A Universal Sampling Method for Reconstructing Signals with Simple Fourier Transforms

Haim Avron
Tel Aviv University
Israel
haimav@post.tau.ac.il

Michael Kapralov
EPFL
Switzerland
michael.kapralov@epfl.ch

Cameron Musco
Microsoft Research
USA
camusco@microsoft.com

Christopher Musco
Princeton University
USA
cmusco@cs.princeton.edu

Ameya Velingker
Google Research
USA
ameyav@google.com

Amir Zandieh
EPFL
Switzerland
amir.zandieh@epfl.ch

ABSTRACT

Reconstructing continuous signals based on a small number of discrete samples is a fundamental problem across science and engineering. We are often interested in signals with "simple" Fourier structure – e.g., those involving frequencies within a bounded range, a small number of frequencies, or a few blocks of frequencies – i.e., bandlimited, sparse, and multiband signals, respectively. More broadly, any prior knowledge on a signal's Fourier power spectrum can constrain its complexity. Intuitively, signals with more highly constrained Fourier structure require fewer samples to reconstruct.

We formalize this intuition by showing that, roughly, a continuous signal from a given class can be approximately reconstructed using a number of samples proportional to the *statistical dimension* of the allowed power spectrum of that class. We prove that, in nearly all settings, this natural measure tightly characterizes the sample complexity of signal reconstruction.

Surprisingly, we also show that, up to log factors, a universal nonuniform sampling strategy can achieve this optimal complexity for *any class of signals*. We present an efficient and general algorithm for recovering a signal from the samples taken. For bandlimited and sparse signals, our method matches the state-of-the-art, while providing the the first computationally and sample efficient solution to a broader range of problems, including multiband signal reconstruction and Gaussian process regression tasks in one dimension.

Our work is based on a novel connection between randomized linear algebra and the problem of reconstructing signals with constrained Fourier structure. We extend tools based on statistical leverage score sampling and column-based matrix reconstruction to the approximation of continuous linear operators that arise in the signal reconstruction problem. We believe these extensions are of independent interest and serve as a foundation for tackling a broad range of continuous time problems using randomized methods.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
STOC '19, June 23–26, 2019, Phoenix, AZ, USA
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6705-9/19/06…$15.00
https://doi.org/10.1145/3313276.3316363

CCS CONCEPTS

• **Theory of computation** → *Numeric approximation algorithms*;
• **Computing methodologies** → *Linear algebra algorithms.*

KEYWORDS

Signal reconstruction, Leverage score sampling, Numerical linear algebra

ACM Reference Format:
Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. 2019. A Universal Sampling Method for Reconstructing Signals with Simple Fourier Transforms. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19), June 23–26, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3313276.3316363

1 INTRODUCTION

Consider the following fundamental function fitting problem, pictured in Figure 1. We can access a continuous signal $y(t)$ at any time $t \in [0, T]$. We wish to select a finite set of sample times $t_1, \ldots, t_q$ such that, by observing the signal values $y(t_1), \ldots, y(t_q)$ at those samples, we are able to find a good approximation $\tilde{y}$ to y over the entire range $[0, T]$. We also study the problem in a noisy setting, where for each sample t_i, we only observe $y(t_i) + n(t_i)$ for some fixed noise function n. We seek to understand:

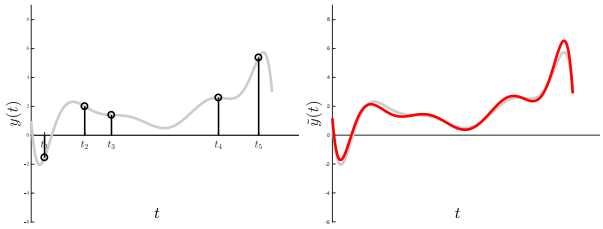
(1) How many samples q are required to approximately reconstruct y and how should we select these samples?
(2) After sampling, how can we find and represent $\tilde{y}$ in a computationally efficient way?

Answering these questions requires assumptions on the underlying signal y. In particular, for the information at our samples $t_1, \ldots, t_q$ to be useful in reconstructing y on the entirety of $[0, T]$, the signal must be smooth, structured, or otherwise "simple" in some way.

Across science and engineering, by far one of the most common ways in which structure arises is through various assumptions about $\hat{y}$, the *Fourier transform* of y:

$$\hat{y}(\xi) = \int_{-\infty}^{\infty} y(t) e^{-2\pi i t \xi} \, dt.$$

Our goal is to understand signal reconstruction under natural constraints on the complexity of $\hat{y}$.



(a) Observed signal y sampled at times $t_1, \ldots, t_q$. **(b) Reconstructed signal $\tilde{y}$ based on samples.**

Figure 1: Our basic function fitting problem requires reconstructing a continuous signal based on a small number of (possibly noisy) discrete samples.

1.1 Classical Theory for Bandlimited Signals

Classically, the most standard example of such a constraint is requiring y to be *bandlimited*, meaning that $\hat{y}$ is only non-zero for frequencies ξ with $|\xi| \leq F$ for some bandlimit F. In this case, we recall the famous sampling theory of Nyquist, Shannon, and others [32, 43, 54, 60]. This theory shows that y can be reconstructed exactly using sinc interpolation (i.e, Whittaker-Shannon interpolation) if $1/2F$ uniformly spaced samples of y are taken per unit of time (the 'Nyquist rate').

Unfortunately, this theory is asymptotic: it requires infinite samples over the entire real line to interpolate y, even at a single point. When a finite number of samples are taken over an interval $[0, T]$, sinc interpolation is not a good reconstruction method, either in theory or in practice [63].[1]

This well-known issue was resolved through a seminal line of work by Slepian, Landau, and Pollak [34, 35, 57], who presented a set of explicit basis functions for interpolating bandlimited functions when a finite number of samples are taken from a finite interval. Their so-called "prolate spheroidal wave functions" can be combined with numerical quadrature methods [31, 56, 64] to obtain sample efficient (and computationally efficient) methods for bandlimited reconstruction. Overall, this work shows that roughly $O(FT + \log(1/\epsilon))$ samples from $[0, T]$ are required to interpolate a signal with bandlimit F to accuracy ϵ on that same interval.[2]

1.2 More General Fourier Structure

While the aforementioned line of work is beautiful and powerful, in today's world, we are interested in far more general constraints than bandlimits. For example, there is wide-spread interest in *Fourier-sparse* signals [20], where $\hat{y}$ is only non-zero for a small number of frequencies, and *multiband* signals, where the Fourier transform is confined to a small number of intervals. Methods for recovering signals in these classes have countless applications in communication, imaging, statistics, and a variety of other disciplines [22].

More generally, in statistical signal processing, a *prior distribution*, specified by some probability measure μ, is often assumed on the frequency content of y [23, 48]. For signals with bandlimit F, μ would be the uniform probability measure on $[-F, F]$. Alternatively, instead of assuming a hard bandlimit, a zero-centered Gaussian prior on $\hat{y}$ can encode knowledge that higher frequencies are less likely to contribute significantly to y, although they may still be present. Such a prior naturally suits a Bayesian approach to signal reconstruction [26] and, in fact, is essentially equivalent to assuming y is a stationary stochastic process with a certain covariance function (see Section 3). Under various names, including "Gaussian process regression" and "kriging," likelihood estimation under a covariance prior is the dominant statistical approach to fitting continuous signals in many scientific disciplines, from geostatistics to economics to medical imaging [50, 52].

1.3 Our Contributions

Despite their clear importance, accurate methods for fitting continuous signals under most common Fourier transform priors are not well understood, even 50 years after the groundbreaking work of Slepian, Landau, and Pollak on the bandlimited problem. The only exception is Fourier sparse signals: the *noiseless* interpolation problem can be solved using classical methods [10, 18, 46], and recent work has resolved the much more difficult noisy case [12, 13].

In this paper, we address the problem far more generally. Our contributions are as follows:

Theorem 1: Up to constant factors, we characterize the information theoretic sample complexity of reconstructing y under any Fourier transform prior, specified by probability measure μ. In essentially all settings, we show this complexity scales near linearly with a natural *statistical dimension* parameter associated with μ.

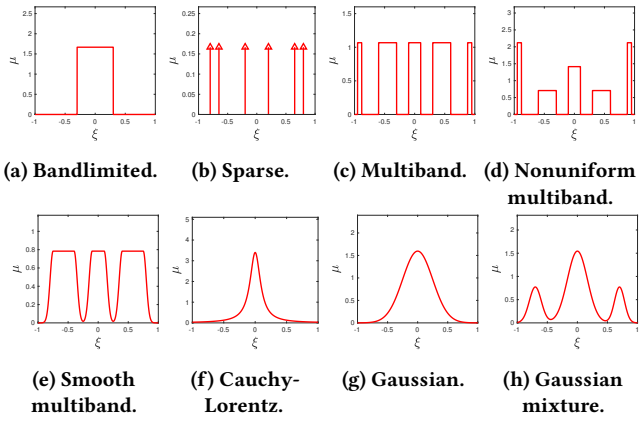
Theorem 2: We present a method for sampling from y that achieves the aforementioned statistical dimension bound to within a polylogarithmic factor. Our approach is randomized and *universal*: we prove that it is possible to draw $t_1, \ldots, t_q$ from a fixed non-uniform distribution over $[0, T]$ that is *independent of* μ, i.e., "spectrum-blind." In other words, the same sampling scheme works for bandlimited, sparse, or more general priors.

Theorem 3: We show that y can be recovered from $t_1, \ldots, t_q$ using a simple and completely general algorithm. In particular, we just solve a kernel ridge regression problem involving $y(t_1), \ldots, y(t_q)$ and an appropriately chosen kernel function for μ. This method runs in $O(q^3)$ time and is already widely used for signal reconstruction, albeit with suboptimal strategies for choosing $t_1, \ldots, t_q$.

Overall, this approach gives the first finite sample, provable approximation bounds for all common Fourier-constrained signal reconstruction problems beyond bandlimited and sparse functions.

Our results are obtained by drawing on a rich set of tools from randomized numerical linear algebra, including sampling methods for approximate regression and deterministic column-based low-rank approximation methods [6, 17]. Many of these methods view matrices as sums of rank-1 outer products and approximate them by sampling or deterministically selecting a subset of these outer products. We adapt these tools to the approximation of continuous operators, which can be written as the (weak) integral of rank-1 operators. For example, our universal time domain sampling distribution is obtained using the notion of *statistical leverage* [1, 21, 58], extended to a continuous Fourier transform operator that arises in the signal reconstruction problem. We hope that, by extending

[1] Approximation bounds can be obtained by truncating the Whittaker-Shannon method; however, they are weak, depending *polynomially*, rather than *logarithmically*, on the desired error ϵ (see full version of this paper [4]).
[2] We formalize our notion of accuracy in Section 2.



(a) Bandlimited. (b) Sparse. (c) Multiband. (d) Nonuniform multiband.

(e) Smooth multiband. (f) Cauchy-Lorentz. (g) Gaussian. (h) Gaussian mixture.

Figure 2: Examples of Fourier transform "priors" induced by various measures μ (we plot the corresponding density). Our algorithm can reconstruct signals under any of these priors.

many of the fundamental contributions of randomized numerical linear algebra to build a toolkit for 'randomized operator theory', our work will offer a starting point for progress on many signal processing problems using randomized methods.

2 FORMAL STATEMENT OF RESULTS

As suggested, we formally capture Fourier structure through any probability measure μ over the reals.[3] We often refer to μ as a "prior", although we will see that it can be understood beyond the context of Bayesian inference. The simplicity of a set of constraints will be quantified by a natural *statistical dimension* parameter for μ, defined in Section 2.1.

For signals with bandlimit F, μ is the uniform probability measure on $[-F, F]$. For multiband signals, it is uniform on the union of k intervals, while for Fourier-sparse functions, μ is uniform on a union of k Dirac measures. More general priors are visualized in Figure 2. Those based on Gaussian or Cauchy-Lorentz distributions are common in scientific applications, and we will discuss examples shortly. For now, we begin with our main problem formulation.

Problem 1. *Given a probability measure μ on $\mathbb{R}$, for any $t \in [0, T]$, define the inverse Fourier transform of $g(\xi)$ with respect to μ as*

$$\left[\mathcal{F}_\mu^* g\right](t) \stackrel{\text{def}}{=} \int_{\mathbb{R}} g(\xi) e^{2\pi i \xi t}\, d\mu(\xi). \tag{1}$$

Suppose our input y can be written as $y = \mathcal{F}_\mu^* x$ for some frequency domain function $x(\xi)$ and, for any chosen t, we can observe $y(t) + n(t)$ for some fixed noise function $n(t)$. Then, for error parameter ϵ, our goal is to recover an approximation $\tilde{y}$ satisfying

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + C\|n\|_T^2, \tag{2}$$

where $\|x\|_\mu^2 \stackrel{\text{def}}{=} \int_{\mathbb{R}} |x(\xi)|^2\, d\mu(\xi)$ is the energy of x with respect to μ, and $\|z\|_T^2 \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T |z(t)|^2 dt$, so that $\|y - \tilde{y}\|_T^2$ and $\|n\|_T^2$ are the mean squared error and noise level respectively. $C \geq 1$ is a fixed constant.

[3] We consider the measure space $(\mathbb{R}, \mathcal{B}, \mu)$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$.

Unlike the $\|x\|_\mu^2$ term in (2), which we can control by adjusting ϵ, we can never hope to recover y to accuracy better than $\|n\|_T^2$. Accordingly, we consider $\|n\|_T^2$ to be small and are happy with any solution of Problem 1 that is within a constant factor of optimal – i.e., where $C = O(1)$.

Problem 1 captures signal reconstruction under all standard Fourier transform constraints, including bandlimited, multiband, and sparse signals.[4] The error in (2) naturally scales with the average energy of the signal over the allowed frequencies. For more general priors, $\|x\|_\mu^2$ will be larger when y contains a significant component of frequencies with low density in μ.[5] For a given number of samples, we would thus incur larger error in (2) in comparison to a signal that uses more "likely" frequencies.

As an alternative to Problem 1, we can formulate signal fitting from a Bayesian perspective. We assume that n is independent random noise, and y is a stationary stochastic process with expected power spectral density μ. This assumption on y's power spectral density is equivalent to assuming that y has covariance function (a.k.a. autocorrelation) $\hat{\mu}(t)$, which is the type of prior used in kriging and Gaussian process regression. While we focus on the formulation of Problem 1 in this work, we also give an informal discussion of the Bayesian setup in the full version [4].

2.0.1 Examples and applications. As discussed in Section 1.2, "hard constraint" versions of Problem 1, such as bandlimited, sparse, and multiband signal reconstruction, have applications in communications, imaging, audio, and other areas of engineering. Generalizations of the multiband problem to non-uniform measures (see Figure 2d) are also useful in various communication problems [38].

On the other hand, "soft constraint" versions of the problem are widely applied in scientific applications. In medical imaging, images are often denoised by setting μ to a heavy-tailed Cauchy-Lorentz measure on frequencies [8, 25, 36]. This corresponds to assuming an exponential covariance function for spatial correlation. Exponential covariance and its generalization, Matérn covariance, are also common in the earth and geosciences [51, 52], as well as in general image processing [45, 49].

A Gaussian prior μ, which corresponds to Gaussian covariance, is also used to model both spatial and temporal correlation in medical imaging [24, 62] and is very common in machine learning. Other choices for μ are practically unlimited. For example, the popular ArcGIS kriging library also supports the following covariance functions: circular, spherical, tetraspherical, pentaspherical, rational quadratic, hole effect, k-bessel, and j-bessel, and stable [30].

2.1 Sample Complexity

With Problem 1 defined, our first goal is to characterize the number of samples required to reconstruct y, as a function of the *accuracy parameter* ϵ, the *range* T, and the *measure* μ. We do so using what we refer to as the *Fourier statistical dimension* of μ, which corresponds

[4] For sparse or multiband signals, Problem 1 assumes frequency or band locations are known *a priori*. There has been significant work on algorithms that can recover y when these locations are not known [12, 37, 39, 47]. Understanding this more complicated problem in the multiband case is an important future direction.

[5] Informally, decreasing $d\mu(\xi)$ by a factor of $c > 1$ requires increasing $x(\xi)$ by a factor of c to give the same time domain signal. This increases $x(\xi)^2$ by a factor of c^2 and so increases its contribution to $\|x\|_\mu^2$ by a factor of $c^2/c = c$.

to the standard notion of statistical or 'effective dimension' for regularized function fitting problems [28, 65].

Definition 2 (Fourier statistical dimension). *For a probability measure μ on $\mathbb{R}$ and time length T, define the kernel operator $\mathcal{K}_\mu : L_2(T) \to L_2(T)$[6] as:*

$$[\mathcal{K}_\mu z](t) \stackrel{\text{def}}{=} \int_{\xi \in \mathbb{R}} e^{2\pi i \xi t} \left[\frac{1}{T} \int_{s \in [0,T]} z(s) e^{-2\pi i \xi s} \, ds \right] d\mu(\xi). \quad (3)$$

Note that $\mathcal{K}_\mu$ is self-adjoint, positive semidefinite and trace-class.[7] *The Fourier statistical dimension for μ, T, and error ϵ is denoted by $s_{\mu,\epsilon}$ and defined as:*

$$s_{\mu,\epsilon} \stackrel{\text{def}}{=} \text{tr}(\mathcal{K}_\mu (\mathcal{K}_\mu + \epsilon \mathcal{I}_T)^{-1}), \quad (4)$$

where $\mathcal{I}_T$ is the identity operator on $L_2(T)$. Letting $\lambda_i(\mathcal{K}_\mu)$ denote the i^{th} largest eigenvalue of $\mathcal{K}_\mu$, we may also write

$$s_{\mu,\epsilon} = \sum_{i=1}^{\infty} \frac{\lambda_i\left(\mathcal{K}_\mu\right)}{\lambda_i\left(\mathcal{K}_\mu\right) + \epsilon}. \quad (5)$$

Note that $\mathcal{K}_\mu$ and $s_{\mu,\epsilon}$, and $\mathcal{F}_\mu$ as defined in Problem 1, all depend on T and thus could naturally be denoted $\mathcal{F}_{\mu,T}$, $\mathcal{K}_{\mu,T}$, and $s_{\mu,\epsilon,T}$. However, since T is fixed throughout our results, for conciseness we do not use T in our notation for these and related notions.

It is not hard to see that $s_{\mu,\epsilon}$ increases as ϵ decreases, meaning that we will require more samples to obtain a more accurate solution to Problem 1. The operator $\mathcal{K}_\mu$ corresponds to taking the Fourier transform of a time domain input $z(t)$, scaling that transform by μ, and then taking the inverse Fourier transform. Readers familiar with the literature on bandlimited signal reconstruction will recognize $\mathcal{K}_\mu$ as the natural generalization of the frequency limiting operator studied in the work of Landau, Slepian, and Pollak on prolate spheroidal wave functions [34, 35, 57]. In that work, it is established that a quantity nearly identical to $s_{\mu,\epsilon}$ bounds the sample complexity of solving Problem 1 for bandlimited functions.

Our first technical result is that this is true *for any prior* μ.

Theorem 1 (Main result, sample complexity). *For any probability measure μ, Problem 1 can be solved using $q = O\left(s_{\mu,\epsilon} \cdot \log s_{\mu,\epsilon}\right)$ noisy signal samples $y(t_1) + n(t_1), \ldots, y(t_q) + n(t_q)$.*

What does Theorem 1 imply for common classes of functions with constrained Fourier transforms? Table 1 includes a list of upper bounds on $s_{\mu,\epsilon}$ for many standard priors.

A complexity of $O(s_{\mu,\epsilon} \cdot \log s_{\mu,\epsilon})$ equates to $\tilde{O}(k)$ samples for k-sparse functions and $\tilde{O}(FT + \log 1/\epsilon)$ for bandlimited functions. Up to log factors, these bounds are tight for these well studied problems. In Section 6, we show that Theorem 1 is actually tight for all common Fourier transform priors: $\Omega(s_{\mu,\epsilon})$ time points are required for solving Problem 1 as long as $s_{\mu,\epsilon}$ grows slower than $1/\epsilon^p$ for some $p < 1$. This property holds for all μ in Table 1. We conjecture that our lower bound can be extended to hold even without this weak assumption.

To compliment the sample complexity bound of Theorem 1, we introduce a *universal method* for selecting samples $t_1, \ldots, t_q$ that nearly matches this complexity. Our method selects samples at random, in a way that *does not depend* on the specific prior μ.

[6] $L_2(T)$ denotes the complex-valued square integrable functions with respect to the uniform measure on $[0, T]$.

[7] See Section 3 for a formal explanation of these facts.

[8] This intuitively matches the asymptotic Landau rate for multiband functions [33].

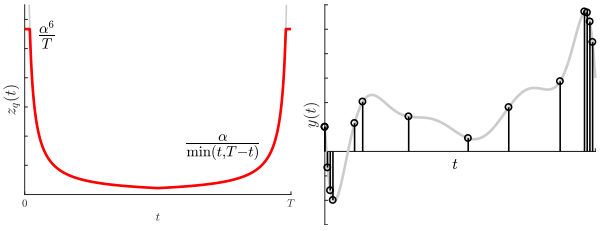
Table 1: Statistical dimension upper bounds for common Fourier interpolation problems. Our result (Theorem 1) requires $O(s_{\mu,\epsilon} \cdot \log s_{\mu,\epsilon})$ samples.

Fourier prior, μ	Statistical dimension, $s_{\mu,\epsilon}$
k-sparse	k
bandlimited to $[-F, F]$	$O\left(FT + \log(1/\epsilon)\right)$
multiband, $F_1, \ldots, F_s$	$O\left(\sum_i F_i T + s \log(1/\epsilon)\right)$[8]
Gaussian, variance F	$O\left(FT\sqrt{\log(1/\epsilon)} + \log(1/\epsilon)\right)$
Cauchy-Lorentz, scale F	$O\left(FT\sqrt{1/\epsilon} + \sqrt{1/\epsilon}\right)$

Theorem 2 (Main result, sampling distribution). *For any sample size q, there is a fixed probability density p_q over $[0, T]$ such that, if q time points $t_1, \ldots, t_q$ are selected independently at random according to p_q, and $q \geq c \cdot s_{\mu,\epsilon} \cdot \log^2 s_{\mu,\epsilon}$ for some fixed constant c, then it is possible to solve Problem 1 with probability 99/100 using the noisy signal samples $y(t_1) + n(t_1), \ldots, y(t_q) + n(t_q)$.*[9]

Theorem 2 is our main technical contribution. By achieving near optimal sample complexity with a universal distribution, it shows that wide range of common Fourier constrained interpolation problems are more closely related than previously understood.

Moreover, p_q (which is formally defined in Theorem 17) is very simple to describe and sample from. As may be intuitive from results on polynomial interpolation, bandlimited approximation, and other function fitting problems, it is more concentrated towards the endpoints of $[0, T]$, so our sampling scheme selects more time points in these regions. The density is shown in Figure 3.



(a) Density for selecting time points. **(b) Example set of nodes sampled according to p_q.**

Figure 3: A plot of the universal sampling distribution, p_q, which can be used to reconstruct a signal under any Fourier transform prior μ. To obtain p_q for a given number of samples q, choose α so that $q = \Theta(\alpha \log^2 \alpha)$. Set $z_q(t) = \alpha / \min(t, T - t)$, except near 0 and T, where the function is capped at $z_q(t) = \alpha^6$. Construct p_q by normalizing so z_q integrates to 1.

[9] In Section 5.4, we formally quantify the tradeoff between success probability and sample complexity.

2.2 Algorithmic Complexity

While Theorem 2 immediately yields an approach for selecting samples $t_1, \ldots, t_q$, it is only useful if we can *efficiently* solve Problem 1 given the noisy measurements $y(t_1) + n(t_1), \ldots, y(t_q) + n(t_q)$. We show that this is possible for a broad class of constraint measures. Specifically, we need only assume that we can efficiently compute the positive-definite kernel function[10]:

$$k_\mu(t_1, t_2) = \int_{\xi \in \mathbb{R}} e^{-2\pi i (t_1 - t_2)\xi} \, d\mu(\xi). \tag{6}$$

The above integral can be approximated via numerical quadrature, but for many of the aforementioned applications, it has a closed-form. For example, when μ is supported on just k frequencies, it is a sum of these frequencies. When μ is uniform on $[-F, F]$, $k_\mu(t_1, t_2) = \text{sinc}(2\pi F(t_1 - t_2))$. For multiband signals with s bands, $k_\mu(t_1, t_2)$ is a sum of s modulated sinc functions. In fact, $k_\mu(t_1, t_2)$ has a closed-form for all μ illustrated in Figure 2. Further details are discussed in full version [4]. Assuming a subroutine for computing $k_\mu(t_1, t_2)$, our main algorithmic result is as follows:

Theorem 3. *(Main result, algorithmic complexity) There is an algorithm that solves Problem 1 with probability* 99/100 *which uses* $O\left(s_{\mu,\epsilon} \cdot \log^2(s_{\mu,\epsilon})\right)$ *time domain samples (sampled according to the distribution given by Theorem 2) and runs in* $\tilde{O}(s_{\mu,\epsilon}^\omega + s_{\mu,\epsilon}^2 \cdot Z)$ *time, assuming the ability to compute* $k_\mu(t_1, t_2)$ *for any* $t_1, t_2 \in [0, T]$ *in* Z *time.*[11] *The algorithm returns a representation of* $\tilde{y}(t)$ *that can be evaluated in* $\tilde{O}(s_{\mu,\epsilon} \cdot Z)$ *time for any* t.

For bandlimited, Gaussian, or Cauchy-Lorentz priors μ, $Z = O(1)$. For s sparse signals or multiband signals with s blocks, $Z = O(s)$.

We note that, while Theorem 3 holds when $\tilde{O}\left(s_{\mu,\epsilon}\right)$ samples are taken, $s_{\mu,\epsilon}$ may be not be known and thus it may be unclear how to set the sample size. In our full statement of the Theorem in Section 5.4 we make it clear that any upper bound on $s_{\mu,\epsilon}$ suffices to set the sample size. The sample complexity will depend on how tight this upper bound is. In the full version we give upper bounds on $s_{\mu,\epsilon}$ for a number of common μ, which can be plugged into Theorem 3.

2.3 Our Approach

Theorems 1, 2, and 3 are achieved through a simple and practical algorithmic framework. In Section 4, we show that Problem 1 can be modeled as a least squares regression problem with ℓ_2 regularization. As long as we can compute $k_\mu(t_1, t_2)$, we can solve this problem using *kernel ridge regression*, a popular function fitting technique in nonparametric statistics [55].

Naively, the kernel regression problem is infinite dimensional: it needs to be solved over the *continuous* time domain $[0, T]$ to reconstruct y. This is where sampling comes in. We need to discretize the problem and establish that our solution over a fixed set of time samples nearly matches the solution over the whole interval. To bound the error of discretization, we turn to a tool from randomized numerical linear algebra: *statistical leverage score sampling* [21, 58]. We show how to *randomly* discretize Problem 1 by sampling time points with probability proportional to an appropriately defined non-uniform leverage score distribution on $[0, T]$. The required number of samples is $O(s_{\mu,\epsilon} \log s_{\mu,\epsilon})$, which proves Theorem 1.

Unfortunately, the leverage score distribution does not have a closed-form, varies depending on ϵ, T, and μ, and likely cannot be sampled from exactly. To prove Theorem 2, we show that for any μ, for large enough q, the closed form distribution p_q *upper bounds* the leverage score distribution. This upper bound closely approximates the true distribution and can thus be used in its place during sampling, losing only a $\log s_{\mu,\epsilon}$ in sample complexity.

The leverage score distribution roughly measures, for each time point t, how large $|y(t)|^2$ can be compared to $\|y\|_T^2$ when y's Fourier transform is constrained by μ (i.e., when $\|x\|_\mu^2$ as defined in Problem 1 is bounded). To upper bound this measure we turn to another powerful result from the randomized numerical linear algebra literature: every matrix contains a small subset of columns that span a near-optimal low-rank approximation to that matrix [9, 19, 53]. In other words, every matrix admits a near-optimal low-rank approximation with *sparse column support*. By extending this result to continuous linear operators, we prove that the smoothness of a signal whose Fourier transform has $\|x\|_\mu^2$ bounded can be bounded by the smoothness of an $O(s_{\mu,\epsilon})$ sparse Fourier function. This lets us apply recent results of [12, 13] that bound $|y(t)|^2$ in terms of $\|y\|_T^2$ for any sparse Fourier function y . Intuitively, our result shows that the simplicity of sparse Fourier functions governs the simplicity of *any class* of Fourier constrained functions.

The above argument yields Theorem 2. Since we can sample from p_q in $O(1)$ time, we can efficiently sample the time domain to $O(s_{\mu,\epsilon} \cdot \log^2 s_{\mu,\epsilon})$ points and then solve Problem 1 by applying kernel ridge regression to these points, which takes $\tilde{O}(s_{\mu,\epsilon}^\omega + s_{\mu,\epsilon}^2 \cdot Z)$ time, assuming the ability to compute $k_\mu(\cdot, \cdot)$ in Z time. This yields the algorithmic result of Theorem 3.

2.4 Roadmap

The rest of this paper is devoted to proving Theorems 1, 2, and 3. In Section 4 we reduce Problem 1 to a kernel ridge regression problem and explain how to randomly discretize and solve this problem via leverage score sampling, proving Theorem 1. In Section 5] we give an upper bound on the leverage score distribution for general priors, proving Theorems 2 and 3. Finally, in Section 6 we prove that, under a mild assumption, the statistical dimension tightly characterizes the sample complexity of solving Problem 1, and thus that our results are nearly optimal. Many details are deferred to the full version of this paper available at [4]. There we also give an in depth overview of related work and prove our extensions of a number of randomized linear algebra primitives to continuous operators. We also bound the statistical dimension for the important case of bandlimited functions. We use this result to prove statistical dimension bounds for multiband, Gaussian, and Cauchy-Lorentz priors (shown in Table 1). Moreover, we show how to compute the kernel function k_μ for these common priors.

3 NOTATION

Let μ be a probability measure on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. Let $L_2(\mu)$ denote the space of complex-valued square integrable functions with respect to μ. For $a, b \in L_2(\mu)$, let $\langle a, b \rangle_\mu$

[10] When y is real valued, it makes sense to consider symmetric μ. In this case, k_μ is also real valued. However, in general it may be complex valued.

[11] For conciseness, we use $\tilde{O}(z)$ to denote $\tilde{O}(z \log^c z)$, where c is some fixed constant (usually ≤ 2). In formal theorem statements we give c explicitly. $\omega < 2.373$ is the current exponent of fast matrix multiplication [61].

denote $\int_{\xi\in\mathbb{R}} a(\xi)^* b(\xi)\, d\mu(\xi)$ where for any $x \in \mathbb{C}$, x^* is its complex conjugate. Let $\|a\|_\mu^2$ denote $\langle a, a\rangle_\mu$. Let $\mathcal{I}_\mu$ denote the identity operator on $L_2(\mu)$. Note that for any μ, $L_2(\mu)$ is a separable Hilbert space and thus has a countably infinite orthonormal basis [29].

We overload notation and use $L_2(T)$ to denote the space of complex-valued square integrable functions with respect to the uniform probability measure on $[0, T]$. It will be clear from context that T is not a measure. For $a, b \in L_2(T)$, let $\langle a, b\rangle_T$ denote $\frac{1}{T}\int_0^T a(t)^* b(t)\, dt$ and let $\|a\|_T^2$ denote $\langle a, a\rangle_T$. Let $\mathcal{I}_T$ denote the identity operator on $L_2(T)$.

Define the Fourier transform operator $\mathcal{F}_\mu : L_2(T) \to L_2(\mu)$ as:

$$\left[\mathcal{F}_\mu f\right](\xi) = \frac{1}{T}\int_0^T f(t) e^{-2\pi i t\xi} dt. \tag{7}$$

The adjoint of $\mathcal{F}_\mu$ is the unique operator $\mathcal{F}_\mu^* : L_2(\mu) \to L_2(T)$ such that for all $f \in L_2(T), g \in L_2(\mu)$ we have $\langle g, \mathcal{F}_\mu f\rangle_\mu = \langle \mathcal{F}_\mu^* g, f\rangle_T$. It is not hard to see that $\mathcal{F}_\mu^*$ is the inverse Fourier transform operator with respect to μ as defined in Section 2, equation (1):

$$\left[\mathcal{F}_\mu^* g\right](t) \stackrel{\text{def}}{=} \int_{\mathbb{R}} g(\xi) e^{2\pi i \xi t}\, d\mu(\xi). \tag{8}$$

Note that the kernel operator $\mathcal{K}_\mu : L_2(T) \to L_2(T)$ originally defined in (3) is equal to

$$\mathcal{K}_\mu = \mathcal{F}_\mu^* \mathcal{F}_\mu.$$

$\mathcal{K}_\mu$ is self-adjoint, positive semidefinite and trace-class and an integral operator with kernel k_μ:

$$[\mathcal{K}_\mu z](t) = \frac{1}{T}\int_0^T k_\mu(s, t) z(s) ds,$$

where k_μ is as defined in (6). The trace of $\mathcal{K}_\mu$ is equal to 1.[12] We will also make use of the Gram operator: $\mathcal{G}_\mu \stackrel{\text{def}}{=} \mathcal{F}_\mu \mathcal{F}_\mu^*$. $\mathcal{G}_\mu$ is also self-adjoint, positive semidefinite, and trace-class.

Remark: It may be useful for the reader to informally regard $\mathcal{F}_\mu$ as an infinite matrix with rows indexed by $\xi \in \mathbb{R}$ and columns indexed by $t \in [0, T]$. Following the definition of $\mathcal{F}_\mu$ above, and assuming that μ has a density p, this infinite matrix has entries given by:

$$\mathcal{F}_\mu(\xi, t) = \sqrt{\frac{p(\xi)}{T}} \cdot e^{-2\pi i t\xi}. \tag{9}$$

The results we apply on leverage score sampling can all be seen as extending results for finite matrices from the randomized numerical linear algebra literature to this infinite matrix.

4 FUNCTION FITTING WITH LEAST SQUARES REGRESSION

Least squares regression provides a natural approach to solving the interpolation task of Problem 1. In particular, consider the following regularized minimization problem over functions $g \in L_2(\mu)$[13]:

$$\min_{g\in L_2(\mu)} \|\mathcal{F}_\mu^* g - (y + n)\|_T^2 + \epsilon\|g\|_\mu^2. \tag{10}$$

The first term encourages us to find a function g whose inverse Fourier transform is close to our measured signal $y + n$. The second term encourages us to find a low energy solution – ultimately, we solve (10) based on only a small number of samples $y(t_1), \ldots, y(t_k)$, and smoother, lower energy solutions will better generalize to the entire interval $[0, T]$. We remark that it is well known that least squares approximations benefit from regularization even in the noiseless case [15].

We first state a straightforward fact: if we minimize (10), even to a coarse approximation, then we are able to solve Problem 1.

Claim 4. *Let $y = \mathcal{F}_\mu^* x$, $n \in L_2(T)$ be an arbitrary noise function, and for any $C \geq 1$, let $\tilde{g} \in L_2(\mu)$ be a function satisfying:*

$$\|\mathcal{F}_\mu^* \tilde{g} - (y + n)\|_T^2 + \epsilon\|\tilde{g}\|_\mu^2 \leq C \cdot \min_{g\in L_2(\mu)} \left[\|\mathcal{F}_\mu^* g - (y + n)\|_T^2 + \epsilon\|g\|_\mu^2\right].$$

Then

$$\|\mathcal{F}_\mu^* \tilde{g} - y\|_T^2 \leq 2C\epsilon\|x\|_\mu^2 + 2(C + 1)\|n\|_T^2.$$

Claim 4 shows that approximately solving the regression problem in (10), with regularization parameter ϵ gives a solution to Problem 1 with parameter $2C\epsilon$ (decreasing the regularization parameter to $\frac{\epsilon}{2C}$ will let us solve with parameter ϵ). But how can we solve the regression problem efficiently? Not only does the problem involve a possibly infinite dimensional parameter vector g, but the objective function also involves the continuous time interval $[0, T]$.

4.1 Random Discretization

The first step is to deal with the latter challenge, i.e., that of a continuous time domain. We show that it is possible to *randomly discretize* the time domain of (10), thereby reducing our problem to a regression problem on a finite set of times $t_1, \ldots, t_q$. In particular, we can sample time points with probability proportional to the so-called *ridge leverage function*, a specific non-uniform distribution that has been applied widely in randomized algorithms for regression and other linear algebra problems on discrete matrices [1, 11, 16, 40, 41].

While we cannot compute the leverage function explicitly for our problem, an issue highlighted in [5], our main result (Theorem 2) uses a simple, but very accurate, closed form approximation in its place. We start with the definition of the ridge leverage function:

Definition 3 (Ridge leverage function). *For a probability measure μ on $\mathbb{R}$, time length $T > 0$, and $\epsilon \geq 0$, we define the ϵ-ridge leverage function for $t \in [0, T]$ as*[14]*:*

$$\tau_{\mu,\epsilon}(t) = \frac{1}{T} \cdot \max_{\{\alpha\in L_2(\mu):\, \|\alpha\|_\mu > 0\}} \frac{\left|[\mathcal{F}_\mu^* \alpha](t)\right|^2}{\|\mathcal{F}_\mu^* \alpha\|_T^2 + \epsilon\|\alpha\|_\mu^2}. \tag{11}$$

[12] Since the kernel is a Fourier transform of a probability measure, it is Hermitian positive definite (Bochner's Theorem). Then we can conclude that $\mathcal{K}_\mu$ is trace-class from Mercer's theorem, and calculate $\mathrm{tr}(\mathcal{K}_\mu) = \frac{1}{T}\int_0^T k_\mu(t, t) dt = 1$.

[13] The fact that the minimum is attainable is a simple consequence of the extreme value theorem, since the search space can be restricted to $\|g\|_\mu^2 \leq \|(y + n)\|_T^2/\epsilon$.

[14] Formally $L_2(T)$ is a space of equivalence classes of functions that differ at a set of points with measure 0. For notational simplicity, here and throughout we use $\mathcal{F}_\mu^* \alpha$ to denote the specific representative of the equivalence class $\mathcal{F}_\mu^* \alpha \in L_2(T)$ given by (8). In this way, we can consider the pointwise value $[\mathcal{F}_\mu^* \alpha](t)$, which we could alternatively express as $\langle \varphi_t, \alpha\rangle_\mu$, for $\varphi_t(\xi) = e^{-2\pi i t\xi}$.

Intuitively, the ridge leverage function at time t is an upper bound on how much a function can "blow up" at t when its Fourier transform is constrained by μ. The denominator term $\|\mathcal{F}_\mu^*\alpha\|_T^2$ is the average squared magnitude of the function $F_\mu^*\alpha$, while the numerator term, $|[\mathcal{F}_\mu^*\alpha](t)|^2$, is the squared magnitude at t. The regularization term $\epsilon\|\alpha\|_\mu^2$ reflects the fact that, to solve (10), we only need to bound the smoothness for functions with bounded Fourier energy under μ. As observed in [44], the leverage function can be viewed as a type of *Christoffel function*, studied in work on orthogonal polynomials and approximation theory [7, 42, 44, 59].

The larger the leverage "score" $\tau_{\mu,\epsilon}(t)$, the higher the probability we will sample time t, to ensure that our sample points well reflect any possibly significant components or 'spikes' of the function y. Ultimately, the integral of the ridge leverage function $\int_0^T \tau_{\mu,\epsilon}(t)dt$ determines how many samples we require to solve (10) to a given accuracy. Theorem 5 below states the already known fact that the ridge leverage function integrates to the statistical dimension [3], which will ultimately allow us to achieve the $\tilde{O}(s_{\mu,\epsilon})$ sample complexity bound of Theorems 1 and 2. Theorem 5 also gives two alternative characterizations of the leverage function that will prove useful. The theorem is proven in the full version, using techniques for finite matrices, adapted to the operator setting.

Theorem 5 (Leverage function properties). *Let $\tau_{\mu,\epsilon}(t)$ be the ridge leverage function (Definition 3) and define $\varphi_t \in L_2(\mu)$ by $\varphi_t(\xi) \stackrel{\text{def}}{=} e^{-2\pi i t\xi}$. We have:*

- *The ridge leverage function integrates to the statistical dimension:*

$$\int_0^T \tau_{\mu,\epsilon}(t)dt = s_{\mu,\epsilon} \stackrel{\text{def}}{=} \text{tr}(\mathcal{K}_\mu(\mathcal{K}_\mu + \epsilon\mathcal{I}_T)^{-1}). \tag{12}$$

- *Inner Product characterization:*

$$\tau_{\mu,\epsilon}(t) = \frac{1}{T}\cdot\langle\varphi_t, (\mathcal{G}_\mu + \epsilon\mathcal{I}_\mu)^{-1}\varphi_t\rangle_\mu. \tag{13}$$

- *Minimization Characterization:*

$$\tau_{\mu,\epsilon}(t) = \frac{1}{T}\cdot\min_{\beta\in L_2(T)}\frac{\|\mathcal{F}_\mu\beta - \varphi_t\|_\mu^2}{\epsilon} + \|\beta\|_T^2. \tag{14}$$

In Theorem 6, we give our formal statement that the ridge leverage function can be used to randomly sample time domain points to discretize the regression problem in (10) and solve it approximately. While complex in appearance, readers familiar with randomized linear algebra will recognize Theorem 6 as closely analogous to standard approximate regression results for leverage score sampling from finite matrices [14]. As discussed, since we are typically unable to sample according to the true ridge leverage function, we give a general result, showing that sampling with any upper bound function with a finite integral suffices.

Theorem 6 (Approximate regression via leverage function sampling). *Assume that $\epsilon \le \|\mathcal{K}_\mu\|_{\text{op}}$.[15] Consider a measurable function $\tilde{\tau}_{\mu,\epsilon}(t)$ with $\tilde{\tau}_{\mu,\epsilon}(t) \ge \tau_{\mu,\epsilon}(t)$ for all t and let $\tilde{s}_{\mu,\epsilon} = \int_0^T \tilde{\tau}_{\mu,\epsilon}(t)dt$. Let $s = c\cdot\tilde{s}_{\mu,\epsilon}\cdot\left(\log\tilde{s}_{\mu,\epsilon} + 1/\delta\right)$ for sufficiently large fixed constant c and let $t_1,\ldots,t_s$ be time points selected by drawing each randomly from $[0,T]$ with probability proportional to $\tilde{\tau}_{\mu,\epsilon}(t)$. For $j\in 1,\ldots,s$, let $w_j = \sqrt{\frac{1}{sT}\cdot\frac{\tilde{s}_{\mu,\epsilon}}{\tilde{\tau}_{\mu,\epsilon}(t_j)}}$. Let $\mathbf{F}:\mathbb{C}^s\to L_2(\mu)$ be the operator defined by:*

$$[\mathbf{F}\,g](\xi) = \sum_{j=1}^s w_j\cdot g(j)\cdot e^{-2\pi i\xi t_j}$$

and $\mathbf{y},\mathbf{n}\in\mathbb{R}^s$ be the vectors with $\mathbf{y}(j) = w_j\cdot y(t_j)$ and $\mathbf{n}(j) = w_j\cdot n(t_j)$. Let:

$$\tilde{g} = \arg\min_{g\in L_2(\mu)}\left[\|\mathbf{F}^*g - (\mathbf{y}+\mathbf{n})\|_2^2 + \epsilon\|g\|_\mu^2\right] \tag{15}$$

With probability $\ge 1-\delta$:

$$\|\mathcal{F}_\mu^*\tilde{g} - (y+n)\|_T^2 + \epsilon\|\tilde{g}\|_\mu^2 \le 3\min_{g\in L_2(\mu)}\left[\|\mathcal{F}_\mu^*g - (y+n)\|_T^2 + \epsilon\|g\|_\mu^2\right]. \tag{16}$$

A generalized version of this result is proven in full version of this paper, which holds even when $\tilde{g}$ is only an approximate minimizer of (15).

Theorem 6 shows that $\tilde{g}$ obtained from solving the discretized regression problem provides an approximate solution to (10) and by Claim 4, $\tilde{y} = \mathcal{F}_\mu^*\tilde{g}$ solves Problem 1 with parameter $\Theta(\epsilon)$. If we have $\tilde{\tau}_{\mu,\epsilon}(t) = \tau_{\mu,\epsilon}(t)$, Theorem 6 combined with Claim 4 shows that Problem 1 with parameter $\Theta(\epsilon)$ can be solved with sample complexity $O\left(s_{\mu,\epsilon}\cdot\log s_{\mu,\epsilon}\right)$, since by (12), $\int_0^T\tau_{\mu,\epsilon}(t)dt = s_{\mu,\epsilon}$. Note that, by simply decreasing the regularization parameter in (10) by a constant factor, we can solve Problem 1 with parameter ϵ. The asymptotic complexity is identical since, by (14), for any $c\le 1$ and any $t\in[0,T]$, $\tau_{\mu,c\epsilon}(t)\le\frac{1}{c}\tau_{\mu,\epsilon}(t)$ and so:

$$s_{\mu,c\epsilon} \le \frac{1}{c}s_{\mu,\epsilon}. \tag{17}$$

This proves the sample complexity result of Theorem 1. However, since it is not clear that sampling according to $\tau_{\mu,\epsilon}(t)$ can be done efficiently (or at all), it does not yet give an algorithm yielding this complexity.[16] This issue will be addressed in Section 5, where we prove Theorem 2.

We show that leverage function sampling satisfies, with good probability, an affine embedding guarantee: that $\|\mathbf{F}^*g - (\mathbf{y}+\mathbf{n})\|_2^2 + \epsilon\|g\|_\mu^2$ closely approximates $\|\mathcal{F}_\mu^*g - (y+n)\|_T^2 + \epsilon\|g\|_\mu^2$ *for all* $g\in L_2(\mu)$. Thus, a (near) optimal solution to the discretized problem,

$$\min_{g\in L_2(\mu)}\left[\|\mathbf{F}^*g - (\mathbf{y}+\mathbf{n})\|_2^2 + \epsilon\|g\|_\mu^2\right],$$

gives a near optimal solution to the original problem,

$$\min_{g\in L_2(\mu)}\left[\|\mathcal{F}_\mu^*g - (y+n)\|_T^2 + \epsilon\|g\|_\mu^2\right].$$

Our proof of the affine embedding property is analogous to existing proofs for finite dimensional matrices [2, 14].

[15] If $\epsilon > \|\mathcal{K}_\mu\|_{\text{op}}$ then (10) is solved to a constant approximation factor by $g = 0$.

[16] We conjecture that the sample complexity can in fact be upper bounded by $O(s_{\mu,\epsilon})$ by adapting deterministic methods for finite matrices to the operator setting [17]

4.2 Efficient Solution of the Discretized Problem

Given an upper bound on the ridge leverage function $\tilde{\tau}_{\mu,\epsilon}(t) \geq \tau_{\mu,\epsilon}(t)$, we can apply Theorem 6 to approximately solve the ridge regression problem of (10) and therefore Problem 1 by Claim 4. In Section 5 we show how to obtain such an upper bound for any μ using a universal distribution.

First, however, we demonstrate how to apply Theorem 6 algorithmically. Specifically, we show how to solve the randomly discretized problem of (15) efficiently. Combined with Theorem 6 and our bound on $\tau_{\mu,\epsilon}(t)$ given in Section 5, this yields a randomized algorithm (Algorithm 1) for Problem 1. The formal analysis of Algorithm 1 is given in Theorem 7.

Algorithm 1 Time Point Sampling and Signal Reconstruction

input: Probability measure $\mu(\xi)$, $\epsilon, \delta > 0$, time bound T, and function $y : [0,T] \to \mathbb{R}$. Ridge leverage function upper bound $\tilde{\tau}_{\mu,\epsilon}(t) \geq \tau_{\mu,\epsilon}(t)$ with $\tilde{s}_{\mu,\epsilon} = \int_0^T \tilde{\tau}_{\mu,\epsilon}(t)dt$.
output: $t_1, \ldots, t_s \in [0,T]$ and $\mathbf{z} \in \mathbb{C}^s$.

1: Let $s = c \cdot \tilde{s}_{\mu,\epsilon} \cdot \left(\log \tilde{s}_{\mu,\epsilon} + \frac{1}{\delta}\right)$ for a large enough constant c.
2: Independently sample $t_1, \ldots, t_s \in [0,T]$ with probability proportional to $\tilde{\tau}_{\mu,\epsilon}(t)$ and set the weight $w_i := \sqrt{\frac{1}{sT} \cdot \frac{\tilde{s}_{\mu,\epsilon}}{\tilde{\tau}_{\mu,\epsilon}(t_i)}}$.
3: Let $\mathbf{K} \in \mathbb{C}^{s \times s}$ be the matrix with $\mathbf{K}(i,j) = w_i w_j \cdot k_\mu(t_i, t_j)$.
4: Let $\bar{\mathbf{y}} \in \mathbb{C}^s$ be the vector with $\bar{\mathbf{y}}(i) = w_i \cdot [y(t_i) + n(t_i)]$.
5: Compute $\bar{\mathbf{z}} := (\mathbf{K} + \epsilon \mathbf{I})^{-1} \bar{\mathbf{y}}$.
6: **return** $t_1, \ldots, t_s \in [0,T]$ and $\mathbf{z} \in \mathbb{C}^s$ with $\mathbf{z}(i) = \bar{\mathbf{z}}(i) \cdot w_i$.

Algorithm 2 Evaluation of Reconstructed Signal

input: Probability measure $\mu(\xi)$, $t_1, \ldots, t_s \in [0,T]$, $\mathbf{z} \in \mathbb{C}^s$, and evaluation point $t \in [0,T]$.
output: Reconstructed function value $\tilde{y}(t)$.

1: For $i \in \{1, \ldots, s\}$, compute $k_\mu(t_i, t) = \int_{\xi \in \mathbb{R}} e^{-2\pi i (t_i - t)} d\mu(\xi)$.
2: **return** $\tilde{y}(t) = \sum_{i=1}^s \mathbf{z}(i) \cdot k_\mu(t_i, t)$.

Theorem 7 (Efficient signal reconstruction given leverage function upper bounds). *Assume that $\epsilon \leq \|\mathcal{K}_\mu\|_{\text{op}}$.[17] Algorithm 1 returns $t_1, \ldots, t_s \in [0,T]$ and $\mathbf{z} \in \mathbb{C}^s$ such that $\tilde{y}(t) = \sum_{i=1}^s \mathbf{z}(i) \cdot k_\mu(t_i, t)$ (as computed in Algorithm 2) satisfies with probability $\geq 1 - \delta$:*

$$\|\tilde{y} - y\|_T^2 \leq 6\epsilon \|x\|_\mu^2 + 8\|n\|_T^2.$$

Suppose we can sample $t \in [0,T]$ with probability proportional to $\tilde{\tau}_{\mu,\epsilon}(t)$ in time W and compute the kernel function $k_\mu(t_1, t_2) = \int_{\xi \in \mathbb{R}} e^{-2\pi i (t_1 - t_2)} d\mu(\xi)$ in time Z. Algorithm 1 queries $y + n$ at $s = O\left(\tilde{s}_{\mu,\epsilon} \cdot \left(\log \tilde{s}_{\mu,\epsilon} + 1/\delta\right)\right)$ points and runs in $O\left(s \cdot W + s^2 \cdot Z + s^\omega\right)$ time[18]. Algorithm 2 evaluates $\tilde{y}(t)$ in $O(s \cdot Z)$ time for any t.

[17] As discussed for Theorem 6, if $\epsilon > \|\mathcal{K}_\mu\|_{\text{op}}$, Problem 1 is trivially solved by $\tilde{y} = 0$.
[18] Here $\omega < 2.373$ is the exponent of fast matrix multiplication. s^ω is the theoretically fastest runtime required to invert a dense $s \times s$ matrix. We note that the s^ω term may be thought of as s^3 in practice, and potentially could be accelerated using a variety of techniques for fast (regularized) linear system solvers.

The proof follows from applying Theorem 6 and Claim 4.

Remark: As discussed, in Section 5 we will give a ridge leverage function upper bound that can be sampled from in $W = O(1)$ time and closely bounds the true leverage function for any μ, giving $\tilde{s}_{\mu,\epsilon} = O(s_{\mu,\epsilon} \log s_{\mu,\epsilon})$. Using this upper bound to sample time domain points, our sample complexity s is thus within a $O(\log s_{\mu,\epsilon})$ factor of the best possible using Theorem 6, which we would achieve if sampling using the true ridge leverage function.

In full version we prove a tighter leverage function bound than the one in Section 5 for bandlimited signals, removing the logarithmic factor in this case. It is not hard to see that for general μ we can also achieve optimal sample complexity by further subsampling $t_1, \ldots, t_s$ using the ridge leverage scores of $\mathbf{K}^{1/2}$. These scores can be computed in $\tilde{O}(s \cdot s_{\mu,\epsilon}^2)$ time using known techniques for finite kernel matrices [40]. Subsampling $O\left(\frac{s_{\mu,\epsilon} \log s_{\mu,\epsilon}}{\delta^2}\right)$ time domain points according to these scores lets us approximately solve the discretized problem of (15) to error $(1+\delta)$.

Applying the more general version of Theorem 6 stated in the full version, this yields an approximate solution to (10) and thus to Problem 1. For constant δ, we need just $O(s_{\mu,\epsilon} \log s_{\mu,\epsilon})$ time samples to solve the subsampled regression problem. By the lower bound given in Section 6, Theorem 19, this complexity is within a $O(\log s_{\mu,\epsilon})$ factor of optimal in nearly all settings. We conjecture that one can achieve within an $O(1)$ factor of the optimal sample complexity by applying deterministic selection methods to $\mathbf{F}$ [17].

5 A NEAR-OPTIMAL SPECTRUM BLIND SAMPLING DISTRIBUTION

In the previous section, we showed how to solve Problem 1 given the ability to sample time points according to the ridge leverage function $\tau_{\mu,\epsilon}$. In general, this function depends on T, μ, and ϵ, and it is not clear if it can be computed or sampled from directly.

Nevertheless, in this section we show that it is possible to efficiently obtain samples from a function that *very closely* approximates the true leverage function for *any* constraint measure μ. In particular we describe a set of closed form functions $\tilde{\tau}_\alpha(t)$, each parameterized by $\alpha > 0$. $\tilde{\tau}_\alpha$ upper bounds the leverage function $\tau_{\mu,\epsilon}$ for any μ and ϵ, as long as the statistical dimension $s_{\mu,\epsilon} \leq O(\alpha)$. Our upper bound satisfies

$$\int_0^T \tilde{\tau}_\alpha(t)dt = O(s_{\mu,\epsilon} \cdot \log s_{\mu,\epsilon}),$$

which means it can be used in place of the true ridge leverage function to give near optimal sample complexity via Theorem 6 and 7. This result is proven formally in Theorem 17, which as a consequence immediately yields our main technical result, Theorem 2. The majority of this section is devoted towards building tools necessary for proving Theorem 17.

5.1 Uniform Leverage Score Bound

We seek a simple closed form function that upper bounds the leverage function $\tau_{\mu,\epsilon}$. Ultimately, we want this upper bound to be very tight, but a natural first question is whether it should exists at all. Is it possible to prove any finite upper bound on $\tau_{\mu,\epsilon}$ without using specific knowledge of μ?

We answer this first question by showing that $\tau_{\mu,\epsilon}$ can be upper bounded by a constant function. Specifically, we show that for $t \in [0,T]$, $\tau_{\mu,\epsilon}(t) \leq C$ for $C = \text{poly}(s_{\mu,\epsilon})$. This upper bound depends on the statistical dimension, but importantly, it does not depend on μ. Formally we show:

Theorem 8 (Uniform leverage function bound). *For all $t \in [0,T]$ and $\epsilon \leq 1$,*[19] $\tau_{\mu,\epsilon}(t) \leq \frac{2^{41}(s_{\mu,\epsilon})^5 \log^3(40 s_{\mu,\epsilon})}{T}$.

While Theorem 8 appears to give a relatively weak bound, proving this statement is a key technical challenge. Ultimately, it is used in Section 5.3 as one of two main ingredients in proving the much tighter leverage function bound that yields Theorems 17 and 2.

Towards a proof of Theorem 8, we consider the operator $\mathcal{F}_\mu$ defined in Section 3. Since $\mathcal{F}_\mu$ has statistical dimension $s_{\mu,\epsilon}$, $\mathcal{K}_\mu = \mathcal{F}_\mu^* \mathcal{F}_\mu$ can have at most $2s_{\mu,\epsilon}$ eigenvalues $\geq \epsilon$:

$$s_{\mu,\epsilon} = \geq \sum_{i:\lambda_i(\mathcal{K}_\mu)\geq\epsilon} \frac{\lambda_i(\mathcal{K}_\mu)}{\lambda_i(\mathcal{K}_\mu)+\epsilon} \geq \frac{\left|i : \lambda_i(\mathcal{K}_\mu) \geq \epsilon\right|}{2}. \tag{18}$$

So, if we project $\mathcal{F}_\mu$ onto $\mathcal{K}_\mu$'s top $2s_{\mu,\epsilon}$ eigenfunctions (when μ is uniform on an interval these are the prolate spherical wave functions of Slepian and Pollak [57]) we will approximate $\mathcal{K}_\mu$ up to its small eigenvalues. The mass of these eigenvalues is at most:

$$\sum_{i:\lambda_i(\mathcal{K}_\mu)\leq\epsilon} \lambda_i(\mathcal{K}_\mu) \leq 2\epsilon \cdot \sum_{i:\lambda_i(\mathcal{K}_\mu)\leq\epsilon} \frac{\lambda_i(\mathcal{K}_\mu)}{\lambda_i(\mathcal{K}_\mu)+\epsilon} \leq 2\epsilon \cdot s_{\mu,\epsilon}.$$

Alternatively, instead of projecting onto the span of the eigenfunctions, we can approximate $\mathcal{K}_\mu$ nearly optimally by projecting $\mathcal{F}_\mu$ onto the span of a subset of $O(s_{\mu,\epsilon})$ of its "rows" – i.e., frequencies in the support of μ. For finite linear operators, is well known that such a subset exists: the problem of finding these subsets has been studied extensively in the literature on randomized low-rank matrix approximation under the name *column subset selection* [9, 19, 53]. In the full version of this paper we show that an analogous result extends to the continuous operator $\mathcal{F}_\mu$:

Theorem 9 (Frequency subset selection). *For some $s \leq \lceil 36 \cdot s_{\mu,\epsilon} \rceil$ there exists a set of distinct frequencies $\xi_1, \ldots, \xi_s \in \mathbb{C}$ such that, if $\mathbf{C}_s : L_2(T) \to \mathbb{C}^s$ and $\mathbf{Z} : L_2(\mu) \to \mathbb{C}^s$ are defined by:*

$$[\mathbf{C}_s g](j) = \frac{1}{T}\int_0^T g(t) e^{-2\pi i \xi_j t} dt \quad \mathbf{Z} = (\mathbf{C}_s \mathbf{C}_s^*)^{-1} \mathbf{C}_s \mathcal{F}_\mu^*,^{20} \tag{19}$$

then

$$\text{tr}(\mathcal{K}_\mu - \mathbf{C}_s^* \mathbf{Z}\mathbf{Z}^* \mathbf{C}_s) \leq 4\epsilon \cdot s_{\mu,\epsilon}. \tag{20}$$

Note that, if $\varphi_t \in L_2(\mu)$ is defined $\varphi_t(\xi) = e^{-2\pi i t \xi}$ and $\boldsymbol{\phi}_t \in \mathbb{C}^s$ is defined $\boldsymbol{\phi}_t(j) = \varphi_t(\xi_j)$, we have:

$$\text{tr}(\mathcal{K}_\mu - \mathbf{C}_s^* \mathbf{Z}\mathbf{Z}^* \mathbf{C}_s) = \frac{1}{T}\int_{t\in[0,T]} \|\varphi_t - \mathbf{Z}^* \boldsymbol{\phi}_t\|_\mu^2 \, dt.$$

Leverage function bound proof sketch. With Theorem 9 in place, we explain how to use this result to prove Theorem 8, i.e., to establish a universal bound on the leverage function of $\mathcal{F}_\mu$. For the sake of exposition, we use the term "row" of an operator $\mathcal{A} : L_2(\mu) \to L_2(T)$ to refer to the corresponding operator restricted to some time t. We use the term "column" of an operator as the adjoint of a row of $\mathcal{A}^* : L_2(T) \to L_2(\mu)$, i.e., the adjoint operator restricted to some frequency ξ.

By Theorem 9, $\mathbf{C}_s^* \mathbf{Z} : L_2(\mu) \to L_2(T)$ (the projection of $\mathcal{F}_\mu^*$ onto the range of $\mathbf{C}_s$) closely approximates the operator $\mathcal{F}_\mu^*$ yet has columns spanned by just $O(s_{\mu,\epsilon})$ frequencies: $\xi_1, \ldots, \xi_s$. So for any $\alpha \in L_2(\mu)$, $\mathbf{C}_s^* \mathbf{Z}\alpha \in L_2(T)$ is just an $O(s_{\mu,\epsilon})$ sparse Fourier function. Using the maximization characterization of Definition 3, we can thus bound the time domain ridge leverage function of $\mathbf{C}_s^* \mathbf{Z}$ by appealing to known smoothness bounds for Fourier sparse functions [13], even for $\epsilon = 0$. When $\epsilon = 0$, the ridge leverage function is known as the *standard leverage function* in the randomized numerical linear algebra literature, and we refer to it as such.

We can use a similar argument to bound the row norms of the residual operator $[\mathcal{F}_\mu^* - \mathbf{C}_s^* \mathbf{Z}]$. The columns of this operator are each spanned by $O(s_{\mu,\epsilon})$ frequencies, and so are again sparse Fourier functions whose smoothness we can bound. This smoothness ensures that no row can have norm significantly higher than average.

Finally, we note that the time domain ridge leverage function of $\mathcal{F}_\mu$ is approximated to within a constant factor by the sum of the standard row leverage function of $\mathbf{C}_s^* \mathbf{Z}$ along with row norms of $\mathcal{F}_\mu - \mathbf{C}_s^* \mathbf{Z}$. This gives us a bound on $\mathcal{F}_\mu$'s ridge leverage function.

Theorem 10 (Ridge leverage function approximation). *Let $\mathbf{C}_s$ and $\mathbf{Z}$ be the operators guaranteed to exist by Theorem 9. Let $\ell(t)$ be the standard leverage function of t in $\mathbf{C}_s^* \mathbf{Z}$:*[21]

$$\ell(t) \stackrel{\text{def}}{=} \max_{\{\alpha \in L_2(\mu):\, \|\alpha\|_\mu > 0\}} \frac{1}{T} \cdot \frac{|[\mathbf{C}_s^* \mathbf{Z}\alpha](t)|^2}{\|\mathbf{C}_s^* \mathbf{Z}\alpha\|_T^2}.$$

Let $r(t)$ be the residual $\frac{1}{T} \cdot \|\varphi_t - \mathbf{Z}^* \boldsymbol{\phi}_t\|_\mu^2$ where φ_t and $\boldsymbol{\phi}_t$ are as defined in Theorem 9. Then for all t:

$$\tau_{\mu,\epsilon}(t) \leq 2 \cdot \left(\ell(t) + \frac{r(t)}{\epsilon}\right)$$

This theorem can be proved by considering the maximization characterization of the ridge leverage function.

With Theorem 10 in place, we now bound $\bar{\tau}_{\mu,\epsilon}(t) = 2\left(\ell(t) + \frac{r(t)}{\epsilon}\right)$, which yields a uniform bound on the true ridge leverage scores.

Lemma 11. *Let $\ell(t), r(t)$ be as defined in Theorem 10 and $\bar{\tau}_{\mu,\epsilon}(t) \stackrel{\text{def}}{=} 2 \cdot \left(\ell(t) + \frac{r(t)}{\epsilon}\right)$. For all $t \in [0,T]$:*

$$\bar{\tau}_{\mu,\epsilon}(t) \leq \frac{15400(36 s_{\mu,\epsilon} + 2)^5 \log^3(36 s_{\mu,\epsilon} + 2)}{T}.$$

Combining Lemma 11 with Theorem 10 yields Theorem 8. We just simplify the constants by noting that for $\epsilon \leq 1$, $s_{\mu,\epsilon} \geq \frac{\text{tr}(\mathcal{K}_\mu)}{2} = \frac{1}{2}$ and so $36 s_{\mu,\epsilon} + 2 \leq 40 s_{\mu,\epsilon}$. Proof of Lemma 11 proceeds by bounding the leverage score $\ell(t)$ and residual $r(t)$ components of $\bar{\tau}_{\mu,\epsilon}(t)$ using a similar argument based on the smoothness of sparse Fourier functions for both. Specifically, for both bounds we employ the following smoothness bound of Chen et al.:

[19] If $\epsilon > 1 = \text{tr}(\mathcal{K}_\mu)$, Problem 1 is trivially solved by returning $\tilde{y} = 0$.

[20] The fact that $\xi_1, \ldots, \xi_s$ are distinct ensures that $(\mathbf{C}_s \mathbf{C}_s^*)^{-1}$ exists.

[21] Analogously to how $[\mathcal{F}_\mu^* \alpha](t)$ is used in Def. 3, while $L_2(T)$ is formally a space of equivalence classes of functions, here we use $\mathbf{C}_s^* \mathbf{Z}\alpha$ to denote the specific representative of the equivalence class $\mathbf{C}_s^* \mathbf{Z}\alpha \in L_2(T)$ given by $[\mathbf{C}_s^* \mathbf{Z}\alpha](t) = \sum_{j=1}^s [\mathbf{Z}\alpha](j) \cdot e^{2\pi i \xi_j t} = \langle \boldsymbol{\phi}_t, \mathbf{Z}\alpha \rangle_{\mathbb{C}^s}$. In this way, we can consider the pointwise value $[\mathbf{C}_s^* \mathbf{Z}\alpha](t)$.

Lemma 12 (Via Lem 5.1 of [12]). *For any $f(t) = \sum_{j=1}^{k} v_j e^{2\pi i \xi_j t}$,*

$$\max_{x\in[0,T]} \frac{|f(x)|^2}{\|f\|_T^2} = 1540 \cdot k^4 \log^3 k.$$

Here we show how to bound the leverage scores $\ell(t)$ of $\mathbf{C}_s^*\mathbf{Z}$. To see how the residuals $r(t)$ is bounded, refer to the full version of this paper. For every $\alpha \in L_2(\mu)$, $\mathbf{C}_s^*\mathbf{Z}\alpha$ is an $s = O(s_{\mu,\epsilon})$ sparse Fourier function. Specifically, we have:

$$[\mathbf{C}_s^*\mathbf{Z}\alpha](t) = \sum_{j=1}^{s} [\mathbf{Z}\alpha](j) \cdot e^{2\pi i \xi_j t},$$

for frequencies $\xi_1, \ldots, \xi_s \in \mathbb{C}$ given by Theorem 9. We can thus directly apply Lemma 12 giving for any $t \in [0, T]$:

$$\begin{aligned}\ell(t) &\overset{\text{def}}{=} \max_{\{\alpha\in L_2(\mu):\|\alpha\|_\mu>0\}} \frac{1}{T}\frac{|[\mathbf{C}_s^*\mathbf{Z}\alpha](t)|^2}{\|\mathbf{C}_s^*\mathbf{Z}\alpha\|_T^2} \\ &\le \max_{\{\alpha\in L_2(\mu):\|\alpha\|_\mu>0\}} \frac{1}{T}\max_{t'\in[0,T]} \frac{|[\mathbf{C}_s^*\mathbf{Z}\alpha](t')|^2}{\|\mathbf{C}_s^*\mathbf{Z}\alpha\|_T^2} \le \frac{1540}{T} s^4 \log^3 s. \end{aligned} \tag{21}$$

Theorem 8 gives a universal uniform bound on the ridge leverage scores corresponding to measure μ in terms of $s_{\mu,\epsilon}$. If we directly sample time points according to the uniform distribution over $[0, T]$, this theorem shows that $\text{poly}(s_{\mu,\epsilon})$ samples and $\text{poly}(s_{\mu,\epsilon})$ runtime suffice to apply Theorem 7 and solve Problem 1 with good probability. This is already a surprising result, showing that the simplest sampling scheme, uniform random sampling, can give bounds in terms of the optimal complexity $s_{\mu,\epsilon}$ for *any* μ. Existing methods with similar complexity, such as those that interpolate bandlimited signals using prolate spheroidal wave functions [56, 64] require nonuniform sampling. Methods that use uniform sampling, such as truncated Whittaker-Shannon, have sample complexity depending polynomially rather than logarithmically on the desired error ϵ.

5.2 Gap-based Leverage Score Bound

Our final result gives a much tighter bound on the ridge leverage scores than the uniform bound of Theorem 8. The key idea is to show that the bound is loose for t bounded away from the edges of $[0, T]$. Specifically we have:

Theorem 13 (Gap-Based Leverage Score Bound). *For all t,*

$$\tau_{\mu,\epsilon}(t) \le \frac{s_{\mu,\epsilon}}{\min(t, T-t)}.$$

Proof. Consider $t \in [0, T/2]$. We will show that $\tau_{\mu,\epsilon}(t) \le \frac{s_{\mu,\epsilon}}{t}$. A symmetric proof will hold for $t \in [T/2, T]$, giving the theorem. We define an auxiliary operator: $\mathcal{F}_{\mu,t} : L_2(T) \to L_2(\mu)$ given by restricting the integration in $\mathcal{F}_\mu$ to $[0, t]$. Specifically, for $f \in L_2(T)$:

$$[\mathcal{F}_{\mu,t} f](\xi) = \frac{1}{T}\int_0^t f(s) e^{-2\pi i s\xi}\, ds. \tag{22}$$

We see that $[\mathcal{F}_{\mu,t}^* g](s) = \int_{\mathbb{R}} g(\xi) e^{2\pi i s\xi}\, d\mu(\xi)$ for $s \in [0, t]$ and $[\mathcal{F}_{\mu,t}^* g](s) = 0$ for $s \in (t, T]$. We will use the leverage score of some $s \in [0, t]$ in the restricted operator $\mathcal{F}_{\mu,t}$ to upper bound those of t in $\mathcal{F}_\mu$. We start by defining these scores as in Definition 3 for $\mathcal{F}_\mu$.

Definition 4 (Restricted ridge leverage scores). *For probability measure μ on $\mathbb{R}$, time length T, $t \in [0, T]$ and $\epsilon \ge 0$, define the ϵ-ridge leverage score of $s \in [0, t]$ in $\mathcal{F}_{\mu,t}$ as:*

$$\tau_{\mu,\epsilon,t}(s) = \frac{1}{T} \cdot \max_{\{\alpha\in L_2(\mu):\|\alpha\|_\mu>0\}} \frac{|[\mathcal{F}_{\mu,t}\alpha](s)|^2}{\|\mathcal{F}_{\mu,t}^*\alpha\|_T^2 + \epsilon\|\alpha\|_\mu^2}.$$

We have the following leverage score properties, analogous to those given for $\mathcal{F}_\mu$ in Theorem 5:

Theorem 14 (Restricted leverage score properties). *Let $\tau_{\mu,\epsilon,t}(s)$ be as defined in Definition 4.*

- *The leverage scores integrate to the statistical dimension:*

$$\int_0^t \tau_{\mu,\epsilon,t}(s)\, ds = s_{\mu,\epsilon,t} \overset{\text{def}}{=} \text{tr}(\mathcal{F}_{\mu,t}^*\mathcal{F}_{\mu,t}(\mathcal{F}_{\mu,t}^*\mathcal{F}_{\mu,t} + \epsilon\mathcal{I}_T)^{-1}). \tag{23}$$

- *Inner Product Characterization: Letting $\varphi_s \in L_2(\mu)$ have $\varphi_s(\xi) = e^{-2\pi i s\xi}$ for $s \in [0, t]$,*

$$\tau_{\mu,\epsilon,t}(s) = \frac{1}{T} \cdot \langle \varphi_s, (\mathcal{F}_{\mu,t}\mathcal{F}_{\mu,t}^* + \epsilon\mathcal{I}_\mu)^{-1}\varphi_s\rangle_\mu. \tag{24}$$

- *Minimization Characterization:*

$$\tau_{\mu,\epsilon,t}(s) = \frac{1}{T} \cdot \min_{\beta\in L_2(T)} \frac{\|\mathcal{F}_{\mu,t}\beta - \varphi_s\|_\mu^2}{\epsilon} + \|\beta\|_T^2. \tag{25}$$

We first note that the restricted leverage scores of Definition 4 are not too large on average.

Claim 15 (Restricted statistical dimension bound).

$$\int_0^T \tau_{\mu,\epsilon,t}(s)\, ds \le s_{\mu,\epsilon}. \tag{26}$$

From Claim 15 we immediately have:

Claim 16. *There exists $s^\star \in [0, t]$ with $\tau_{\mu,\epsilon,t}(s^\star) \le \frac{s_{\mu,\epsilon}}{t}$.*

We now show that the leverage score of $s^\star$ in $\mathcal{F}_{\mu,t}$ upper bounds the leverage score of t in $\mathcal{F}_\mu$, completing the proof of Theorem 13. We apply the minimization characterization of Theorem 14, equation (25), showing that by simply shifting an optimal solution for $s^\star$ we can show the existence of a good solution for t, upper bounding its leverage score by that of $s^\star$ and giving $\tau_{\mu,\epsilon}(t) \le \tau_{\mu,\epsilon,t}(s^\star) \le \frac{s_{\mu,\epsilon}}{t}$ by Claim 16.

Formally, by Claim 16 and (25), there is some $\beta^\star \in L_2(T)$ giving:

$$\frac{1}{T} \cdot \frac{\|\mathcal{F}_{\mu,t}\beta^\star - \varphi_{s^\star}\|_\mu^2}{\epsilon} + \|\beta^\star\|_T^2 = \tau_{\mu,\epsilon,t}(s^\star) \le \frac{s_{\mu,\epsilon}}{t}. \tag{27}$$

We can assume without loss of generality that $\beta^\star(s) = 0$ for $s \notin [0, t]$, since $\mathcal{F}_{\mu,t}\beta^\star$ is unchanged if we set $\beta^\star(s) = 0$ on this range and since doing this cannot increase $\|\beta\|_T^2$. Now, let $\bar\beta \in L_2(T)$ be given by $\bar\beta(s) = \beta^\star(s - (t - s^\star))$. That is, $\bar\beta$ is just $\beta^\star$ shifted from the range $[0, t]$ to the range $[t - s^\star, 2t - s^\star]$. Note that since we are

assuming $t \le T/2$, $[t - s^\star, 2t - s^\star] \subset [0, T]$. For any ξ:

$$\begin{aligned}[\mathcal{F}_\mu \bar{\beta}](\xi) &= \frac{1}{T}\int_0^T \bar{\beta}(s) e^{-2\pi i s\xi} ds \\ &= \frac{1}{T}\int_{t-s^\star}^{2t-s^\star} \beta^\star(s-(t-s^\star)) e^{-2\pi i s\xi} ds \\ &= \frac{1}{T}\int_0^t \beta^\star(s) e^{-2\pi i (s+(t-s^\star))\xi} ds \\ &= [\mathcal{F}_{\mu,t}\beta^\star](\xi)\cdot e^{-2\pi i (t-s^\star)\xi}. \end{aligned} \tag{28}$$

We can write $\varphi_t(\xi) = e^{-2\pi i t\xi} = e^{-2\pi i (t-s^\star)\xi}\cdot \varphi_{s^\star}(\xi)$, which combined with (28) gives:

$$\begin{aligned}\|\mathcal{F}_\mu \bar{\beta} - \varphi_t\|_\mu^2 &= \int_\xi \left|[\mathcal{F}_\mu\bar{\beta}](\xi) - \varphi_t\right|^2 d\mu(\xi) \\ &= \int_\xi \left|([\mathcal{F}_{\mu,t}\beta^\star](\xi) - \varphi_{s^\star})\cdot e^{-2\pi i (t-s^\star)\xi}\right|^2 d\mu(\xi) \\ &= \int_\xi \left|([\mathcal{F}_{\mu,t}\beta^\star](\xi) - \varphi_{s^\star})\right|^2 d\mu(\xi) \\ &= \|\mathcal{F}_{\mu,t}\beta^\star - \varphi_{s^\star}\|_\mu^2. \end{aligned} \tag{29}$$

Finally, since $\|\bar{\beta}\|_T = \|\beta^\star\|_T$, applying the minimization characterization of Theorem 5 the bound in (29) along with (27) gives:

$$\tau_{\mu,\epsilon}(t) \le \frac{\|\mathcal{F}_{\mu,t}\beta^\star - \varphi_{s^\star}\|_\mu^2}{\epsilon} + \|\beta^\star\|_T^2 \le \frac{s_{\mu,\epsilon}}{t},$$

which completes the theorem. □

5.3 Nearly Tight Leverage Score Bound

Combining Theorems 8 and 13 gives our tight, spectrum blind leverage score bound:

Theorem 17 (Spectrum Blind Leverage Score Bound). *For any $\alpha, T \ge 0$ let $\tilde{\tau}_\alpha(t)$ be given by:*

$$\tilde{\tau}_\alpha(t) = \begin{cases} \frac{\alpha}{256\cdot\min(t, T-t)} \text{ for } t \in [T/\alpha^6, T(1-1/\alpha^6)] \\ \frac{\alpha^6}{T} \text{ for } t \in [0, T/\alpha^6] \cup [T(1-1/\alpha^6), T]. \end{cases}$$

For any probability measure μ, $T \ge 0$, $0 \le \epsilon \le 1$ and $t \in [0, T]$, if $\alpha \ge 256 \cdot s_{\mu,\epsilon}$:

$$\tau_{\mu,\epsilon}(t) \le \tilde{\tau}_\alpha(t) \text{ and } \tilde{s}_\alpha \overset{\text{def}}{=} \int_0^T \tilde{\tau}_\alpha(t)\, dt \le \frac{\alpha \cdot \log \alpha}{19}.$$

A visualization of $\tilde{\tau}_\alpha$ is given in Figure 3.

5.4 Putting It All Together

Finally, we combine the leverage score bound of Theorem 17 with Theorem 7 to give our main algorithmic result, Theorem 3 (and as a corollary, Theorem 2). We state the full theorem below:

Theorem 3 (Main result, algorithmic complexity). *Consider any measure μ, for which we can compute the kernel function $k_\mu(t_1, t_2) = \int_{\xi\in\mathbb{R}} e^{-2\pi i (t_1 - t_2)} d\mu(\xi)$ for any $t_1, t_2 \in [0, T]$ in time Z.*

Let $\tilde{\tau}_\alpha(t)$ be as defined in Theorem 17. For any $\epsilon \le \|\mathcal{K}_\mu\|_{\text{op}}$ and $T > 0$, let $\tilde{\tau}_{\mu,\epsilon}(t) = \tilde{\tau}_\alpha(t)$ for $\alpha = \beta \cdot s_{\mu,\epsilon}$ with $\beta \ge 256$. Alg. 1 applied with $\tilde{\tau}_{\mu,\epsilon}(t)$ and failure probability δ returns $t_1, \ldots, t_s \in [0, T]$ and $\mathbf{z} \in \mathbb{C}^s$ such that $\tilde{y}(t) = \sum_{i=1}^s \mathbf{z}(i)\cdot k_\mu(t_i, t)$ solves Problem 1 with parameter 6ϵ and probability $\ge 1 - \delta$. I.e., with probability $\ge 1 - \delta$:

$$\|\tilde{y} - y\|_T^2 \le 6\epsilon\|x\|_\mu^2 + 8\|n\|_T^2.$$

The algorithm queries $y + n$ at s points and runs in $O\left(s^2 \cdot Z + s^\omega\right)$ time where

$$s = O\left(\beta \cdot s_{\mu,\epsilon}\log(\beta\cdot s_{\mu,\epsilon})\cdot[\log(\beta\cdot s_{\mu,\epsilon}) + 1/\delta]\right) = \tilde{O}\left(\frac{\beta\cdot s_{\mu,\epsilon}}{\delta}\right).$$

The output $\tilde{y}(t)$ can be evaluated in $O(sZ)$ time for any t with Alg. 2.

Note that if we want to solve Problem 1 with parameter ϵ, it suffices to apply Theorem 3 with parameter $\epsilon' = \epsilon/6$. The asymptotic complexity will be identical since, by (17), $s_{\mu,\epsilon/6} \le 6 s_{\mu,\epsilon}$.

6 LOWER BOUND

We conclude by showing that the statistical dimension $s_{\mu,\epsilon}$ tightly characterizes the sample complexity of solving Problem 1, under a mild assumption on μ that holds for all natural constraints we discuss in this paper. Thus, Thm. 1 is tight up to logarithmic factors. We first define a natural lower bound on $s_{\mu,\epsilon}$:

$$n_{\mu,\epsilon} \overset{\text{def}}{=} \sum_{i=1}^\infty \mathbb{I}[\lambda_i(\mathcal{K}_\mu) \ge \epsilon]. \tag{30}$$

That is, $n_{\mu,\epsilon}$ is the number of $\mathcal{K}_\mu$'s eigenvalues $\ge \epsilon$. By (18), we always have $n_{\mu,\epsilon} \le 2 s_{\mu,\epsilon}$. We show that solving Problem 1 requires $\Omega(n_{\mu,\epsilon})$ samples. In turn, under a very mild constraint on μ (which holds for all μ we consider including sparse, bandlimited, multiband, Gaussian, and Cauchy-Lorentz), $n_{\mu,\epsilon} = \Omega(s_{\mu,\epsilon})$. Thus, $s_{\mu,\epsilon}$ gives a tight bound on the query complexity of solving Problem 1.

Theorem 18 (Lower bound in terms of eigenvalue count). *Consider a measure μ, an error parameter $\epsilon > 0$, and any (possibly randomized) algorithm that solves Problem 1 with probability $\ge 2/3$ for any function y and makes at most r (possibly adaptive) queries on any input. Then $r \ge n_{\mu,72\epsilon}/20$.*

6.1 Statistical Dimension Lower Bound

We now use Theorem 18 to prove that the statistical dimension tightly characterizes the sample complexity of solving Problem 1 for any constraint measure μ satisfying a simple condition: we must have $s_{\mu,\epsilon} = O(1/\epsilon^p)$ for some $p < 1$. Note that this assumption holds for all μ considered in this work (including bandlimited, multiband, sparse, Gaussian, and Cauchy-Lorentz), where $s_{\mu,\epsilon}$ either grows as $\log(1/\epsilon)$ or $1/\sqrt{\epsilon}$. Also note that by (5) we can always bound $s_{\mu,\epsilon} \le \text{tr}(\mathcal{K}_\mu)/\epsilon = 1/\epsilon$. So this assumption holds whenever we have a nontrivial upper bound on $s_{\mu,\epsilon}$.

Theorem 19 (Statistical Dimension Lower Bound). *Consider any probability measure μ, with $s_{\mu,\epsilon} = O(1/\epsilon^p)$ for constant $p < 1$. Consider any (possibly randomized) algorithm that solves Problem 1 with probability $\ge 2/3$ for any function y and any $\epsilon > 0$ and makes $\le r_{\mu,\epsilon}$ (possibly adaptive) queries on any input. Then $r_{\mu,\epsilon} = \Omega(s_{\mu,\epsilon})$.*[22]

[22] Here we follow the Hardy-Littlewood definition [27], using $f(\epsilon) = \Omega(g(\epsilon))$ to denote that $\limsup_{x\to\infty} \frac{f(\epsilon)}{g(\epsilon)} > 0$. Thus the lower bound shows that, for some fixed constant $c > 0$, for every ϵ, there is at least some $\epsilon' < \epsilon$ where the number of queries used by any algorithm solving Problem 1 with probability $\ge 2/3$ is at least $c \cdot s_{\mu,\epsilon}$. I.e., the lower bound rules out the possibility that the number of queries is $o(s_{\mu,\epsilon})$.

Remark A similar technique to Thm. 19 can be used to show that $n_{\mu,\epsilon} = \Omega(s_{\mu,\epsilon}/\epsilon^p)$ for any $p > 0$, without any assumptions on $s_{\mu,\epsilon}$.

7 CONCLUSION AND OPEN PROBLEMS

We view our work as the starting point for further exploring the application of techniques from the randomized numerical linear algebra literature (such as leverage score sampling, column based matrix reconstruction, and random projection) in signal processing. In the full version we lay out a number of open directions related to higher dimensional setting, learning μ from the samples, and derandomization of our techniques.

ACKNOWLEDGMENTS

We thank Ron Levie for helpful discussions on weak integrals in Hilbert spaces, Zhao Song for discussions on smoothness bounds for sparse Fourier functions, and Yonina Eldar for general discussion and pointers to related work. Haim Avron's work is supported in part by Israel Science Foundation (grant no. 1272/17) and United States-Israel Binational Science Foundation (grant no. 2017698). Michael Kapralov is supported in part by ERC Starting Grant 759471.

REFERENCES

[1] Ahmed Alaoui and Michael W. Mahoney. 2015. Fast Randomized Kernel Ridge Regression with Statistical Guarantees. In *NIPS 2015*. 775–783.
[2] Haim Avron, Kenneth L. Clarkson, and David P. Woodruff. 2017. Sharper Bounds for Regularized Data Fitting. In *RANDOM 2017*.
[3] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. 2017. Random Fourier Features for Kernel Ridge Regression: Approximation Bounds and Statistical Guarantees. In *ICML 2017*.
[4] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. 2018. A Universal Sampling Method for Reconstructing Signals with Simple Fourier Transforms. (2018). arXiv:1812.08723
[5] Francis Bach. 2017. On the Equivalence Between Kernel Quadrature Rules and Random Feature Expansions. *Journal of Machine Learning Research* 18, 21 (2017).
[6] Joshua Batson, Daniel Spielman, and Nikhil Srivastava. 2014. Twice-Ramanujan Sparsifiers. *SIAM Rev.* 56, 2 (2014), 315–334.
[7] Peter Borwein and Tamás Erdélyi. 2012. *Polynomials and Polynomial Inequalities*. Vol. 161. Springer Science & Business Media.
[8] Marc Bourgeois, Frank T. A. W. Wajer, Dirk van Ormondt, and Danielle Graveron-Demilly. 2001. *Reconstruction of MRI Images from Non-Uniform Sampling and Its Application to Intrascan Motion Correction in Functional MRI*. Birkhäuser Boston.
[9] Christos Boutsidis, Michael W. Mahoney, and Petros Drineas. 2009. An Improved approximation algorithm for the column subset selection problem. In *SODA 2009*.
[10] Yoram Bresler and Alber Macovski. 1986. Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34, 5 (1986), 1081–1089.
[11] Daniele Calandriello, Alessandro Lazaric, and Michal Valko. 2016. Analysis of Nyström Method with Sequential Ridge Leverage Score Sampling. In *UAI 2016*.
[12] Xue Chen, Daniel M. Kane, Eric Price, and Zhao Song. 2016. Fourier-Sparse Interpolation without a Frequency Gap. In *FOCS 2016*. 741–750.
[13] Xue Chen and Eric Price. 2018. Active Regression via Linear-Sample Sparsification. *arXiv 1711.10051* (2018).
[14] Kenneth L. Clarkson and David P. Woodruff. 2013. Low Rank Approximation and Regression in Input Sparsity Time. In *STOC 2013*. 81–90.
[15] Albert Cohen, Mark A. Davenport, and Dany Leviatan. 2013. On the Stability and Accuracy of Least Squares Approximations. *Foundations of Computational Mathematics* 13, 5 (01 Oct 2013), 819–834.
[16] Michael B. Cohen, Cameron Musco, and Christopher Musco. 2017. Input Sparsity Time Low-Rank Approximation via Ridge Leverage Score Sampling. In *SODA 2017*.
[17] Michael B. Cohen, Jelani Nelson, and David P. Woodruff. 2016. Optimal Approximate Matrix Product in Terms of Stable Rank. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Dagstuhl, Germany.
[18] Gaspard Riche de Prony. 1795. Essay experimental et analytique: sur les lois de la dilatabilite de fluides elastique et sur celles de la force expansive de la vapeur de l'alcool, a differentes temperatures. *Journal de l'Ecole Polytechnique* (1795).
[19] Amit Deshpande and Luis Rademacher. 2010. Efficient Volume Sampling for Row/Column Subset Selection. In *FOCS 2010*. 329–338.
[20] David L. Donoho. 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52, 4 (2006), 1289–1306.
[21] Petros Drineas and Michael W. Mahoney. 2016. RandNLA: Randomized Numerical Linear Algebra. *Commun. ACM* 59, 6 (2016).
[22] Yonina C. Eldar. 2015. *Sampling Theory: Beyond Bandlimited Systems* (1st ed.). Cambridge University Press, New York, NY, USA.
[23] Yonina C. Eldar and Michael Unser. 2006. Nonideal Sampling and Interpolation from Noisy Observations in Shift-invariant Spaces. *IEEE Transactions on Signal Processing* 54, 7 (2006), 2636–2651.
[24] Karl J. Friston, Peter Jezzard, and Robert Turner. 1994. Analysis of Functional MRI Time-series. *Human Brain Mapping* 1, 2 (1994), 153–171.
[25] Miha Fuderer. 1989. Ringing Artifact Reduction by an Efficient Likelihood Improvement Method. In *Science and Engineering of Medical Imaging*, Vol. 1137.
[26] Mark S. Handcock and Michael L. Stein. 1993. A Bayesian Analysis of Kriging. *Technometrics* 35, 4 (1993), 403–410.
[27] Godfrey Harold Hardy and John Edensor Littlewood. 1914. Some Problems of Diophantine Approximation. *Acta mathematica* 37, 1 (1914), 155–191.
[28] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2002. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.). Springer.
[29] John K. Hunter and Bruno Nachtergaele. 2001. *Applied Analysis*.
[30] Environmental Systems Research Institute. 2018. ArcGIS Desktop: Release 10.
[31] Santhosh Karnik, Zhihui Zhu, Michael B. Wakin, Justin Romberg, and Mark A. Davenport. 2017. The Fast Slepian Transform. *Applied and Computational Harmonic Analysis* (2017).
[32] Vladimir A. Kotelnikov. 1933. On the Carrying Capacity of the Ether and Wire in Telecommunications. *Material for the First All-Union Conference on Questions of Communication, Izd. Red. Upr. Svyazi RKKA* (1933).
[33] Henry J. Landau. 1967. Sampling, Data Transmission, and the Nyquist Rate. *Proc. IEEE* 55, 10 (1967), 1701–1706.
[34] Henry J. Landau and Henry O. Pollak. 1961. Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty – II. *The Bell System Technical Journal* (1961).
[35] Henry J. Landau and Henry O. Pollak. 1962. Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty – III: The Dimension of the Space of Essentially Time- and Band-limited Signals. *The Bell System Technical Journal* 41, 4 (1962).
[36] Alan H. Lettington and Qi He Hong. 1995. Image Restoration Using a Lorentzian Probability Model. *Journal of Modern Optics* 42, 7 (1995), 1367–1376.
[37] Moshe Mishali and Yonina C. Eldar. 2009. Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals. *IEEE Trans. on Signal Processing* (2009).
[38] Moshe Mishali and Yonina C. Eldar. 2010. From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals. *IEEE Journal of Selected Topics in Signal Processing* 4 (2010), 375–391.
[39] Ankur Moitra. 2015. Super-resolution, Extremal Functions and the Condition Number of Vandermonde Matrices. In *STOC 2015*. 821–830.
[40] Cameron Musco and Christopher Musco. 2017. Recursive Sampling for the Nyström Method. In *NIPS 2017*. 3833–3845.
[41] Cameron Musco and David P. Woodruff. 2017. Sublinear Time Low-Rank Approximation of Positive Semidefinite Matrices. *FOCS 2017* (2017).
[42] Paul Nevai. 1986. Géza Freud, Orthogonal Polynomials and Christoffel Functions. A case study. *Journal of Approximation Theory* 48, 1 (1986), 3–167.
[43] Harry Nyquist. 1928. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers* 47, 2 (1928), 617–644.
[44] Edouard Pauwels, Francis Bach, and Jean-Philippe Vert. 2018. Relating Leverage Scores and Density using Regularized Christoffel Functions. In *NIPS 2018*.
[45] Béatrices Pesquet-Popescu and Jacques L. Vehel. 2002. Stochastic fractal models for image processing. *IEEE Signal Processing Magazine* 19, 5 (2002), 48–62.
[46] Vladilen F. Pisarenko. 1973. The Retrieval of Harmonics from a Covariance Function. *Geophysical Journal International* 33, 3 (1973), 347–366.
[47] Eric Price and Zhao Song. 2015. A Robust Sparse Fourier Transform in the Continuous Setting. In *FOCS 2015*. 583–600.
[48] Sathish Ramani, Dimitri van de Ville, and Michael Unser. 2005. Sampling in Practice: is the Best Reconstruction Space Bandlimited?. In *IEEE International Conference on Image Processing*.
[49] S Ramani, D Van De Ville, and M Unser. 2006. Non-Ideal Sampling and Adapted Reconstruction Using the Stochastic Matern Model. In *ICASSP 2006*.
[50] Carl Edward Rasmussen and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.
[51] B. Ripley. 1989. *Statistical Inference for Spatial Processes*. Cambridge Univ. Press.
[52] B. Ripley. 2005. *Spatial statistics*. John Wiley & Sons.
[53] Tamas Sarlos. 2006. Improved Approximation Algorithms for Large Matrices via Random Projections. In *FOCS 2006*. 143–152.
[54] Claude E. Shannon. 1949. Communication in the Presence of Noise. *Proceedings of the Institute of Radio Engineers* 37, 1 (1949), 10–21.
[55] John Shawe-Taylor and Nello Cristianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.
[56] Yoel Shkolnisky, Mark Tygert, and Vladimir Rokhlin. 2006. Approximation of Bandlimited Functions. *Applied and Computational Harmonic Analysis* (2006).
[57] David Slepian and Henry O. Pollak. 1961. Prolate spheroidal wave functions, Fourier analysis and uncertainty – I. *The Bell System Technical Journal* (1961).

[58] Daniel A. Spielman and Nikhil Srivastava. 2011. Graph Sparsification by Effective Resistances. *SIAM J. Comput.* 40, 6 (2011), 1913–1926. STOC 2008.
[59] Vilmos Totik. 2000. Asymptotics for Christoffel Functions for General Measures on the Real Line. *Journal d'Analyse Mathématique* 81, 1 (2000), 283–303.
[60] E T. Whittaker. 1915. On the Functions Which are Represented by the Expansions of the Interpolation Theory. *Proc. of the Royal Soc. of Edinburgh* (1915).
[61] Virginia Vassilevska Williams. 2012. Multiplying Matrices Faster than Coppersmith-Winograd. In *STOC 2012*. ACM, 887–898.
[62] Keith J. Worsley, Sean Marrett, Peter Neelin, Alain C. Vandal, Karl J. Friston, and Alan C. Evans. 1996. A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation. *Human Brain Mapping* 4, 1 (1996).
[63] Hong Xiao. 2001. *Prolate Spheroidal Wavefunctions, Quadrature, Interpolation, and Asymptotic Formulae.* Ph.D. Dissertation. Yale University.
[64] Hong Xiao, Vladimir Rokhlin, and Norman Yarvin. 2001. Prolate Spheroidal Wavefunctions, Quadrature and Interpolation. *Inverse Problems* 17, 4 (2001).
[65] Tong Zhang. 2005. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation* 17, 9 (2005), 2077–2098.