron et al. (2019). However, the multidimensional setting we consider has a fundamental difference from the unidimensional setting there, which requires new ideas and new tools to deal with. Our result reveals that the (near-)optimal sampling strategy depends crucially on the *geometry* of the sampling region Ω, a phenomenon that manifests itself only in the multidimensional setting for the following reason. In multidimensional setting, one has to take into account the fact that the region Ω on which the signal is to be sampled and reconstructed can have very different shapes, or geometries. This does not happen in unidimensional setting, where all regions are of the form $[a, b]$ which are all geometrically equivalent. In fact, this difference makes it more difficult to design near-optimal sampling strategy for a general region Ω which calls for new ideas not present in the unidimensional case and also new tools from convex geometry to accomplish better generality in presence of various geometries, c.f. Section 4.

1.2 Basic Notations

By a *region* we mean a compact set in $\mathbb{R}^d$ whose interior is non-empty. For a region Ω, denote its volume by $|\Omega|$. A *convex body* is a convex region. Denote $a \wedge b \triangleq \min(a, b)$, $\tilde{O}(N) = O(N \log^{O(1)} N)$. By $a \asymp b$ we mean that $a = O(b)$ and $b = O(a)$ hold simultaneously. By C_d or $C(d)$ we mean a constant that depending only on d, and $a = O_d(b)$ means $a \leq C(d)b$ for some constant $C(d)$ depending only on d. Similarly we define $\tilde{O}_d(N) = O_d(\log^{O_d(1)} N)$, and by $a \asymp_d b$ we mean $a = O_d(b)$ and $b = O_d(a)$. By $a = \mathrm{poly}(b)$ (resp. $a = \mathrm{poly}_d(b)$) we mean $a = O(b^{O(1)})$ (resp. $a = O_d(b^{O_d(1)})$). Other notations will be introduced upon their first occurrence.

2 BACKGROUNDS

In this section we discuss our basic settings, which are natural extensions of those in Avron et al. (2019) to multidimensions. Recall that we are interested in signals f of the form (1) where μ is a probability measure on $\mathbb{R}^d$. Furthermore, we consider the noisy version of (1):

$$\tilde{f}(x) = \int \alpha(\xi) \mathrm{e}^{i\langle \xi, x\rangle} d\mu(\xi) + n(x), \tag{2}$$

where n denotes the noise. We allow the noise to be adversarial, i.e. n can be an arbitrary function. We would like to reconstruct f on a given region Ω. To measure the reconstruction accuracy we introduce the following standard norms. Denote by $\|\cdot\|_\Omega$ and $\|\cdot\|_\mu$ the L^2-norm associated to the normalized Lebesgue measure on Ω and to μ respectively, i.e.

$$\|g\|_\Omega \triangleq \left(\frac{1}{|\Omega|}\int_\Omega |g(x)|^2 dx\right)^{1/2},$$
$$\|\beta\|_\mu \triangleq \left(\int |\beta(\xi)|^2 d\mu(\xi)\right)^{1/2},$$

where $|\Omega|$ denotes the volume of Ω. Similarly we introduce the notations $\langle\cdot,\cdot\rangle_\Omega$ and $\langle\cdot,\cdot\rangle_\mu$ for inner products and the notations $L^2(\Omega)$, $L^2(\mu)$ for the corresponding Hilbert spaces. We seek to prove reconstruction guarantees in the following form:

Definition 2.1. *A function $\hat{f}$ is said to be an* (ϵ, K)-accurate *approximation of f if*

$$\|\hat{f} - f\|_\Omega^2 \leq \epsilon\|\alpha\|_\mu^2 + K\|n\|_\Omega^2. \tag{3}$$

Throughout this paper we will simply take K to be a sufficiently large constant and omit its appearance; that is, $\hat{f}$ is said to be an ϵ-accurate *approximation of f if* (3) *holds for some K that is sufficiently large and fixed throughout this paper.*

The above form of reconstruction error and its variants are pervasive in relevant literature (Bhaskar et al., 2013; Avron et al., 2019; Jin et al., 2020). The term $\|n\|_\Omega^2$ is the total energy of the noise. Generally, it is not possible to guarantee a reconstruction error less than $\|n\|_\Omega^2$, so up to a multiplicative constant the reconstruction accuracy is characterized by ϵ, justifying the simplified notation "ϵ-accurate". The term $\|\alpha\|_\mu^2$ can be interpreted as the "energy" of the signal f. For more discussion on Definition 2.1, please refer to Avron et al. (2019).

For the sequel we denote by $\mathcal{F}$ the Fourier transform operator $L^2(\mu) \to L^2(\Omega)$:

$$\mathcal{F}\beta = \int \beta(\xi)\mathrm{e}^{i\langle\xi,\cdot\rangle} d\mu(\xi), \quad \forall \beta \in L^2(\mu).$$

The adjoint of $\mathcal{F}$ can be explicitly expressed as

$$\mathcal{F}^* g = \frac{1}{|\Omega|}\int_\Omega g(x)\mathrm{e}^{-i\langle\cdot,x\rangle} dx, \quad \forall g \in L^2(\Omega).$$

Note that by Cauchy-Schwarz inequality, $\mathcal{F}\beta$ (resp. $\mathcal{F}^* g$) can be defined pointwise rather than as an element in $L^2(\Omega)$ (resp. $L^2(\mu)$). We will use notations like $(\mathcal{F}\beta)(x)$ freely in the sequel.

2.1 Reconstruction via Ridge Regression

Following Avron et al. (2019), we reconstruct f using ridge regression. For pedagogical reason, we first consider the case where we have full knowledge of $\tilde{f}$. We may then find an estimate of f by solving

$$\hat{\beta} = \operatorname*{arg\,min}_{\beta \in L^2(\mu)} \|\mathcal{F}\beta - \tilde{f}\|_\Omega^2 + \epsilon\|\beta\|_\mu^2, \tag{4}$$
$$\hat{f} = \mathcal{F}\hat{\beta}. \tag{5}$$

It was shown in Avron et al. (2019) (and is well-known) that the $\hat{f}$ obtained in (5) is a $C\epsilon$-accurate reconstruction of f for some absolute constant $C > 0$, even if $\hat{\beta}$ is only an $\sqrt{C}$-approximate solution[2] of (4).

In reality we only have access to a few samples $\tilde{f}(x_1), \ldots, \tilde{f}(x_k)$ of $\tilde{f}$. In this case we need to properly discretize (4). It turns out that it is very useful to assign to each sample a certain weight and then approximate the continuous integral $\|\cdot\|_\Omega^2$ by a weighted discrete sum. In

[2]This means $\|\mathcal{F}\hat{\beta} - \tilde{f}\|_\Omega^2 + \epsilon\|\hat{\beta}\|_\mu^2 \leq \sqrt{C}\min_\beta(\|\mathcal{F}\beta - \tilde{f}\|_\Omega^2 + \epsilon\|\beta\|_\mu^2)$.

particular, we will choose some weights $w_1,\dots,w_k \ge 0$, and define a norm $\|\cdot\|_w$ on $\mathbb{C}^k$ by

$$\|z\|_w^2 = \sum_{j=1}^k w_j |z_j|^2$$

where $z = (z_1,\dots,z_k)$. Denote by $\tilde{f}_X = (\tilde{f}(x_j))_{j=1}^k$ the vector formed by the samples at our disposal, and by $\mathcal{F}_X : L^2(\mu) \to \mathbb{C}^k$ the operator obtained by sampling $\mathcal{F}$ at $x_1,\dots,x_k$, e.g. $\mathcal{F}_X\beta = \left(\int \beta(\xi) \mathrm{e}^{i\langle \xi, x_j\rangle} d\mu(\xi)\right)_{j=1}^k$. The discrete version of (4) can be expressed as

$$\hat{\beta} = \underset{\beta \in L^2(\mu)}{\arg\min} \|\mathcal{F}_X\beta - \tilde{f}_X\|_w^2 + \epsilon\|\beta\|_\mu^2, \tag{6}$$

meanwhile $\hat{f}$ is still obtained as in (5). Though (6) is infinite-dimensional in nature, it does not need to be solved in practice since we are only interested in $\hat{f}$. By elementary algebraic manipulation one may obtain an effectively computable analytic expression for $\hat{f}$ assuming access to an oracle which computes the kernel $k_\mu(x,x') \triangleq \int \mathrm{e}^{i\langle \xi, x-x'\rangle} d\mu(\xi)$. In many practical scenarios such an oracle exists and each access to the oracle takes $O(1)$ time. For more details on algorithmic aspect of the ridge regression approach to reconstruction we refer the interested readers to Avron et al. (2019) or the extended version of this paper.

Algorithm 1 Reconstruct

Input: Samples $\tilde{f}_X = (\tilde{f}(x_1),\dots,\tilde{f}(x_k))$, sample locations $x_1,\dots,x_k$, weights $w_1,\dots,w_k \ge 0$.
Output: An estimation $\hat{f}$ of f.
1. Solve (6) to find the optimal $\hat{\beta}$.
2. Let $\hat{f} = \mathcal{F}\hat{\beta}$.
3. Output $\hat{f}$.

2.2 Leverage Score Sampling

Next we introduce an important method, called *leverage score sampling* (Spielman and Srivastava, 2011; Drineas and Mahoney, 2016; Cohen and Migliorati, 2017), to choose the sample points x_i's and the weights w_i's near-optimally.

Definition 2.2 (Leverage score)**.** *Given $\epsilon > 0$, the* leverage score function $\tau_{\mu,\Omega,\epsilon}$ *is defined on Ω as*

$$\tau_{\mu,\Omega,\epsilon}(x) \triangleq \frac{1}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|(\mathcal{F}\beta)(x)|^2}{\|\mathcal{F}\beta\|_\Omega^2 + \epsilon\|\beta\|_\mu^2}$$

for $x \in \Omega$. The effective dimension $s_{\mu,\Omega,\epsilon}$ *is defined as*

$$s_{\mu,\Omega,\epsilon} = \operatorname{tr}(\mathcal{F}(\mathcal{F}^*\mathcal{F} + \epsilon\mathcal{I})^{-1}\mathcal{F}^*).$$

We put the definitions of $\tau_{\mu,\Omega,\epsilon}$ and $s_{\mu,\Omega,\epsilon}$ together for good reasons. In fact, it is not hard (Avron et al., 2019) to show that $\tau_{\mu,\Omega,\epsilon}$ is continuous, and

$$s_{\mu,\Omega,\epsilon} = \int_\Omega \tau_{\mu,\Omega,\epsilon}(x)dx. \tag{7}$$

What we would like to stress here is that $\tau_{\mu,\Omega,\epsilon}$, or an upper bound of it, provides very useful information to choose x_i's and w_i's:

Lemma 2.1 (Near-optimality of leverage score sampling)**.** *There exists some absolute constant $C > 0$ such that the following holds. Assume $\tilde{\tau}$ (called* leverage score bound*) is a measurable function satisfying $\tilde{\tau} \ge \tau_{\mu,\Omega,\epsilon}$. Let $\tilde{s} = \int_\Omega \tilde{\tau}(x)dx$. Let $x_1,\dots,x_k$ be i.i.d. samples drawn from the distribution on Ω given by p.d.f. $\tilde{\tau}/\tilde{s}$. Let $w_j = \frac{1}{k|\Omega|}\frac{\tilde{s}}{\tilde{\tau}(x_j)}$ for $j = 1,\dots,k$. For any $\delta > 0$, whenever $k \ge C\tilde{s}(\log\tilde{s} + 1/\delta)$, the solution to (6) is a 3-approximate solution of (4) with probability at least $1-\delta$. In that case, the estimate $\hat{f}$ obtained by (6) and (5) is $C'\epsilon$-accurate for some absolute constant $C' > 0$. On the other hand, under mild assumptions on μ and Ω, at least $\Omega(s_{\mu,\Omega,\epsilon})$ samples are required to achieve $C\epsilon$-accuracy.*

Algorithm 2 LeverageScoreSamp

Input: Leverage score bound $\tilde{\tau}$, sampling region Ω, number of samples k, an oracle to sample $\tilde{f}$ at any given point in Ω.
Output: Sample locations $x_1,\dots,x_k \in \Omega$, samples $\tilde{f}_X$, weights $w_1,\dots,w_k$.
1. Compute $\tilde{s} = \int_\Omega \tilde{\tau}(x)dx$.
2. Take k i.i.d. samples $x_1,\dots,x_k$ from the distribution on Ω with p.d.f. $\tilde{\tau}/\tilde{s}$.
3. Sample $\tilde{f}$ at $x_1,\dots,x_k$, forming a vector $\tilde{f}_X = (\tilde{f}(x_1),\dots,\tilde{f}(x_k))$.
4. Compute $w_j = \frac{1}{k|\Omega|}\frac{\tilde{s}}{\tilde{\tau}(x_j)}$ for $j = 1,\dots,k$.
5. Output $x_1,\dots,x_k$, $\tilde{f}_X$ and $w_1,\dots,w_k$.

The proof of the upper bound is a simple application of operator Bernstein inequality (Minsker, 2017), c.f. Avron et al. (2019) where a formal statement and the proof of the lower bound can also be found[3]. By Lemma 2.1, to obtain a near-optimal sampling strategy, one only needs to find an upper bound $\tilde{\tau}$ of $\tau_{\mu,\Omega,\epsilon}$ satisfying $\tilde{s} = \int_\Omega \tilde{\tau}(x)dx = \tilde{O}(s_{\mu,\Omega,\epsilon})$. The rest part of this paper is in fact devoted to proving such bounds.

3 SAMPLING ON THE UNIT CUBE

In this section we consider the most fundamental case where $\Omega = [0,1]^d$ is the unit cube. Let $\rho(x) = 1/(x \wedge (1-x))$ on $(0,1)$ and $\rho(0) = \rho(1) = +\infty$, making ρ a function defined on $[0,1]$. One may recognize that $\kappa(\kappa^5 \wedge \rho)$

[3]Note that Avron et al. (2019) did not consider the role of Ω, but since Lemma 2.1 is a highly abstract statement hiding all explicit dependences on Ω, the proof remains verbatim.

is the unidimensional leverage score bound obtained in Avron et al. (2019). Our first result shows this can be naturally generalized to a tight leverage score bound on the unit cube (Theorem 3.1), yielding a near-optimal sampling strategy on unit cube (Corollary 3.1; see also Fig. 1(b)). Note that this generalization is not a trivial one that follows from some tensorization trick: though the unit cube can be viewed as some "tensorized" version of $[0,1]$, the Fourier prior μ may not be a tensor product of unidimensional distributions. Neither is the leverage score bound $\tilde{\tau}_\kappa$ below a tensorization of the bound in unidimensional setting. One has to delve into a certain technical depth to reach such generalization.

Theorem 3.1. *Let $\rho_d(x) = \prod_{j=1}^d \rho(x_j)$ for $x = (x_1, \ldots, x_d) \in \Omega$. Let*

$$\tilde{\tau}_\kappa = \kappa(\kappa^{5d} \wedge \rho_d), \tag{8}$$

with $\kappa > 0$ being a parameter to be chosen later. Then $\tilde{s}_\kappa \triangleq \int_\Omega \tilde{\tau}_\kappa(x)dx = \tilde{O}_d(\kappa)$. Furthermore, there exists some absolute constant $C > 0$ such that $\tau_{\mu,\Omega,\epsilon} \leq \tilde{\tau}_\kappa$ if $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$.

The implication of this theorem on sampling strategy will be given soon in Corollary 3.1 below. Before that we introduce a useful notation stemming from Theorem 3.1:

Definition 3.1. *A* well-parametrized family of leverage score bounds *on Ω is a parametrized function $\tilde{\tau}_\kappa$ defined on Ω with parameter $\kappa > 0$, satisfying $\tilde{s} \triangleq \tilde{\tau}_\kappa(x)dx = \tilde{O}_d(\kappa)$ and that for any probability measure μ it holds $\tau_{\mu,\Omega,\epsilon} \leq \tilde{\tau}_\kappa$ whenever $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$.*

With this definition, Theorem 3.1 provides a well-parametrized family of leverage score bounds on the unit cube. The importance of well-parametrized family of leverage score bounds can be seen from the following

Lemma 3.1. *Let $\{\tilde{\tau}_\kappa\}_\kappa$ be a well-parametrized family of leverage score bounds on Ω', where Ω' is an arbitrary region. Then* LeverageScoreSamp *(Algorithm 2) with leverage score bound $\tilde{\tau}_\kappa$ takes $\tilde{O}_d(\kappa+1/\delta)$ (where $\delta > 0$ is arbitrary) samples and guarantees $C\epsilon$-accurate reconstruction using* Reconstruct *(Algorithm 1) with probability at least $1-\delta$ whenever $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$.*

By Lemma 2.1 we know that $\Omega(s_{\mu,\Omega,\epsilon})$ is a lower-bound for most cases, thus well-parametrized family of leverage score bounds provides a way to derive near-optimal sampling strategies.

Corollary 3.1 (Near-optimal sampling on a cube)**.** *With $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$ and $\tilde{\tau}_\kappa$ as in Theorem 3.1,* LeverageScoreSamp *(Algorithm 2) with leverage score bound $\tilde{\tau}_\kappa$ takes $\tilde{O}_d(\kappa + 1/\delta)$ (where $\delta > 0$ is arbitrary) samples and guarantees $C\epsilon$-accurate reconstruction with probability using* Reconstruct *(Algorithm 1) at least $1-\delta$.*

The requirement $\kappa \geq Cs_{\mu,\Omega,\epsilon}$ does not imply that the sampling strategy needs to depend on μ. Rather, one should think of κ as a parameter controlling the number of samples: larger κ corresponds to more samples and leads to smaller ϵ, i.e. better reconstruction accuracy.

4 SAMPLING ON GENERAL REGIONS

In this section we present several techniques to generalize the results in Section 3 to different reconstruction region Ω's. The simplest generalization follows from the fact that our sampling strategy is equivariant with respect to affine transforms, which allows Ω to be any parallelepiped (Corollary 4.1). For more general region Ω, we provide a leverage score bound in terms of the boundary behavior of Ω which, when applied to polyhedra with "not too many" vertices, still yield near-optimal results (Theorem 4.3). However, for some Ω it is not clear how to find near-optimal sampling strategy with this approach, even though the sub-optimal sampling strategy given by this approach for general convex regions (Theorem 4.1) is already highly non-trivial and has not been obtained before. Fortunately, this can be remedied by assuming we have access to samples of $\tilde{f}$ *outside* Ω. An alternative viewpoint of this assumption is to think that we can sample $\tilde{f}$ on a larger region Ω' and would like to reconstruct $\tilde{f}$ on a smaller region $\Omega \subset \Omega'$. In this case, we may still attain near-optimal sampling under mild assumptions on Ω (Theorem 4.4). To facilitate understanding, we depict in Fig. 1 the probability density used for our sampling strategy for different regions.

4.1 Affine Equivariance

Any affine transform can be expressed as a composition of a linear transform and a translation. We observe that the leverage score function is equivariant with respect to linear transforms and translations, hence to affine transforms.

Lemma 4.1. *Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be an invertible linear transform, and $b \in \mathbb{R}^d$ be a vector. Denote $A\Omega = \{Ax : x \in \Omega\}$ and $\Omega + b = \{x + b : x \in \Omega\}$. Then*

$$\tau_{\mu,A\Omega,\epsilon}(x) = \frac{1}{|\det A|}\tau_{\mu\circ A^{*-1},\Omega,\epsilon}(A^{-1}x),\ \forall x \in A\Omega,$$

$$\tau_{\mu,\Omega+b,\epsilon}(x) = \tau_{\mu,\Omega,\epsilon}(x-b),\ \forall x \in \Omega + b,$$

where $\mu \circ A^{*-1}$ denotes the push-forward of μ by A^*, i.e. $\mu \circ A^{*-1}(E) = \mu(A^{*-1}E)$ for any measurable set E.

Assume we have a well-parametrized family of leverage score bounds for some region Ω_0 as in Theorem 3.1. Then for any Ω that is an image of Ω_0 under some affine transform, we may use Lemma 4.1 to produce a well-parametrized family of leverage score bounds for Ω_0. This process is summarized in Algorithm 3:

By Lemma 4.1 it is easy to prove the following

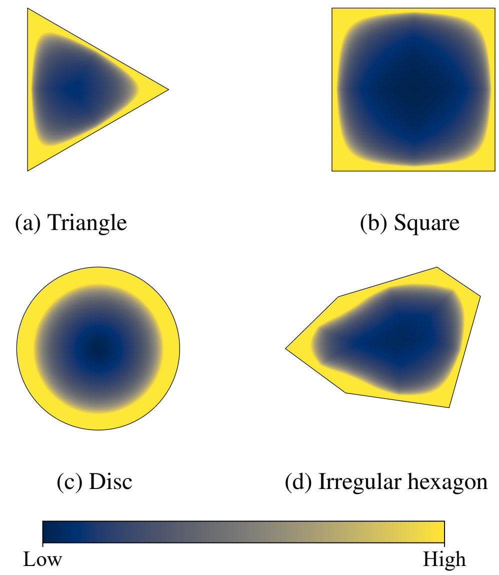

Figure 1: Sampling probability density

Algorithm 3 AffineReduce

Input: A region Ω_0; a well-parametrized family of leverage score bounds $\{\tilde{\tau}^0_\kappa\}_\kappa$ on Ω_0; an invertible linear operator A; a vector $b \in \mathbb{R}^d$; the affine-transformed region $\Omega = A\Omega_0 + b$; an oracle to sample $\tilde{f}$ at any given point in Ω.

Output: A parametrized family of leverage score bounds $\{\tilde{\tau}_\kappa\}_\kappa$ on Ω;

1. Compute $B = A^{-1}$, $\sigma = |\det A|$.
2. Let $\tilde{\tau}_\kappa = \tilde{\tau}^0_\kappa(B(x-b))/\sigma$.
3. Output $\{\tilde{\tau}_\kappa\}_\kappa$.

Lemma 4.2. *The output of* AffineReduce *(Algorithm 3) is a well-parametrized family of leverage score bounds on* Ω.

Applying this argument to $\Omega_0 = [0,1]^d$, the conclusion of 3.1 and Corollary 3.1 can be directly generalized to the case where Ω is a parallelepiped.

Corollary 4.1 (Near-optimal sampling on a parallelepiped)**.** *With* $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$, *running* ParallelepipedSamp *(Algorithm 4) takes* $\tilde{O}_d(\kappa + 1/\delta)$ *(where* $\delta > 0$ *is arbitrary) samples and guarantees* $C\epsilon$*-accurate reconstruction using* Reconstruct *(Algorithm 1) with probability at least* $1-\delta$.

4.2 A General Bound

In this part we consider a general region Ω and proves a bound that generalizes previous results but are not necessarily optimal in some cases. We first set up some technical notations. For a point $x \in \Omega$, denote by Ω^{sym}_x the convex body of maximum volume contained in Ω that is symmetric with respect to x, and by Ω^{st}_x the maximal star-shaped set with respect to x contained in Ω. More precisely, denoting by $[u,v]$ the segment connecting u and v for $u,v \in \mathbb{R}^d$,

$$\Omega^{\mathsf{st}}_x \triangleq \{z \in \Omega : [x,z] \subset \Omega\}.$$

Observe that when Ω is convex, $\Omega^{\mathsf{st}}_x = \Omega$, and

$$\Omega^{\mathsf{sym}}_x = \{z \in \Omega : [2x-z, z] \subset \Omega\}.$$

A technical leverage score bound can be described with the above notations.

Theorem 4.1. *For* $x \in \Omega$*, let*

$$\tilde{\tau}_\kappa = \kappa \min(\kappa^{5d}/|\Omega^{\mathsf{st}}_x|, 1/|\Omega^{\mathsf{sym}}_x|)$$

with $\kappa > 0$ *being a parameter to be chosen later. Then there exists some absolute constant* $C > 0$ *such that* $\tau_{\mu,\Omega,\epsilon} \leq \tilde{\tau}_\kappa$ *if* $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$.

This already implies a non-trivial sample complexity bound for any convex region Ω, of which, to the best of our knowledge, no comparable result has been obtained before.

Theorem 4.2 (Sub-optimal sampling on convex regions)**.** *Assume* Ω *is convex. With* $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$, *running* LeverageScoreSamp *(Algorithm 2) with* $\tilde{\tau}_\kappa \equiv \kappa^{5d+1}/|\Omega|$ *takes* $\mathrm{poly}_d(s_{\mu,\Omega,\epsilon}) + O(1/\delta)$ *samples and guarantees* $C\epsilon$*-accurate reconstruction with probability at least* $1-\delta$.

The above theorem indicates that regardless of the form of μ, it is possible to achieve accurate reconstruction with uniformly random sampling over Ω, taking only $\mathrm{poly}_d(s_{\mu,\Omega,\epsilon})$ number of samples which can be fully controlled by the optimal sample complexity $s_{\mu,\Omega,\epsilon}$. Considering that so few have been known for the sampling complexity on general regions Ω and for general μ in multidimensions, this is already a remarkably strong result.

Algorithm 4 ParallelepipedSamp

Input: A parallelepiped Ω, number of samples k, an oracle to sample $\tilde{f}$ at any given point in Ω.

Output: Sample locations $x_1, \ldots, x_k \in \Omega$, samples $\tilde{f}_X$, weights $w_1, \ldots, w_k$.

1. Compute an invertible matrix A and a vector b such that $\Omega = A\Omega_0 + b$, where $\Omega_0 = [0,1]^d$ is the unit cube.
2. Let $\{\tilde{\tau}^0_\kappa\}_\kappa$ be the well-parametrized family on Ω_0 given by the right hand side of (8).
3. Apply AffineReduce (Algorithm 3) to Ω_0, $\{\tilde{\tau}^0_\kappa\}_\kappa$, A, b, Ω and obtain a parametrized family of leverage score bounds $\{\tilde{\tau}_\kappa\}_\kappa$ on Ω.
4. Output the result of LeverageScoreSamp (Algorithm 2) applied to $\tilde{\tau}_\kappa$, Ω, k with a specified κ.

Next we discuss whether it is possible to attain the near-optimal sample complexity $\tilde{O}_d(s_{\mu,\Omega,\epsilon})$ instead of $\mathrm{poly}_d(s_{\mu,\Omega,\epsilon})$. In Theorem 4.1 we do not provide a bound for $\tilde{s}_\kappa \triangleq \int_\Omega \tilde{\tau}_\kappa(x)dx$. In general such a bound can be useful only if $|\Omega_x^{\mathsf{st}}|$ and $|\Omega_x^{\mathsf{sym}}|$ behave well. The following lemma is a quantified version of this idea.

Lemma 4.3. *Assume $|\Omega|/|\Omega_x^{\mathsf{st}}| = O_d(1)$. Assume for some constant $C(d) > 0$ depending only on d and for any $\lambda > 2$,*

$$\frac{1}{|\Omega|}\int_\Omega \lambda \wedge \frac{|\Omega|}{|\Omega_x^{\mathsf{sym}}|}dx \le \log^{C(d)}\lambda. \tag{9}$$

Then $\tilde{s}_\kappa \triangleq \int_\Omega \tilde{\tau}_\kappa(x)dx = \tilde{O}_d(\kappa)$ with $\tilde{\tau}_\kappa$, κ defined as in Theorem 4.1.

Corollary 4.2. *With assumptions and notations as in Theorem 4.1 and Lemma 4.3, leverage score sampling with leverage score bound $\tilde{\tau}_\kappa$ (see Lemma 2.1) takes $\tilde{O}_d(\kappa + 1/\delta)$ (where $\delta > 0$ is arbitrary) samples and guarantees $C\epsilon$-accurate reconstruction with probability at least $1-\delta$.*

We discuss different settings where the assumptions of Lemma 4.3 hold and fail. Arguably Ω being convex is the most important case in practice. As we have observed, $\Omega_x^{\mathsf{st}} = \Omega$ if Ω is convex, thus in this case $|\Omega|/|\Omega_x^{\mathsf{st}}| = 1$ holds. On the other hand, the behavior of $|\Omega_x^{\mathsf{sym}}|$ is closely related to the boundary behavior of Ω. For Ω being a parallelepiped, one may easily check that (9) holds with $C(d) = d$. In fact, reducing to the case where Ω is the unit cube by affine equivariance, one may check that $1/|\Omega_x^{\mathsf{sym}}|$ is exactly equal to the function $\rho_d(x)$ defined in Theorem 3.1. With some tedious computations one may also show these assumptions are satisfied for Ω being a simplex or other polyhedron with $O_d(1)$ vertices:

Theorem 4.3 (Near-optimal sampling on a polyhedron)**.** *If Ω is a polyhedron with $O_d(1)$ vertices, with $\kappa \ge C(d)s_{\mu,\Omega,\epsilon}$, running* LeverageScoreSamp *(Algorithm 2) with leverage score bound $\tilde{\tau}_\kappa$ defined as in Theorem 4.1 takes $\tilde{O}_d(\kappa + 1/\delta)$ samples and guarantees $C\epsilon$-accurate reconstruction with probability at least $1-\delta$.*

Remark 1. From an algorithmic perspective it is reasonable to ask how $\tilde{\tau}_\kappa$ can be efficiently computed. One may indeed show that for Ω being a polyhedron with $O_d(1)$ vertices as above, it is possible to upper-bound $\tilde{\tau}_\kappa$ by another well-parametrized family $\hat{\tau}_\kappa$ which can be efficiently computed (given that Ω can be efficiently represented). This provides a method to implement the above near-optimal sampling scheme efficiently. Details can be found in the extended version of this paper.

Remark 2. Using techniques introduced later in Section 4.3, the above theorem can be generalized to the case where Ω is a simplicial complex (i.e., almost disjoint union of polyhedra) with $O_d(1)$ vertices. For ease of presentation we will not present the details here.

For many practical purposes the above theorem should be sufficiently general and tight. However, when Ω is, for example, the d-dimensional unit ball, one has $|\Omega_x^{\mathsf{sym}}| \asymp (1-|x|)^{-(d+1)/2}$, and (9) fails since the left hand side is polynomial in λ. In effect, this leads to a sub-optimal sampling strategy on such Ω which requires $\mathrm{poly}_d(s_{\mu,\Omega,\epsilon})$ samples rather than the near-optimal $\tilde{O}_d(s_{\mu,\Omega,\epsilon})$ samples.

The reason for the aforementioned non-optimality for some Ω may be due to the fact that our sampling strategy (leverage score sampling), together with the leverage score bound we proved, puts particular emphasis on the "corners" of Ω. This can be seen clearly when Ω is the unit cube: the sampling density function $\tilde{\tau}$ can be viewed as a truncation of ρ_d by some constant, and ρ_d grows fastest around the corners of the unit cube. This indicates that samples around the "corners" of Ω are crucial to the reconstruction performance. However, for Ω with smooth boundary, e.g. when Ω is the unit ball, no "corners" exist, making it difficult to design an effective sampling strategy. It is currently not clear whether this difficulty can be resolved within the framework of leverage score sampling, i.e. by improving upon Theorem 4.1. We leave this direction for future work, possibly by utilizing the techniques in the next subsection. For the rest of this paper we consider another approach, which assumes access to samples outside Ω. One may imagine that it is then possible to embed Ω into a slightly larger region with sufficiently many "corners", and achieve near-optimal sampling and reconstruction on the larger region. If the larger region is not too large, the restriction of the reconstructed function to Ω is the a good estimation of f on Ω. This is indeed the case, as we will show soon.

4.3 Near-Optimal Bound with Outside Samples

In this section we assume that (2) holds for all $x \in \mathbb{R}^d$, and we may sample $\tilde{f}$ for any $x \in \mathbb{R}^d$. Meanwhile, we still aim at reconstructing f on a given region Ω. As will be shown in our proofs, these assumptions can be significantly relaxed: we only need to sample $\tilde{f}$ on a region that is slightly larger than Ω, but for ease of presentation we choose to omit these details.

Unless specified otherwise, throughout this section we assume that Ω can be expressed by the *almost disjoint* union of convex regions $\Omega_1, \dots, \Omega_n$. By *almost disjoint* we mean $|\Omega_j \cap \Omega_l| = 0$ whenever $j \ne l$. The reason for this assumption is that we have to assume basic regularity of Ω to obtain such a strong result as a near-optimal sampling strategy on Ω, and convexity as a regularity assumption is amenable to analysis as well as being able to cover most practical purposes. Here we further allow Ω to be a finite union of convex sets, which should be sufficiently general for practical use. We first present a basic property of this model.

Lemma 4.4. *For $\Omega = \cup_{j=1}^n \Omega_j$ with $|\Omega_j \cap \Omega_l| = 0$ whenever $j \ne l$ (no convexity is assumed), setting $\epsilon_j =$*

$\epsilon|\Omega|/|\Omega_j|$, we have

$$\max_{1\le j\le n} s_{\mu,\Omega_j,\epsilon_j} \le s_{\mu,\Omega,\epsilon} \le \sum_{j=1}^{n} s_{\mu,\Omega_j,\epsilon_j}.$$

Consequently, if $O_d(1)$, we have

$$s_{\mu,\Omega,\epsilon} \asymp_d \max_{1\le j\le n} s_{\mu,\Omega_j,\epsilon_j}.$$

We will use Lemma 4.4 to reduce our general case to the simpler case where Ω is convex, by showing that it suffices to sample and reconstruct on the components Ω_j. This simpler case will then be handled by the following tools originated from convex geometry (Brazitikos et al., 2014):

Lemma 4.5. *Let Ω be a convex body. Then there exists a parallelepiped Ω' in $\mathbb{R}^d$ such that $\Omega \subset \Omega'$ and*

$$|\Omega'| \le C_d|\Omega|.$$

Moreover, Ω' can be effectively computed given knowledge of $|\Omega|$, $\int_\Omega x_j dx$, $\int_\Omega x_j x_l dx$ for $1 \le j,l \le d$ and an oracle that for any affine transform f computes $\sup\{|f(x)| : x \in \Omega\}$. The algorithm of computing Ω' is shown in Algorithm 5.

Algorithm 5 FindParallelepiped

Input: A convex region Ω, quantities $|\Omega|$, $\int_\Omega x_j dx$, $\int_\Omega x_j x_l dx$ for $1 \le j,l \le d$, an oracle that for any affine transform f computes $\sup\{|f(x)| : x \in \Omega\}$.
Output: A parallelepiped Ω' fulfilling Lemma 4.5.
1. Let $b = (b_1, \dots, b_d)$, where $b_j = \int_\Omega x_j dx/|\Omega|$.
2. Compute $A = \frac{1}{|\Omega|}\int_\Omega x_j x_l dx - b_j b_l$, which should be positive definite (Brazitikos et al., 2014).
3. Compute $B = A^{-1/2}$.
4. Using the oracle, compute $R = \sup\{|B(x-b)| : x \in \Omega\}$.
5. Let $\mathcal{B}_\infty(R) = \{x \in \mathbb{R}^d : \|x\|_\infty \le R\}$.
6. Output $\Omega' = B^{-1}\mathcal{B}_\infty(R) + b$.

For each Ω_j, Lemma 4.5 gives a parallelepiped Ω'_j that is slightly larger than Ω_j in the sense that $|\Omega'_j| \le C_d|\Omega_j|$. We may then apply the procedure ParallelepipedSamp in Algorithm 4 to each Ω'_j with $\kappa > 0$ to be chosen later and using these samples to compute a reconstruction $\hat{f}_j$ of f on Ω'_j using Reconstruct (Algorithm 1). Ignoring boundary values, letting $\hat{f} = \hat{f}_j$ on Ω_j for $j = 1, \dots, n$, we obtain an $\hat{f}$ defined on the whole Ω. The next lemma show that $\hat{f}$ is $O_d(\epsilon)$-accurate given $\kappa \ge Cs_{\mu,\Omega,\epsilon}$ and demonstrate the near-optimality of the sampling strategy.

Theorem 4.4. *The procedure described above takes $\tilde{O}_d(n(\kappa + n/\delta))$ samples and outputs a $C_d\epsilon$-accurate reconstruction $\hat{f}$ of f with probability at least if $\kappa \ge C(d)ns_{\mu,\Omega,\epsilon}$. In particular, when $n = O_d(1)$, it takes $\tilde{O}_d(\kappa + n/\delta)$ samples and outputs a $O_d(\epsilon)$-accurate reconstruction given $\kappa \ge C(d)s_{\mu,\Omega,\epsilon}$.*

Instead of gluing piecewise defined functions $\hat{f}_j$, one may simply use all samples to reconstruct f on $\cup_{j=1}^n \Omega'_j$ and then restrict to Ω. This algorithm has almost identical theoretical guarantee as that in Theorem 4.4. We omit the corresponding result here to avoid repetition.

5 CONCLUSION

In this work we present a oblivious near-optimal sampling strategy for multidimensional signals with Fourier constraints that can be expressed by a prior distribution on Fourier power spectrum. Built on a well-known randomized algorithm called leverage score sampling, we derive our sampling strategy by proving upper bounds for leverage score functions on different sampling regions. While our bounds are tight and yield near-optimal sampling strategies for many useful regions such as parallelepipeds, it fails to be optimal when the boundary of the region is smooth and has no enough “corners”. Using tools from convex geometry, we are able to resolve this challenge assuming that we may sample the signal outside the region on which we would like to reconstruct the signal. In this case, we show that it is still possible to obtain oblivious near-optimal sampling strategy under very mild assumptions on the reconstruction region. A few problems are left open. Is it possible to obtain oblivious near-optimal sampling strategy without sampling outside the reconstruction region? Is it possible to design a deterministic sampling strategy with the same performance? Is it possible to tightening the dependence on the dimension d so that the results can be used in high dimensions? We leave these problems for future work.

6 ACKNOWLEDGEMENTS

The authors are with Beijing National Research Center for Information Science and Technology (BNRist) and the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. This work was supported in part by the National Natural Science Foundation of China under Grant U2230201 and 61971266, Grant from the Guoqiang Institute, Tsinghua University, and in part by the Clinical Medicine Development Fund of Tsinghua University.

References

H. Avron, M. Kapralov, C. Musco, C. Musco, A. Velingker, and A. Zandieh. A universal sampling method for reconstructing signals with simple fourier transforms. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1051–1063, 2019.

J. Batson, D. A. Spielman, and N. Srivastava. Twice-ramanujan sparsifiers. *SIAM Journal on Computing*, 41 (6):1704–1721, 2012.

B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.

R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*, volume 196. American Mathematical Soc., 2014.

X. Chen, D. M. Kane, E. Price, and Z. Song. Fourier-sparse interpolation without a frequency gap. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 741–750. IEEE, 2016.

A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *The SIAM journal of computational mathematics*, 3:181–203, 2017.

D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

P. Drineas and M. W. Mahoney. Randnla: randomized numerical linear algebra. *Communications of the ACM*, 59 (6):80–90, 2016.

Y. C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge university press, 2012.

Y. C. Eldar and M. Unser. Nonideal sampling and interpolation from noisy observations in shift-invariant spaces. *IEEE Transactions on Signal Processing*, 54(7):2636–2651, 2006.

K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional mri time-series. *Human brain mapping*, 1(2):153–171, 1994.

Y. Jin, D. Liu, and Z. Song. A robust multi-dimensional sparse fourier transform in the continuous setting. *arXiv preprint arXiv:2005.06156*, 2020.

V. A. Kotelnikov. On the carrying capacity of the" either" and wire in telecommunications. In *Material for the First All-Union Conference on Questions of Communication (Russian), Izd. Red. Upr. Svyzai RKKA, Moscow, 1933*, 1933.

H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty—ii. *Bell System Technical Journal*, 40(1):65–84, 1961.

S. Minsker. On some extensions of bernstein's inequality for self-adjoint operators. *Statistics & Probability Letters*, 127:111–119, 2017.

M. Mishali and Y. C. Eldar. From theory to practice: Sub-nyquist sampling of sparse wideband analog signals. *IEEE Journal of selected topics in signal processing*, 4(2):375–391, 2010.

H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

A. Osipov and V. Rokhlin. On the evaluation of prolate spheroidal wave functions and associated quadrature rules. *Applied and Computational Harmonic Analysis*, 36(1):108–142, 2014.

B. Pesquet-Popescu and J. L. Véhel. Stochastic fractal models for image processing. *IEEE signal processing magazine*, 19(5):48–62, 2002.

S. Ramani, D. Van De Ville, and M. Unser. Sampling in practice: Is the best reconstruction space bandlimited? In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–153. IEEE, 2005.

S. Ramani, D. Van De Ville, and M. Unser. Non-ideal sampling and adapted reconstruction using the stochastic matern model. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II. IEEE, 2006.

C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

B. D. Ripley. *Spatial statistics*, volume 575. John Wiley & Sons, 2005.

R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2nd edition, 2013.

C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

D. Slepian. Prolate spheroidal wave functions, fourier analysis and uncertainty—iv: extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal*, 43(6):3009–3057, 1964.

D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty—i. *Bell System Technical Journal*, 40(1):43–63, 1961.

D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40 (6):1913–1926, 2011.

H. Xiao, V. Rokhlin, and N. Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse problems*, 17(4):805, 2001.

Z. Zhu and M. B. Wakin. Approximating sampled sinusoids and multiband signals using multiband modulated dpss dictionaries. *Journal of Fourier Analysis and Applications*, 23(6):1263–1310, 2017.

A EXAMPLES OF PRIOR DISTRIBUTIONS

We list some commonly-used μ in model (1) and (2). For one-dimensional signals this was investigated by Avron et al. (2019). Here we focus on the multidimensional setting.

A.1 Prior Distributions From Wave Theory

In this subsection we use a notation slightly different from the main text to conform to wave-theoretic conventions. We view a 4-dimensional signal as a signal in two variables x, t, where $x \in \mathbb{R}^3$ and $t \in \mathbb{R}$ means respectively the spatial domain and the time domain. The Fourier transform of such a signal, also defined on $\mathbb{R}^4$, is considered to be a function in two variables k, ω where $k \in \mathbb{R}^3$, $\omega \in \mathbb{R}$, corresponding respectively to the spatial domain and the time domain.

In wave theory, a plane wave can be thought of as a signal defined on $\mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}$:

$$f(x,t) = A\mathrm{e}^{i(\langle k,x\rangle - \omega t)}$$

with $|k| = \omega/c$, c being the speed of light. Obviously, such a signal corresponds to

$$\mu = \delta_k \otimes \delta_\omega$$

where δ denotes the Dirac measure.

This model can be generalized in multiple ways. For example, a single-frequency wave arrived from an unknown direction can be modelled by

$$\mu = \mathsf{Unif}\left(\frac{\omega_0}{c}\mathbb{S}^2\right) \otimes \delta_{\omega_0}$$

where ω_0 denotes the frequency of the wave, $\mathsf{Unif}\left(\frac{\omega_0}{c}\mathbb{S}^2\right)$ denotes the uniform measure on $\frac{\omega_0}{c}\mathbb{S}^2 = \{z \in \mathbb{R}^3 : |z| = \omega_0/c\}$, and δ_{ω_0} denotes the Dirac measure centered at ω_0.

In the above formulation we have assumed a flat prior on the direction of the wave. Instead, we may assume we have some knowledge on the possible directions expressed by a prior distribution μ_0 on $\frac{\omega_0}{c}\mathbb{S}^2$. In this case we can set

$$\mu = \mu_0 \otimes \delta_{\omega_0}.$$

Next we extend the above discussions to waves consisting of many frequencies. For example, we consider waves that are bandlimited, i.e. consisting of frequencies $\omega \in [\omega_0, \omega_1]$. In this case, assuming a flat prior on the directions of the wave, we have

$$\mu = \frac{1}{\omega_1 - \omega_0}\int_{\omega_0}^{\omega_1} \mathsf{Unif}\left(\frac{\omega}{c}\mathbb{S}^2\right) \otimes \delta_\omega d\omega.$$

Similarly, one may model multiband signals with direction priors, frequency-sparse signals with direction priors, etc., by generalizing the corresponding models for one-dimensional frequency domain prior in Avron et al. (2019) in the above way.

A.2 Prior Distributions From Gaussian Process Regression

Another viewpoint to the prior distribution μ is to think of μ as a prior on the power spectrum of a Gaussian process, that is, the Fourier transform of the autocorrelation of the Gaussian process, see Avron et al. (2019). Sampling and reconstruction in this setting is coined the name "Gaussian process regression", a dominantly useful technique in various fields in science and engineering. With this viewpoint many μ can be proposed for different scenarios of Gaussian process regression. In geostatistics and image processing, the following (isotropic) Matern model (Ramani et al., 2006) of the p.d.f. of μ is often used:

$$p(\xi) = \frac{\sigma_0^2}{(\alpha^2 + |\xi|^2)^{\nu + d/2}},$$

where α, ν are parameters and σ_0 are normalizing constants. Generalized versions of Matern model, e.g. anisotropic Matern model, can also be formulated in terms of a prior distribution μ, see Ramani et al. (2006). In Pesquet-Popescu and

Véhel (2002), many more complicated priors are considered for image processing (where $d = 2$), e.g. the 2D FARIMA is a prior with p.d.f.

$$p(\xi) \propto \left(\sin^2 \frac{\xi_1}{2} + \sin^2 \frac{\xi_2}{2}\right)^{-H},$$

where H is the so-called Hurst parameter. For more examples in this vein, please refer to Ripley (2005).

B AN EFFICIENT RECONSTRUCTION ALGORITHM

We briefly describes an efficient version of Reconstruct (Algorithm 1). Denote by $L^2(w)$ the (finite-dimensional) Hilbert space $\mathbb{C}^k$ equipped with the inner product

$$\langle u, v\rangle_w = \sum_{j=1}^{k} w_j \overline{u_j} v_j.$$

We may view $\mathcal{F}_X$ as an operator from $L^2(\mu)$ to $L^2(w)$ and talk about its adjoint $\mathcal{F}_X^*$. With this notation the solution of (6) can be expressed as:

$$\hat{\beta} = (\mathcal{F}_X^* \mathcal{F}_X + \epsilon\mathcal{I})^{-1} \mathcal{F}_X^* \tilde{f}_X.$$

By elementary linear algebra one has $(\mathcal{F}_X^* \mathcal{F}_X + \epsilon\mathcal{I})^{-1} \mathcal{F}_X^* = \mathcal{F}_X^* (\mathcal{F}_X \mathcal{F}_X^* + \epsilon\mathcal{I})^{-1}$ (note that the $\mathcal{I}$'s in the left hand side and in the right hand side are different: they are identities on different Hilbert spaces $L^2(\mu)$ and $L^2(w)$). The reason we make this transform is that $\mathcal{F}_X \mathcal{F}_X^*$ is simply an endomorphism on the finite-dimensional space $\mathbb{C}^k$, which can be computed directly in matrix form:

$$\mathcal{F}_X \mathcal{F}_X^* = (k_\mu(x_j, x_l) w_l)_{1\le j,l\le k}$$

where

$$k_\mu(x, x') \triangleq \int \mathrm{e}^{i\langle \xi, x - x'\rangle} d\mu(\xi).$$

Moreover, for any $u \in \mathbb{C}^k$ we have

$$(\mathcal{F}\mathcal{F}_X^* u)(x) = \sum_{j=1}^{k} w_j u_j k_\mu(x, x_j).$$

Denote $K = (k_\mu(x_j, x_l))_{1\le j,l\le k}$, $W = \mathrm{diag}(w)$, then $\mathcal{F}_X \mathcal{F}_X^* = KW$. We deduce from $\hat{f} = \mathcal{F}\hat{\beta}$ and the above equations that

$$\hat{f}(x) = \mathcal{F}\mathcal{F}_X^* (\mathcal{F}_X \mathcal{F}_X^* + \epsilon\mathcal{I})^{-1} \tilde{f}_X = \sum_{j=1}^{k} (W(KW + \epsilon\mathcal{I})^{-1} \tilde{f}_X)_j k_\mu(x, x_j)$$

which can be computed efficiently assuming access to an oracle that computes $k_m u(x, x')$ for any given $x, x' \in \mathbb{R}^d$. Also note that if $W \succ 0$, one may write $KW + \epsilon\mathcal{I} = W^{-1/2}(W^{1/2} K W^{1/2} + \epsilon\mathcal{I}) W^{1/2}$, thus

$$\hat{f}(x) = \mathcal{F}\mathcal{F}_X^* (\mathcal{F}_X \mathcal{F}_X^* + \epsilon\mathcal{I})^{-1} \tilde{f}_X = \sum_{j=1}^{k} (W^{1/2}(W^{1/2} K W^{1/2} + \epsilon\mathcal{I})^{-1} W^{1/2} \tilde{f}_X)_j k_\mu(x, x_j).$$

This last expression has the advantage that $W^{1/2} K W^{1/2} \epsilon\mathcal{I}$ is a Hermitian matrix, which may allow for faster linear solver to compute $(W^{1/2} K W^{1/2} + \epsilon\mathcal{I})^{-1} W^{1/2} \tilde{f}_X$.

We remark that for any given x the above approach computes $\hat{f}(x)$ with $O(k^2)$ calls to the oracle computing k_μ, and with $O(k^\omega)$ additions and multiplications, where $2 \le \omega < 2.373$ is the exponent for the matrix multiplication. Moreover, if we may reuse K and W, e.g. if we have computed K and W for some Ω and would like to sample and reconstruct signals on a translation of Ω, say $\Omega' = \Omega + b$ for some $b \in \mathbb{R}^d$, then (by affine equivariance) the algorithm takes $O(k)$ calls to the oracle computing k_μ and uses $O(k^2)$ additions and multiplications.

C SOME MISSING PROOFS

We present the proofs of the main results in the main text. First we prove those results which require less technical background and leave the more involved proofs of Theorem 4.1 (see Section D) and Theorem 4.3 (see Section E) for later.

C.1 Proof of Thereom 3.1

The bound $\tilde{s}_\kappa \triangleq \int_\Omega \tilde{\tau}_\kappa(x)dx = \tilde{O}_d(\kappa)$ is elementary. Define

$$K = \{x = (x_1, \dots, x_d) \in \Omega : \min(x_j, 1 - x_j) \le \kappa^{-5d},\ j = 1, \dots, d\}.$$

It is evident that $\rho_d \ge \kappa^{5d}$ on K, thus

$$\kappa^{5d} \wedge \rho_d(x) \le \begin{cases} \kappa^{5d}, & x \in K, \\ \rho_d(x), & x \in \Omega \setminus K. \end{cases} \tag{10}$$

Now we evaluate the integral $\int_K \kappa^{5d}dx$ and $\int_{\Omega\setminus K} \rho_d(x)dx$ respectively.

Since all points of K are within κ^{-d} distance to the boundary of the convex region Ω, we have $|K| \le \kappa^{-5d}|\partial\Omega|$ where $\partial\Omega$ denotes the boundary of Ω and $|\partial\Omega|$ is the surface area of $\partial\Omega$. It is well-known that $|\partial\Omega| = 2d = O_d(1)$, thus

$$\int_K \kappa^{5d}dx = \kappa^{5d}|K| \le |\partial\Omega| = O_d(1). \tag{11}$$

On the other hand, observe that

$$\Omega \setminus K = \{x = (x_1, \dots, x_d) \in \Omega : \kappa^{-5d} \le x_j \le 1 - \kappa^{-5d},\ j = 1, \dots, d\}.$$

Thus

$$\int_{\Omega\setminus K} \rho_d(x)dx = \prod_{j=1}^d \left(\int_{\kappa^{-5d}}^{1-\kappa^{-5d}} \rho(x_j)dx_j \right).$$

But by elementary integration it follows that $\int_\varepsilon^{1-\varepsilon} \rho(t)dt = 2\log(2/\varepsilon)$ for $\varepsilon \in (0, 1/2)$. This implies

$$\int_{\Omega\setminus K} \rho_d(x) = 2^d \log^d(2\kappa^{5d}) = \operatorname{polylog}_d(\kappa). \tag{12}$$

Putting together (10), (11) and (12), we obtain

$$\int_\Omega \kappa(\kappa^{5d} \wedge \rho_d(x))dx \le \kappa O_d(1) + \kappa \operatorname{polylog}_d(\kappa) = \tilde{O}_d(\kappa), \tag{13}$$

as desired.

The second part of Theorem 3.1 is a special case of Theorem 4.1, which will be proved later in Section D. In fact, assuming in Theorem 4.1 that Ω is the unit cube, it is clear that $\Omega_x^{\mathsf{st}} = \Omega$. Moreover, for a point $x = (x_1, \dots, x_d) \in \Omega$, it is clear that the region

$$\{z = (z_1, \dots, z_d) \in \Omega : |z_j - x_j| \le \min(x_j, 1 - x_j),\ j = 1, \dots, d\}$$

is a convex set (in fact, a rectangular parallelepiped) and is symmetric with respect to x. It has volume

$$\prod_{j=1}^d (2\min(x_j, 1 - x_j)) = 2^d \rho_d^{-1}(x),$$

thus $|\Omega_x^{\mathsf{sym}}| \ge 2^d \rho_d^{-1}(x)$. By Theorem 4.1 we know

$$\hat{\tau}_\kappa = \kappa(\kappa^{5d} \wedge (2^{-d}\rho_d)) = 2^{-d}\kappa((2^{1/5}\kappa)^{5d} \wedge \rho_d)$$

satisfies $\tau_{\mu,\Omega,\epsilon} \le \hat{\tau}_\kappa$ if $\kappa \ge C(d)s_{\mu,\Omega,\epsilon}$. Recalling the definition of $\tilde{\tau}_\kappa = \kappa(\kappa^{5d} \wedge \rho_d)$ in Theorem 3.1, it is clear that $\tilde{\tau}_\kappa \ge \hat{\tau}_{\kappa/2}$. Thus for $\kappa \ge 2C(d)s_{\mu,\Omega,\epsilon}$ we have $\tau_{\mu,\Omega,\epsilon} \le \tilde{\tau}_\kappa$, as desired.

C.2 Proof of Lemma 3.1 and Corollary 3.1

Lemma 3.1 is nothing more than a straightforward application of Lemma 2.1. Corollary 3.1 follows from combining Theorem 3.1 and Lemma 3.1.

C.3 Proof of Lemma 4.1

In this proof we will use notations like $\mathcal{F}_\Omega$ instead of $\mathcal{F}$ as in the main text to emphasize the dependence of related operators on the region of interest. Of course, for a different region Ω', $\mathcal{F}_{\Omega'}$ is the Fourier transform $L^2(\mu) \to L^2(\Omega')$ (note that $L^2(\Omega')$ is endowed with the normalized Lebesgue measure on Ω'). If we furthermore wish to emphasize the dependence on μ, we will use notations like $\mathcal{F}_{\mu,\Omega}$.

To prove the first equation we note that, for any $x \in \Omega$ and any $\beta \in L^2(\mu)$ we have

$$\mathcal{F}_{\mu,A\Omega}\beta(Ax) = \int \beta(\xi)\mathrm{e}^{i\langle \xi, Ax\rangle} d\mu(\xi) = \int \beta(\xi)\mathrm{e}^{i\langle A^*\xi, x\rangle} d\mu(\xi) = \int \beta(A^{*-1}\eta)\mathrm{e}^{i\langle \eta, x\rangle} d(\mu \circ A^{*-1})(\eta),$$

where the last equality follows from change-of-variable formula. For any $\beta \in L^2(\mu)$, denote

$$\tilde{\beta}(\eta) = \beta(A^{*-1}\eta), \quad \tilde{\mu} = \mu \circ A^{*-1}.$$

It is clear that $\tilde{\beta} \in L^2(\tilde{\mu})$. It is also clear that the mapping $\beta \mapsto \tilde{\beta}$ is an isometric isomorphism between $L^2(\mu)$ and $L^2(\tilde{\mu})$. We have proved

$$\mathcal{F}_{\mu,A\Omega}\beta(Ax) = \mathcal{F}_{\tilde{\mu},\Omega}\tilde{\beta}(x). \tag{14}$$

Taking squares and averaging with respect to $x \in \Omega$ yields

$$\frac{1}{|A\Omega|}\int_{A\Omega} |\mathcal{F}_{\mu,A\Omega}\beta(x')|^2 dx' = \frac{1}{|\Omega|}\int_\Omega |\mathcal{F}_{\mu,A\Omega}\beta(Ax)|^2 dx = \frac{1}{|\Omega|}\int_\Omega |\mathcal{F}_{\tilde{\mu},\Omega}\tilde{\beta}(x)|^2 dx. \tag{15}$$

By definition we know that for any $x \in \Omega$,

$$\begin{aligned}
\tau_{\mu,A\Omega,\epsilon}(Ax) &= \frac{1}{|A\Omega|}\sup_{\beta \in L^2(\mu), \beta \neq 0} \frac{|\mathcal{F}_{\mu,A\Omega}\beta(Ax)|^2}{\frac{1}{|A\Omega|}\int_{A\Omega}|\mathcal{F}_{\mu,A\Omega}\beta(x')|^2 dx' + \epsilon\|\beta\|_\mu^2} \\
&= \frac{1}{|A\Omega|}\sup_{\tilde{\beta} \in L^2(\tilde{\mu}), \beta \neq 0} \frac{|\mathcal{F}_{\tilde{\mu},\Omega}\tilde{\beta}(x)|^2}{\frac{1}{|\Omega|}\int_{\Omega}|\mathcal{F}_{\tilde{\mu},\Omega}\tilde{\beta}(x)|^2 dx + \epsilon\|\tilde{\beta}\|_\mu^2} \\
&= \frac{1}{|\det A|}\tau_{\tilde{\mu},\Omega,\epsilon}(x),
\end{aligned}$$

which proves the first equation of Lemma 4.1. The second equation of Lemma 4.1 can be proved in a similar but much easier way.

C.4 Proof of Lemma 4.2 and Corollary 4.1

Lemma 4.2 is a straight-forward consequence of Lemma 4.1. Corollary 4.1 is then obtained by combining Lemma 3.1.

C.5 Proof of Theorem 4.2

When Ω is convex, it is obvious that $\Omega_x^{\mathsf{st}} = \Omega$ for all $x \in \Omega$. We then infer from Theorem 4.1 that $\tau_{\mu,\Omega,\epsilon} \leq \kappa^{5d+1}/|\Omega|$ for $\kappa \geq C(d)s_{\mu,\Omega,\epsilon}$. The conclusion then follows from Lemma 2.1, where $\tilde{s} = \int_\Omega \kappa^{5d+1}dx/|\Omega| = \kappa^{5d+1}$.

C.6 Proof of Lemma 4.3 and Corollary 4.2

Under the assumption $|\Omega|/|\Omega_x^{\mathsf{st}}| = O_d(1)$, the $\tilde{\tau}_\kappa$ defined in Theorem 4.1 satisfies

$$\tilde{\tau}_\kappa \leq \frac{C(d)}{|\Omega|}\kappa\left(\kappa \wedge \frac{|\Omega|}{|\Omega_x^{\mathsf{sym}}|}\right).$$

It is then obvious that $\tilde{s}_\kappa = \int \tilde{\tau}_\kappa(x)dx = \tilde{O}_d(\kappa)$ under the assumptions in Lemma 4.3.

Corollary 4.2 follows from combining Theorem 4.1, Lemma 4.3 and Lemma 3.1.

C.7 Proof of Lemma 4.4

The proof will make crucial use of operator analysis, for which Bhatia (2013) is a useful reference.

As in Section C.3 we will use notations like $\mathcal{F}_\Omega$ instead of $\mathcal{F}$ as in the main text to emphasize the dependence of related operators on the region of interest. Recall that by the definition and by our choice of ϵ_j we have

$$s_{\mu,\Omega_j,\epsilon_j} = \operatorname{tr}(\mathcal{F}_{\Omega_j}\mathcal{F}_{\Omega_j}^*(\mathcal{F}_{\Omega_j}\mathcal{F}_{\Omega_j}^* + \epsilon_j\mathcal{I})^{-1}) = \operatorname{tr}(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^* + \epsilon|\Omega|\mathcal{I})^{-1}) \tag{16}$$

where $\tilde{\mathcal{F}}_{\Omega_j} = \sqrt{|\Omega_j|}\mathcal{F}_{\Omega_j}$ denotes the unnormalized Fourier transform. This new notation has the advantage that

$$(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^* f)(x) = \int d\mu(\xi)\int_{\Omega_j} f(x')\mathrm{e}^{i\langle\xi,x-x'\rangle}dx'$$

does not explicitly contain $|\Omega_j|$. In particular, one may check from this expression (and the corresponding expression for Ω) that $\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*$ is the compression[4] of $\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*$ to the subspace $L^2(\Omega_j) \subset L^2(\Omega)$. Since $\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*$ is positive, it follows from Courant-Fischer-Weyl minimax principle or Cauchy's interlacing law (Bhatia, 2013) that $\lambda_k(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*) \leq \lambda_k(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*)$, where $\lambda_k(\mathcal{A})$ denotes the k-th largest eigenvalue of a compact Hermitian operator $\mathcal{A}$. Thus

$$s_{\mu,\Omega_j,\epsilon_j} = \operatorname{tr}(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^* + \epsilon|\Omega|\mathcal{I})^{-1}) = \sum_{k=1}^{\infty}\frac{\lambda_k(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*)}{\lambda_k(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*) + \epsilon|\Omega|} \leq \sum_{k=1}^{\infty}\frac{\lambda_k(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*)}{\lambda_k(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*) + \epsilon|\Omega|} = s_{\mu,\Omega,\epsilon}$$

since the function $\lambda \mapsto \lambda/(\lambda + \epsilon|\Omega|)$ is increasing on $[0,\infty)$. This proves the first inequality in the Lemma.

To prove the last inequality in the Lemma, we further note that $L^2(\Omega_j)$ are mutually orthogonal by the assumption $|\Omega_j \cap \Omega_l| = 0$ for $j \neq l$, and together with $\Omega = \cup_j\Omega_j$ this implies

$$L^2(\Omega) = \bigoplus_{j=1}^{n} L^2(\Omega_j).$$

Making this identification, we observe that $\bigoplus_j \tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*$ can be identified as a pinching[5] of $\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*$. By a well-known majorization inequality of pinchings (Bhatia, 2013, Problem II.5.5) and the concavity of the function $\pi : \lambda \mapsto \lambda(\lambda+\epsilon|\Omega|)^{-1}$ on $[0,\infty)$, we have

$$\sum_{j=1}^{n}\sum_{k=1}^{\infty}\frac{\lambda_k(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*)}{\lambda_k(\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*) + \epsilon|\Omega|} = \operatorname{tr}\pi(\lambda(\oplus_j\tilde{\mathcal{F}}_{\Omega_j}\tilde{\mathcal{F}}_{\Omega_j}^*)) \leq \operatorname{tr}\pi(\lambda(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*)) = \sum_{k=1}^{\infty}\frac{\lambda_k(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*)}{\lambda_k(\tilde{\mathcal{F}}_\Omega\tilde{\mathcal{F}}_\Omega^*) + \epsilon|\Omega|}$$

where the notation $\lambda(\mathcal{A})$ denotes the vector formed by all the eigenvalues (with multiplicity) of a compact Hermitian operator $\mathcal{A}$, $\pi(v)$ denotes the entrywise application of π on a vector v, and $\operatorname{tr}(v)$ denotes the sum of all entries of a vector v, following the notations in Bhatia (2013). This proves $\sum_{j=1}^n s_{\mu,\Omega_j,\epsilon_j} \leq s_{\mu,\Omega,\epsilon}$, as desired.

C.8 Proof of Lemma 4.5

It is clear that $\mathcal{B}_\infty(R) \supset B(\Omega - b)$, where $B(\Omega - b) = \{B(x-b) : x \in \Omega\}$. Thus

$$\Omega' = B^{-1}\mathcal{B}_\infty(R) + b \supset B^{-1}B(\Omega - b) + b = \Omega.$$

Furthermore, the region Ω', which is an affine-transformed cube, is evidently a parallelepiped. It remains to prove the inequality $|\Omega'| \leq C(d)|\Omega|$.

We consider an auxiliary affine-transform $x \mapsto B(x-b)/|B\Omega|^{1/d}$ taking Ω to a new region $\tilde{\Omega}$ which is an isotropic convex body of volume 1. For such a convex body, it is known (Brazitikos et al., 2014) that

$$R_{\tilde{\Omega}} \overset{\triangle}{=} \sup\{|x| : x' \in \Omega\} \leq CdL_{\tilde{\Omega}},$$

[4]For an operator $\mathcal{A}$ on a Hilbert space $\mathcal{H}$ and $\mathcal{H}'$ a subspace of $\mathcal{H}$, the *compression* of $\mathcal{A}$ to $\mathcal{H}'$ is an operator on $\mathcal{H}'$ defined as $\mathcal{V}^*\mathcal{A}\mathcal{V}$, where $\mathcal{V}$ is the inclusion map $\mathcal{H}' \hookrightarrow \mathcal{H}$. This can be viewed as a generalization of the notion of principal submatrix in finite dimensions.

[5]For an operator $\mathcal{A}$ on a Hilbert space $\mathcal{H} = \bigoplus_{j=1}^n \mathcal{H}_j$, the pinching of $\mathcal{A}$ (with respect to the decomposition $\mathcal{H} = \bigoplus_{j=1}^n \mathcal{H}_j$) is an operator on $\mathcal{H}$ defined by $\sum_{j=1}^n \mathcal{P}_j\mathcal{A}\mathcal{P}_j$, where $\mathcal{P}_j$ is the orthogonal projection on $\mathcal{H}_j$.

where $L_{\tilde{\Omega}}$ is the *isotropic constant* of $\tilde{\Omega}$. It is an ubiquitous conjecture in convex geometry that $L_{\tilde{\Omega}} = O(1)$. For our purpose, it suffices to know that $L_{\tilde{\Omega}} = O_d(1)$, e.g. $L_{\tilde{\Omega}} \le Cd^{1/4}\log d$ (Brazitikos et al., 2014). As a consequence, we have $R_{\tilde{\Omega}} = O_d(1)$.

Let $\tilde{\Omega}' = R_{\tilde{\Omega}}\mathcal{B}_\infty$ where $\mathcal{B}_\infty$ denotes the ℓ_∞ unit ball. Then $\tilde{\Omega}'$ is a cube, $\tilde{\Omega} \subset \tilde{\Omega}'$, and

$$|\tilde{\Omega}'| = R_{\tilde{\Omega}}^d = O_d(1).$$

Return to the orignal algorithm, we may observe that the intermediate quantity R obtained in the algorithm FindParallelepiped is equal to $|B\Omega|^{1/d}R_{\tilde{\Omega}}$, and $\mathcal{B}_\infty(R)$ there is equal to $|B\Omega|^{1/d}\tilde{\Omega}'$. Thus $\Omega' = |B\Omega|^{1/d}B^{-1}\tilde{\Omega}' + b$. Consequently,

$$|\Omega'| = |B\Omega||B^{-1}\tilde{\Omega}'| = (\det B)(\det B^{-1})|\Omega||\tilde{\Omega}'| = |\tilde{\Omega}'||\Omega|,$$

which is no more than $C(d)|\Omega|$ by the above argument. This completes the proof.

C.9 Proof of Theorem 4.4

In the proof we will need to consider the noise $n(x) \in L^2(\Omega)$, so it is better to use an alternative notation N instead of n for the number of regions in the assumption on Ω. This means $\Omega = \cup_{j=1}^N \Omega_j$. Let $\epsilon_j = \epsilon|\Omega|/|\Omega_j|$. From Corollary 4.1 we know that, given $\kappa \ge C(d)\max_{1\le j\le N} s_{\mu,\Omega_j',\epsilon_j}$ our sampling strategy takes $\tilde{O}_d(\kappa + 1/\delta)$ samples on each Ω_j' and outputs $\hat{f}_j \in L^2(\Omega_j')$ (recall that Ω_j' is the parallelepiped obtained by FindParallelepiped (Algorithm 5) applied to Ω_j) such that

$$\frac{1}{|\Omega_j'|}\int_{\Omega_j'} |\hat{f}_j(x) - f(x)|^2 dx \le C\epsilon_j\|\alpha\|_\mu^2 + K\|n\|_{\Omega_j'}^2.$$

By Lemma 4.5 we know that $|\Omega_j'| \le C_d|\Omega|$. Combined with $\Omega_j \subset \Omega_j'$ we obtain

$$\int_{\Omega_j} |\hat{f}_j(x) - f(x)|^2 dx \le C_d\epsilon_j|\Omega_j|\|\alpha\|_\mu^2 + C_d\int_{\Omega_j'} |n(x)|^2 dx = C_d\epsilon|\Omega|\|\alpha\|_\mu^2 + C_d\int_{\Omega_j'} |n(x)|^2 dx.$$

Recall that our choice of $\hat{f}$ is obtained by gluing together all $\hat{f}_j$ on Ω_j. Thus

$$\int_\Omega |\hat{f}(x) - f(x)|^2 dx = \sum_{j=1}^n \int_{\Omega_j} |\hat{f}_j(x) - f(x)|^2 dx \le C_d N\epsilon|\Omega|\|\alpha\|_\mu^2 + C_d\sum_{j=1}^n \int_{\Omega_j'} |n(x)|^2 dx,$$

By setting $\Omega' = \cup_j \Omega_j'$, we have

$$\|\hat{f} - f\|_\Omega^2 \le C_d N\epsilon\|\alpha\|_\mu^2 + C_d N\|n\|_{\Omega'}^2,$$

which proves that the reconstruction is $O_d(N\epsilon)$-accurate [6], in particular, $O_d(\epsilon)$-accurate if $N = O_d(1)$.

It remains to bound the sample complexity. As mentioned in the beginning, we need $\tilde{O}_d(\kappa + 1/\delta)$ samples on each Ω_j', hence $\tilde{O}_d(N\kappa + N/\delta)$ samples in total. The assumption on κ is $\kappa \ge C(d)\max_j s_{\mu,\Omega_j',\epsilon_j}$. It suffices to show that this is implied by $\kappa \ge C(d)s_{\mu,\Omega,\epsilon}$, which amounts to bounding $s_{\mu,\Omega_j',\epsilon_j}$. To this end we need more tools from convex geometry.

Lemma C.1. *The region Ω_j' can be decomposed into $O_d(1)$ almost disjoint parallelepipeds $\Omega_{j,1}, \ldots, \Omega_{j,M}$, where $M = O_d(1)$, such that all of them are translations of the same parallelepiped, say Ω_j^*. Furthermore, $|\Omega_j^*| \ge c(d)|\Omega_j|$ and $\Omega_{j,1} \subset \Omega_j$.*

Proof. Recalling the proof of Lemma 4.5 (Section C.8), by an affine transform we may reduce to the case where Ω_j is centered isotropic and Ω_j' is a cube $\mathcal{B}_\infty(R)$. Let $r = \sup\{r' \ge 0 : \mathcal{B}_\infty(r') \subset \Omega_j\}$. By a lower bound on inradius (Brazitikos et al., 2014, Eqn. (3.2.1)) we have $r \ge c(d)$, hence $R/r = O_d(1)$. Now note that $\Omega_j' = [-R, R]^d$ and divide $[-R, R]$ evenly into $\lceil 2R/r\rceil$ intervals $I_1, \ldots, I_K$, $K = \lceil 2R/r\rceil = O_d(1)$. Each interval has length $2R\lceil 2R/r\rceil \le r$, thus there is some interval $I_{k_*} \subset [-r, r]$. Taking d-fold Cartesian products of this division gives a decomposition of $[-R, R]^d = \Omega_j'$ into $\lceil 2R/r\rceil^d = O_d(1)$ cubes with side length $2R/\lceil 2R/r\rceil$. We verify that this decomposition has all desired properties:

[6] Strictly speaking, we have enlarged the constant coefficient of $\|n\|_\Omega^2$ and moreover replaced $\|n\|_\Omega^2$ with $\|n\|_{\Omega'}^2$. The first point is not really important, and for the latter point we argue as following: it is reasonable to expect that the noise n distributes "randomly", so some form of ergodicity would imply the average energy $\|n\|_\Omega^2$ does not vary much for different "large" regions Ω, hence $\|n\|_\Omega^2 \asymp \|n\|_{\Omega'}^2$.

- Each of these cubes is a tranlation of the cube $\mathcal{B}_\infty(R/\lceil 2R/r\rceil)$.
- The volume of such a cube is $|\Omega_j'|/\lceil 2R/r\rceil^d \geq c(d)|\Omega_j'| \geq c(d)|\Omega_j|$.
- One of these cubes, namely $I_{k_*}^d$, is contained in $[-r,r]^d = \mathcal{B}_\infty(r) \subset \Omega$. We may index that cube by $\Omega_{j,1}$ and the desired property follows.

This completes the proof. $\square$

Return to our task of bounding $s_{\mu,\Omega_j',\epsilon_j}$. Take a decomposition $\Omega_j' = \cup_{k=1}^M \Omega_{j,k}$ as in the above lemma. Let $\lambda = |\Omega_j'|/|\Omega_{j,1}| \geq 1$. By affine equivariance (Lemma 4.3), we know $s_{\mu,\Omega_{j,k},\lambda\epsilon_j}$ does not depend on k. Thus by Lemma 4.4 we have

$$s_{\mu,\Omega_j',\epsilon_j} \leq O_d(1) s_{\mu,\Omega_{j,1},\lambda\epsilon_j}.$$

Let $\lambda' = |\Omega_j|/|\Omega_{j,1}| = O_d(1)$. Since $\Omega_{j,1} \subset \Omega_j$, by Lemma 4.4 applied to $\Omega_j = \Omega_{j,1} \cup (\Omega_j \setminus \Omega_{j,1})$ we know

$$s_{\mu,\Omega_{j,1},\lambda\epsilon_j} \leq s_{\mu,\Omega_j,\lambda\epsilon_j/\lambda'}.$$

Note that $\lambda/\lambda' \geq c_d > 0$. The above two equations together imply

$$s_{\mu,\Omega_j',\epsilon_j} \leq C(d) s_{\mu,\Omega_j,\lambda\epsilon_j/\lambda'} \leq C_d s_{\mu,\Omega_j,c_d\epsilon_j},$$

where we used $s_{\mu,\Omega_j,\epsilon}$ is, as can be clearly seen from (16), decreasing in ϵ.

Now we may apply Lemma 4.4 to obtain

$$\max_{1\leq j\leq N} s_{\mu,\Omega_j',\epsilon_j} \leq C_d \max_{1\leq j\leq N} s_{\mu,\Omega_j,c_d\epsilon_j} \leq C_d s_{\mu,\Omega_j,c_d\epsilon},\ \forall\epsilon > 0.$$

Thus $\kappa \geq C(d) s_{\mu,\Omega,c_d\epsilon}$ would be sufficient to imply $\kappa \geq C(d)\max_j s_{\mu,\Omega_j',\epsilon_j}$, and hence an accurate reconstruction.

Collecting what we have proved so far, we know that given $\kappa \geq C(d) s_{\mu,\Omega,c_d\epsilon}$, the proposed sampling and reconstruction strategy requires $\tilde{O}_d(\kappa + 1/\delta)$ samples and guarantees $O_d(N\epsilon)$-accurate reconstruction (which reduces to a $O_d(\epsilon)$-accurate reconstruction if the number of regions $N = O_d(1)$). Replacing ϵ by ϵ/c_d, we arrived at the desired conclusion.

D PROOF OF THEOREM 4.1

It suffices to prove the "uniform" bound

$$\tau_{\mu,\Omega,\epsilon}(x) \leq C^d s_{\mu,\Omega,\epsilon}^{5d+1} \cdot \frac{1}{|\Omega_x^{\mathsf{st}}|} \tag{17}$$

and the "relative" bound

$$\tau_{\mu,\Omega,\epsilon}(x) \leq C^d s_{\mu,\Omega,\epsilon}/|\Omega_x^{\mathsf{sym}}|, \tag{18}$$

which entail different approaches presented below.

D.1 Uniform Bound: Smoothness of Fourier-Sparse Functions

The overall idea is to reduce the uniform bound (17) to a smoothness estimate of Fourier-sparse functions by a sparsifier. Denote by

$$\mathscr{S}_s \triangleq \left\{ f \in \mathcal{C}^\infty(\mathbb{R}^d) : f(x) = \sum_{j=1}^s c_j \exp(i\langle \xi_j, x\rangle), c_j \in \mathbb{C}, \xi_j \in \mathbb{R}^d \right\} \tag{19}$$

the set of d-dimensional s-Fourier-sparse functions, where s is a positive integer. Using a celebrated sparsification technique (Batson et al., 2012) one may prove the following result.

Lemma D.1 (Sparsification)**.** *Assume there is some monotonically increasing function* p *such that the following bound holds for all* $s > 0$*,* $f \in \mathscr{S}_s$ *and for all region* Ω*:*

$$|f(x)|^2 \le p(s)\|f\|^2_{\Omega_x^{\mathrm{st}}},\ \forall x \in \Omega, \tag{20}$$

then for some universal constant $C > 0$ *we have, for any region* Ω*, that*

$$\tau_{\mu,\Omega,\epsilon}(x) \le C s_{\mu,\Omega,\epsilon} p(C s_{\mu,\Omega,\epsilon}) \cdot \frac{1}{|\Omega_x^{\mathrm{st}}|}.$$

The next lemma shows the assumption (20) does hold with $p(s) = C s^{2d+2} \log^{d+2} s$.

Lemma D.2 (Smoothness estimate)**.** *There exists a universal constant* $C > 0$ *such that for any* $s > 0$*,* $f \in \mathscr{S}_s$ *and for any region* Ω*, we have, for any* $x \in \Omega$*, that*

$$|f(x)|^2 \le C s^{2d+2} (\log^{d+2} s) \|f\|^2_{\Omega_x^{\mathrm{st}}}.$$

It is clear that (17) follows from combining the above two lemmas. The rest of this part is devoted to the proof of these lemmas.

D.1.1 Proof of Sparsification

The basic ingredient of the proof is the following result, essentially proved in Batson et al. (2012). A formal proof in the unidimensional setting can be found in Avron et al. (2019, Theorem 9), and the modifications required to prove the multidimensional version are mostly notational (hence omitted here).

Lemma D.3. *There exists some frequency vectors* $\xi_1, \xi_2, \ldots, \xi_K \in \mathbb{R}^d$ *with* $K = O(s_{\mu,\Omega,\epsilon})$ *such that the following holds. Let* Ξ *be the subspace of* $L^2(\Omega)$ *spanned by* $(\mathrm{e}^{i\langle \xi_j, x\rangle})_{j=1}^K$*, and* $\mathcal{P}_\Xi$ *be the orthogonal projection onto* Ξ*. Then*

$$\mathrm{tr}((\mathcal{I} - \mathcal{P}_\Xi)\mathcal{F}\mathcal{F}^*(\mathcal{I} - \mathcal{P}_\Xi)) \le 4\epsilon s_{\mu,\Omega,\epsilon}. \tag{21}$$

where the trace is well-defined since $\mathcal{F}$ *is an integral operator between finite-measure spaces with a uniformly bounded kernel, hence a Hilbert-Schmidt operator.*

Remark 3. Note that Ξ is spanned by a finite collection of smooth functions, thus any element in Ξ, being a finite sum of smooth functions, can be considered pointwise defined. Moreover, for any $f \in L^2(\Omega)$, we may say $\mathcal{P}_\Xi f \in L^2(\Omega) \cap \mathscr{S}_K$.

Remark 4. In the sense of the above remark, one may formally deduce for any $\beta \in L^2(\mu)$ that

$$\mathcal{P}_\Xi \mathcal{F}\beta = \mathcal{P}_\Xi \int \beta(\xi) \mathrm{e}^{i\langle \xi, \cdot\rangle} d\mu(\xi) = \int \beta(\xi)(\mathcal{P}_\Xi \mathrm{e}^{i\langle \xi, \cdot\rangle}) d\mu(\xi),$$

which can be verified by standard functional analysis. To clarify the notations we denote by ϕ_ξ the smooth function in $L^2(\Omega)$ given by $\phi_x i(x) = \exp(i\langle \xi, x\rangle)$. Note that ϕ_ξ is also smooth in ξ. The above identity can be written as

$$\mathcal{P}_\Xi \mathcal{F}\beta = \int \beta(\xi) \mathcal{P}_\Xi \phi_\xi d\mu(\xi).$$

By a well-known expression of Hilbert-Schmidt norm of integral operators it is easy to check the following identity:

$$\mathrm{tr}((\mathcal{I} - \mathcal{P}_\Xi)\mathcal{F}\mathcal{F}^*(\mathcal{I} - \mathcal{P}_\Xi)) = \frac{1}{|\Omega|} \int d\mu(\xi) \int_\Omega \left| \mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x) \right|^2 dx, \tag{22}$$

which is no more than $4\epsilon s_{\mu,\Omega,\epsilon}$ by Lemma D.3.

Now we decompose

$$\begin{aligned}
\tau_{\mu,\Omega,\epsilon}(x) &= \frac{1}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|(\mathcal{F}\beta)(x)|^2}{\|\mathcal{F}\beta\|^2_\Omega + \epsilon\|\beta\|^2_\mu} \\
&\le \frac{2}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2 + |((\mathcal{F} - \mathcal{P}_\Xi \mathcal{F})\beta)(x)|^2}{\|\mathcal{F}\beta\|^2_\Omega + \epsilon\|\beta\|^2_\mu} \\
&\le \frac{2}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2}{\|\mathcal{F}\beta\|^2_\Omega} + \frac{|((\mathcal{F} - \mathcal{P}_\Xi \mathcal{F})\beta)(x)|^2}{\epsilon\|\beta\|^2_\mu}
\end{aligned} \tag{23}$$

We handle the last two terms separately. For the first term, note that $\mathcal{P}_\Xi \mathcal{F}\beta \in \mathscr{S}_K$, which by the assumption of Lemma D.1 satisfies

$$\begin{aligned}
|(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2 \leq p(K)\|\mathcal{P}_\Xi \mathcal{F}\beta\|^2_{\Omega_x^{\mathsf{st}}} &= \frac{p(K)}{|\Omega_x^{\mathsf{st}}|}\int_{\Omega_x^{\mathsf{st}}} |(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2 dx \\
&\leq \frac{p(K)}{|\Omega_x^{\mathsf{st}}|}\int_{\Omega} |(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2 dx \\
&= \frac{|\Omega|}{|\Omega_x^{\mathsf{st}}|} p(K)\|\mathcal{P}_\Xi \mathcal{F}\beta\|^2_{\Omega} \\
&\leq \frac{|\Omega|}{|\Omega_x^{\mathsf{st}}|} p(K)\|\mathcal{F}\beta\|^2_{\Omega},
\end{aligned}$$

where the last inequality follows from the definition that $\mathcal{P}_\Xi$ is an orthogonal projection in $L^2(\Omega)$. This proves

$$\frac{|(\mathcal{P}_\Xi \mathcal{F}\beta)(x)|^2}{\|\mathcal{F}\beta\|^2_\Omega} \leq \frac{|\Omega|}{|\Omega_x^{\mathsf{st}}|} p(K). \tag{24}$$

We turn to the second term in (23). Recall the formulas in Remark 4, we have

$$\begin{aligned}
|((\mathcal{F} - \mathcal{P}_\Xi \mathcal{F})\beta)(x)|^2 &= \left|\int \beta(\xi)(\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)) d\mu(\xi)\right|^2 \\
&\leq \|\beta\|^2_\mu \int \left|\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)\right|^2 d\mu(\xi)
\end{aligned} \tag{25}$$

by Cauchy-Schwarz inequality. Now note that, since $\mathcal{P}_\Xi \phi_\xi \in \mathscr{S}_K$, it follows that $\mathrm{e}^{i\langle \xi, \cdot\rangle} - (P_\Xi \phi_\xi)(\cdot) \in \mathscr{S}_{K+1}$, thus by the assumption of Lemma D.1 we have

$$\left|\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)\right|^2 \leq p(K+1)\left\|\mathrm{e}^{i\langle \xi, \cdot\rangle} - (\mathcal{P}_\Xi \phi_\xi)(\cdot)\right\|^2_{\Omega_x^{\mathsf{st}}} \leq \frac{p(K+1)}{|\Omega_x^{\mathsf{st}}|}\int_\Omega \left|\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)\right|^2 dx.$$

Integrating with respect to $d\mu(\xi)$, we obtain

$$\begin{aligned}
\int \left|\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)\right|^2 d\mu(\xi) &\leq \frac{p(K+1)}{|\Omega_x^{\mathsf{st}}|}\int d\mu(\xi)\int_\Omega \left|\mathrm{e}^{i\langle \xi, x\rangle} - (\mathcal{P}_\Xi \phi_\xi)(x)\right|^2 dx \\
&\leq \frac{|\Omega|}{|\Omega_x^{\mathsf{st}}|}(4\epsilon s_{\mu,\Omega,\epsilon}) p(K+1)
\end{aligned}$$

by (22) and Lemma D.3. Plugging back into (25), it follows

$$\frac{|((\mathcal{F} - \mathcal{P}_\Xi \mathcal{F})\beta)(x)|^2}{\epsilon\|\beta\|^2_\mu} \leq 4 \cdot \frac{|\Omega|}{|\Omega_x^{\mathsf{st}}|} p(K+1) s_{\mu,\Omega,\epsilon}. \tag{26}$$

Combining (23), (24), (26), we arrive at

$$\tau_{\mu,\Omega,\epsilon} \leq C s_{\mu,\Omega,\epsilon}(p(K) + p(K+1)) \cdot \frac{1}{|\Omega_x^{\mathsf{st}}|}$$

from which the conclusion of Lemma D.1 readily follows since p is monotonically increasing and $K = O(s_{\mu,\Omega,\epsilon})$.

D.1.2 Proof of Smoothness Estimate

A key tool of the proof is the following universal self-bounding property of recursive sequences with characteristic roots of modulo 1, proved in Chen et al. (2016).

Lemma D.4 (Universal self-bounding). *There exists some universal constant $C > 0$ such that the following holds. Let $a_m = \sum_{j=1}^s \alpha_j z_j^m$ where $\alpha_j, z_j \in \mathbb{C}$, $|z_j| = 1$. Then*

$$|a_0| \leq C \sum_{m=1}^{\lfloor Cs^2 \log s\rfloor} |a_m|.$$

We are now ready to prove Lemma D.2, which can be regarded as a generalization of a result in Chen et al. (2016) from the unidimensional setting to the multidimensional setting.

Proof of Lemma D.2. Let $\mathcal{K} = (\Omega_x^{\mathsf{st}} - x)/(Cs^2 \log s)$. By Lemma D.4, for any $t \in \mathcal{K}$ we have

$$|f(x)| \le C \sum_{1 \le m \le Cs^2 \log s} |f(x+mt)|.$$

Integrating with respect to t inside $\mathcal{K}$ we obtain

$$\begin{aligned} |\mathcal{K}||f(x)| &\le C \sum_{1 \le m \le Cs^2 \log s} \int_{m\mathcal{K}} m^{-d} |f(x+t)| dt \\ &\le C \sum_{1 \le m \le Cs^2 \log s} m^{-d} \sqrt{|m\mathcal{K}|} \left(\int_{m\mathcal{K}} |f(x+t)|^2 dt \right)^{1/2} \\ &\le C \sum_{1 \le m \le Cs^2 \log s} m^{-d/2} \sqrt{|\mathcal{K}|} \sqrt{|\Omega_x^{\mathsf{st}}|} \|f\|_{\Omega_x^{\mathsf{st}}}, \end{aligned}$$

where in the last inequality we used the fact that $x + m\mathcal{K} \subset \Omega_x^{\mathsf{st}}$ by definition of Ω_x^{st} and $\mathcal{K}$. Since $|\mathcal{K}| = |\Omega_x^{\mathsf{st}}|/(Cs^2 \log s)^d$, it follows that

$$|f(x)| \le (Cs^2 \log s)^{d/2} \|f\|_{\Omega_x^{\mathsf{st}}} \sum_{1 \le m \le Cs^2 \log s} m^{-d/2}$$

The last factor is less than $Cs\sqrt{\log s}$ when $d = 1$ and is less than $C \log(s^2 \log s) \le C \log s$ when $d = 2$. When $d > 2$, it is $O(1)$. Anyway, we arrive at the desired conclusion. $\square$

D.2 Relative Bound: Comparison Principles

In proving the relative bound we will need to compare the leverage score functions on different regions. To this end we use notations like $\mathcal{F}_\Omega$ instead of $\mathcal{F}$ as in the main text to emphasize the dependence of related operators on the region of interest. Of course, for a different region Ω', $\mathcal{F}_{\Omega'}$ is the Fourier transform $L^2(\mu) \to L^2(\Omega')$, (note that $L^2(\Omega')$ is endowed with the normalized Lebesgue measure on Ω'). The relative bound is based on the following dual representation of leverage score function.

Lemma D.5. *The leverage score function $\tau_{\mu,\Omega,\epsilon}$ can be alternatively expressed as*

$$\tau_{\mu,\Omega,\epsilon}(x) = \frac{1}{\epsilon|\Omega|} \inf_{g \in L^2(\Omega)} \|\mathcal{F}_\Omega^* g - \varphi_x\|_\mu^2 + \epsilon \|g\|_\Omega^2, \tag{27}$$

where $\varphi_x \triangleq \exp(-i\langle \cdot, x\rangle)$ is regarded as an element in $L^2(\mu)$.

Proof. The ridge regression problem (27) can be solved explicitly. The optimal solution is $g = (\mathcal{F}_\Omega \mathcal{F}_\Omega^* + \epsilon \mathcal{I})^{-1} \mathcal{F}_\Omega \varphi_x$. This leads to a calculation that shows the right hand side is equal to $|\Omega|^{-1} \langle \varphi_x, (\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon)^{-1} \varphi_x \rangle$. On the other hand, in the original definition one may show

$$\begin{aligned} \frac{1}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|(\mathcal{F}_\Omega \beta)(x)|^2}{\|\mathcal{F}_\Omega \beta\|_\Omega^2 + \epsilon \|\beta\|^2} &= \frac{1}{|\Omega|} \sup_{\beta \in L^2(\mu), \beta \ne 0} \frac{|\langle \varphi_x, \beta \rangle|^2}{\|(\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon \mathcal{I})^{1/2} \beta\|_\mu^2} \\ &= \frac{1}{|\Omega|} \sup_{\beta' \in L^2(\mu), \beta' \ne 0} \frac{|\langle (\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon \mathcal{I})^{-1/2} \varphi_x, \beta' \rangle|^2}{\|\beta'\|_\mu^2} \end{aligned}$$

where we substitute $\beta' = (\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon)^{-1/2} \beta$. The last quantity is a standard eigenvalue problem for a rank-one operator, which can be shown to be $|\Omega|^{-1} \|(\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon)^{-1/2} \varphi_x\|_\mu^2 = |\Omega|^{-1} \langle \varphi_x, (\mathcal{F}_\Omega^* \mathcal{F}_\Omega + \epsilon)^{-1} \varphi_x \rangle$, exactly equal to the value of the right hand side of (27) as computed above. $\square$

Let $\Omega' = \frac{1}{2}x + \frac{1}{2}\Omega_x^{\mathsf{sym}} = x + \frac{1}{2}(\Omega_x^{\mathsf{sym}} - x)$. Let $\epsilon' = \epsilon|\Omega|/|\Omega'|$. Since $x \in \Omega_x^{\mathsf{sym}}$, it follows from the definition of Ω_x^{sym} (which is required to be a convex body symmetric with respect to x) we have $\Omega' \subset \Omega_x^{\mathsf{sym}} \subset \Omega$.

To prove the relative bound, it turns out useful to consider an intermediate quantity $s_{\mu,\Omega',\epsilon'}$. By Lemma 4.4, we know $s_{\mu,\Omega',\epsilon'} \le s_{\mu,\Omega,\epsilon}$. On the other hand, by definition we have $s_{\mu,\Omega',\epsilon'} = \int_{\Omega'} \tau_{\mu,\Omega',\epsilon'} dx$. This (together with the continuity of leverage score function, which can be inferred from the expression given in the proof of Lemma D.5) implies the existence of some $x_\star \in \Omega'$ such that

$$\tau_{s,\Omega',\epsilon'}(x_\star) \le \frac{1}{|\Omega'|} s_{\mu,\Omega,\epsilon}.$$

Using Lemma D.5, this in turn implies the existence of some $g_\star \in L^2(\Omega')$ such that

$$\|\mathcal{F}^*_{\Omega'} g_\star - \varphi_{x_\star}\|^2_\mu + \epsilon' \|g_\star\|^2_{\Omega'} \le \epsilon' s_{\mu,\Omega,\epsilon}.$$

Now define a function $g \in L^2(\Omega)$ in the following way. If $u \in (x - x_\star) + \Omega'$, then $g(u) = \frac{|\Omega|}{|\Omega'|} g_\star(u + x_\star - x)$, which is well-defined since $u + x_\star - x \in \Omega'$. Otherwise, let $g(u) = 0$. Now, since Ω^{sym}_x is convex and symmetric with respect to x, it follows that $\Omega' = x + \frac{1}{2}(\Omega^{\mathsf{sym}}_x - x)$ is also convex and symmetric with respect to x. Thus we have $x - x_\star \in x - \Omega' = \Omega' - x$, and

$$(x - x_\star) + \Omega' \subset \Omega' - x + \Omega' = 2\Omega' - x = \Omega^{\mathsf{sym}}_x \subset \Omega.$$

This yields

$$\begin{aligned}
(\mathcal{F}^*_\Omega g)(\xi) = \frac{1}{|\Omega|} \int_\Omega g(u) \mathrm{e}^{-i\langle \xi, u\rangle} du &= \frac{1}{|\Omega|} \int_{(x - x_\star) + \Omega'} g(u) \mathrm{e}^{-i\langle \xi, u\rangle} du \\
&= \frac{1}{|\Omega|} \int_{\Omega'} \frac{|\Omega|}{|\Omega'|} g_\star(u') \mathrm{e}^{-i\langle \xi, u' + (x - x_\star)\rangle} du' \\
&= (\mathcal{F}^*_{\Omega'} g_\star)(\xi) \mathrm{e}^{-i\langle \xi, x - x_\star\rangle},
\end{aligned}$$

It is also clear from similar computations that $\|g\|^2_\Omega = \frac{|\Omega|}{|\Omega'|} \|g_\star\|^2_\Omega$, hence

$$\begin{aligned}
\|\mathcal{F}^*_\Omega g - \varphi_x\|^2_\mu + \epsilon \|g\|^2_\Omega &= \int \left| (\mathcal{F}^*_{\Omega'} g_\star)(\xi) \mathrm{e}^{-i\langle \xi, x - x_\star\rangle} - \mathrm{e}^{-i\langle \xi, x\rangle} \right|^2 d\mu(\xi) + \epsilon \frac{|\Omega|}{|\Omega'|} \|g_\star\|^2_\Omega \\
&= \int \left| (\mathcal{F}^*_{\Omega'} g_\star)(\xi) - \mathrm{e}^{-i\langle \xi, x_\star\rangle} \right|^2 d\mu(\xi) + \epsilon' \|g_\star\|^2_\Omega \\
&= \|\mathcal{F}^*_{\Omega'} g_\star - \varphi_{x_\star}\|^2_\mu + \epsilon' \|g_\star\|^2_\Omega \\
&\le \epsilon' s_{\mu,\Omega,\epsilon}.
\end{aligned}$$

Again, by Lemma D.5 applied to region Ω, this implies

$$\tau_{\mu,\Omega,\epsilon}(x) \le \frac{1}{\epsilon|\Omega|} (\|\mathcal{F}^*_\Omega g - \varphi_x\|^2_\mu + \epsilon \|g\|^2_\Omega) \le \frac{1}{\epsilon|\Omega|} \epsilon' s_{\mu,\Omega,\epsilon} = \frac{1}{|\Omega'|} s_{\mu,\Omega,\epsilon}.$$

The desired bound (18) now follows from the obvious fact that $|\Omega'| = 2^{-d} |\Omega^{\mathsf{sym}}_x|$ since Ω' is a translation of $\frac{1}{2}\Omega^{\mathsf{sym}}_x$.

E PROOF OF THEOREM 4.3

We begin with the simplest case where Ω is a polyhedron with $d+1$ vertices in $\mathbb{R}^d$ where it is easier to see the heart of the matter. Later we will present a trick to reduce the general case to this simplest case using basic convex geometry.

E.1 A Toy Case: Polyhedra with Least Vertices

Assume Ω is a polyhedron with $d+1$ vertices, namely the convex hull of points $\{v_1, \ldots, v_{d+1}\}$ where the vertices $\{v_j\}_{j=1}^{d+1}$ are in generic position. In this case $\Omega^{\mathsf{st}}_x = \Omega$, and we only need to evaluate the integral (by Theorem 4.1)

$$\int_\Omega \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega^{\mathsf{sym}}_x|} \right) dx = \frac{\kappa}{|\Omega|} \int_\Omega \kappa^{5d} \wedge \frac{|\Omega|}{|\Omega^{\mathsf{sym}}_x|} dx. \tag{28}$$

With Theorem 4.1 in mind, we need to show that the above integral is $\tilde{O}_d(\kappa)$ (recalling Lemma 3.1). Our first observation is that it suffices to consider a fixed choice of Ω, since all polyhedra with $d+1$ vertices are affine-isomorphic, and since the related quantities such as $|\Omega^{\mathsf{sym}}_x|$ are equivariant under affine transforms.

Lemma E.1 (Affine equivariance)**.** *Let A be an invertible linear operator on $\mathbb{R}^d$ and $b \in \mathbb{R}^d$. For any region $\Omega \subset \mathbb{R}^d$, we have*

$$(A\Omega + b)^{\mathsf{st}}_{Ax+b} = A\Omega^{\mathsf{st}}_x + b, \quad (A\Omega + b)^{\mathsf{sym}}_{Ax+b} = A\Omega^{\mathsf{sym}}_x + b.$$

Consequently, if Ω is convex, we have

$$\frac{\kappa}{|\Omega|} \int_\Omega \kappa^{5d} \wedge \frac{|\Omega|}{|\Omega^{\mathsf{sym}}_x|} dx = \frac{\kappa}{|A\Omega + b|} \int_{A\Omega+b} \kappa^{5d} \wedge \frac{|A\Omega + b|}{|(A\Omega + b)^{\mathsf{sym}}_x|} dx.$$

Proof. The first two equations can be checked by definitions. The last equation then follows from the change-of-variable formula. $\square$

Denote the canonical basis of $\mathbb{R}^d$ by $e_1, \ldots, e_d$. Note that e_j is simply the vector with all entries zero but the j-th entry 1. By Lemma E.1, it is without loss of generality to assume

$$v_j = \begin{cases} e_j, & 1 \le j \le d, \\ 0, & j = d+1. \end{cases} \tag{29}$$

It will be helpful to keep in mind that with this choice of v_j we have $|\Omega| = 1/d! \ge c(d)$, where $c(d) > 0$ means a constant depending only on d.

Since Ω is the convex hull of $\{v_1, \ldots, v_{d+1}\}$, for any $x \in \Omega$ we have the following representation

$$x = \sum_{j=1}^{d+1} \theta_j v_j,\ \theta_j \ge 0, \sum_{j=1}^{d+1} \theta_j = 1. \tag{30}$$

With this representation we will show that

Lemma E.2. *There exists a constant $c(d) > 0$ depending only on d such that*

$$|\Omega^{\mathsf{sym}}_x| \ge c(d) \prod_{j=1}^{d+1} \theta_j.$$

Proof. Since $\sum_j \theta_j = 1$, at most one of θ_j can exceed $1/2$. Let $j_* = \arg\max_j \theta_j$, then $\theta_j \le 1/2$ for all $j \ne j_*$. Now consider the set

$$K_x \triangleq \left\{ \sum_{j=1}^{d+1} \theta'_j v_j : 0 \le \theta'_j \le 2\theta_j; (\theta'_1, \ldots, \theta'_{d+1}) \in \Delta^d \right\}.$$

It is clear that K_x is convex and symmetric with respect to x, hence $|\Omega^{\mathsf{sym}}_x| \ge |K_x|$. To estimate $|K_x|$ we divide into two cases.

Case 1: $\theta_{j_*} \ge 1/2$**.** In this case the constraint $0 \le \theta'_{j_*} \le 2\theta_{j_*}$ in the definition of K_x is trivially true given that $(\theta'_1, \ldots, \theta'_{d+1}) \in \Delta^d$. Since $\theta'_{j_*} = 1 - \sum_{j \ne j_*} \theta'_j$, one may see

$$K_x = \left\{ v_{j_*} + \sum_{j \ne j_*} \theta'_j (v_j - v_{j_*}) : 0 \le \theta'_j \le 2\theta_j \right\},$$

which is a parallelepiped with sides parallel to $v_j - v_{j_*}$, $j \ne j_*$. By elementary geometry (or a calculation via determinant), such a parallelepiped would have volume $d!|\Omega| = 1$ if all its sides have length 1. Thus

$$|K_x| = 2^d d! |\Omega| \prod_{j \ne j_*} \theta_j \ge 2^d \prod_{j=1}^{d+1} \theta_j,$$

where the last inequality follows from $\theta_{j_*} \le 1$.

Case 2: $\theta_{j_*} < 1/2$. Using $\theta'_{j_*} = 1 - \sum_{j \neq j_*} \theta'_j$ again, we have

$$K_x = \left\{ v_{j_*} + \sum_{j \neq j_*} \theta'_j (v_j - v_{j_*}) : 0 \leq \theta'_j \leq 2\theta_j, 1 - 2\theta_{j_*} \leq \sum_{j \neq j_*} \theta'_j \leq 1 \right\}.$$

Using further $\theta_{j_*} = 1 - \sum_{j \neq j_*}$ and the assumption $\theta_{j_*} < 1/2$, one may check that K_x contains a parallelepiped:

$$K_x \supset \left\{ v_{j_*} + \sum_{j \neq j_*} \theta'_j (v_j - v_{j_*}) : \frac{1 - 2\theta_{j_*}}{1 - \theta_{j_*}} \theta_j \leq \theta'_j \leq \frac{1}{1 - \theta_{j_*}} \theta_j, \forall j \neq j_* \right\}$$

This parallelepiped again has sides parallel to $v_j - v_{j_*}$ and has side lengths $(2\theta_{j_*})\theta_j/(1 - \theta_{j_*})$, thus has volume

$$\left(\frac{2\theta_{j_*}}{1 - \theta_{j_*}} \right)^d d! |\Omega| \prod_{j \neq j_*} \theta_j.$$

Note that by the definition of j_* we have $\theta_{j_*} \geq \sum_j \theta_j/(d+1) = 1/(d+1)$, thus $\theta_{j_*}^{d-1} \geq c(d)$. Combining these arguments we obtain

$$|K_x| \geq \left(\frac{2\theta_{j_*}}{1 - \theta_{j_*}} \right)^d d! |\Omega| \prod_{j \neq j_*} \theta_j \geq 2^d d! |\Omega| \theta_{j_*}^d \prod_{j \neq j_*} \theta_j \geq c(d) \prod_{j=1}^{d+1} \theta_j,$$

as claimed. □

Return to our proof of the toy case. In light of Lemma E.2 we have

$$\kappa \min \left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|} \right) \leq c(d)^{-1} \kappa \min \left(\kappa^{5d}, \frac{1}{\prod_{j=1}^{d+1} \theta_j} \right), \tag{31}$$

where we also used $|\Omega| = 1/d! \geq c(d)$. The main goal of bounding (28) is thus reduced to bounding

$$I = \int_\Omega \kappa^{5d} \wedge \frac{1}{\prod_{j=1}^{d+1} \theta_j} dx, \tag{32}$$

where $\theta_j = \theta_j(x)$ should be regarded as functions in x. With v_j as in (29), we may explicitly compute for $x = (x_1, \ldots, x_d) \in \Omega$ that

$$\theta_j(x) = x_j, j \leq d; \quad \theta_{d+1}(x) = 1 - \sum_{j=1}^d x_j. \tag{33}$$

The evaluation of I should be elementary, but the fastest way utilizes idea from integration theory on the manifold $\mathbb{S}^d$. We cover Ω by $d+1$ "charts":

$$\Omega_j \triangleq \{x \in \Omega : \theta_j(x) = \max_{j'} \theta_{j'}(x)\}. \tag{34}$$

The constraint $\theta_j(x) = \max_{j'} \theta_{j'}(x)$ can be expressed as a set of affine-linear inequalities, thus Ω_j is again a polyhedron. Note that on Ω_j we have $\theta_j \geq \sum_{j'} \theta_{j'}/(d+1) = 1/(d+1)$, thus

$$\int_{\Omega_j} \kappa^{5d} \wedge \frac{1}{\prod_{j'=1}^{d+1} \theta_{j'}} dx \leq C(d) \int_{\Omega_j} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} dx.$$

It is also clear that $\Omega = \cup_j \Omega_j$, thus

$$I \leq C(d) \sum_{j=1}^{d+1} \int_{\Omega_j} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} dx. \tag{35}$$

On Ω_j we use the "local coordinates" $(\theta_1, \ldots, \theta_{j-1}, \theta_j, \ldots, \theta_{d+1})$. This is possible since by (30) every $x \in \Omega$ can be expressed as

$$x = v_j + \sum_{j' \neq j} \theta_{j'}(v_{j'} - v_j), \theta_{j'} \geq 0, \sum_{j' \neq j} \theta_{j'} \leq 1. \tag{36}$$

By the same argument as in the proof of Lemma E.2, we know that the affine-linear coordinate transform

$$\psi_j : (\theta_1, \ldots, \theta_{j-1}, \theta_j, \ldots, \theta_{d+1}) \mapsto x = v_j + \sum_{j' \neq j} \theta_{j'}(v_{j'} - v_j)$$

has Jacobian equal to $d!|\Omega| = 1$, thus the change-of-variable formula reads

$$\int_{\Omega_j} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} dx = \int_{\psi_j^{-1}(\Omega_j)} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} \prod_{j' \neq j} d\theta_{j'}.$$

It follows trivially from (36) that $\psi_j^{-1}(\Omega) \subset [0,1]^d$, hence $\psi_j^{-1}(\Omega_j) \subset [0,1]^d$. Thus we have

$$\int_{\psi_j^{-1}(\Omega_j)} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} \prod_{j' \neq j} d\theta_{j'} \leq \int_{[0,1]^d} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} \prod_{j' \neq j} d\theta_{j'} = \int_{[0,1]^d} \kappa^{5d} \wedge \frac{1}{\prod_{l=1}^d x'_l} \prod_{l=1}^d x'_l,$$

where we have changed the label of coordinates to reflect the fact that the middle term is actually independent of j. For simplicity denote $x' = (x'_1, \ldots, x'_d) \in [0,1]^d$. A good news is that a larger integral has already been estimated in the proof of Theorem 3.1. In fact, recalling the definition of the function $\rho(x) = 1/(x \wedge (1-x))$ and $\rho_d(x') = \prod_l \rho(x'_l)$ in Theorem 3.1, it is clear that $1/x'_l \leq \rho(x'_l)$, thus $1/\prod_l x'_l \leq \rho_d(x')$, and

$$\int_{[0,1]^d} \kappa^{5d} \wedge \frac{1}{\prod_{l=1}^d x'_l} \prod_{l=1}^d x'_l \leq \int_{[0,1]^d} \kappa^{5d} \wedge \rho_d(x') dx'.$$

By (11) and (12), we know

$$\int_{[0,1]^d} \kappa^{5d} \wedge \rho_d(x') dx' \leq O_d(1) + \mathrm{polylog}_d(\kappa) \leq \mathrm{polylog}_d(\kappa).$$

The above equations together imply

$$\int_{\Omega_j} \kappa^{5d} \wedge \frac{1}{\prod_{j' \neq j} \theta_{j'}} dx \leq \mathrm{polylog}_d(\kappa),$$

which can be plugged into (35) to deduce

$$I \leq \mathrm{polylog}_d(\kappa).$$

By (31), we finally obtain

$$\int_\Omega \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|}\right) dx \leq C(d)\kappa\, \mathrm{polylog}_d(\kappa) = \tilde{O}_d(\kappa),$$

as desired.

E.2 The General Case

Now we consider the general case where Ω is a polyhedron with $O_d(1)$ vertices. Denote by V its vertex set and by $V^{[d+1]}$ the collection of all subsets of V containing exactly $(d+1)$ elements. In this case, we know that $|V^{[d+1]}| = O_d(1)$ (note that here $|V^{[d+1]}|$ denotes the cardinality of $V^{[d+1]}$ instead of the volume).

For any $V' \in V^{[d+1]}$, the convex hull $P_{V'} = \mathrm{conv}(V')$ is a possibly-degenerate polyhedron with $d+1$ vertices. If it is non-degenerate, it would be affine-isomorphic to the polyhedron already investigated in Section E.1, thus

$$\int_{P_{V'}} \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|}\right) dx = \tilde{O}_d(\kappa). \tag{37}$$

On the other hand, if $P_{V'}$ is degenerate, it simply has volume 0, so

$$\int_{P_{V'}} \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|}\right) dx = 0. \tag{38}$$

By Carathéodory's theorem (Schneider, 2013) we have $\Omega = \cup_{V' \in V^{[d+1]}} P_{V'}$. Hence

$$\int_{\Omega} \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|}\right) dx \leq \sum_{V' \in V^{[d+1]}} \int_{P_{V'}} \kappa \min\left(\frac{\kappa^{5d}}{|\Omega|}, \frac{1}{|\Omega_x^{\mathsf{sym}}|}\right) dx \leq |V^{[d+1]}| \tilde{O}_d(\kappa) = O_d(1)\tilde{O}_d(\kappa) = \tilde{O}_d(\kappa),$$

as desired.

E.3 Remarks on Computational Issues

In the proof of the simplest case in Section E.1 we actually constructed an efficiently computable and near-optimal upper bound of $\tilde{\tau}_\kappa$ given by (31), where θ_j can be computed via (33).

For the general case, by inspecting the proof we may conclude that if we can compute a Carathéodory representation efficiently for any $x \in \Omega$, i.e. finding $V' \in V^{[d+1]}$ such that $x \in \mathrm{conv}(V')$, then we may use Lemma E.1 to transform the efficiently computable bound in the simplest case to the general case. Obviously, if the polyhedron is efficiently represented by its vertex set with size $O_d(1)$, a Carathéodory representation can be found by an exhaustive enumeration in $O_d(1)$ time. This shows that the sampling scheme indicated by Theorem 4.3 is indeed efficiently implementable.