

Compressed Sensing for Analog Signals

Bernhard G. Bodmann Axel Flinth Gitta Kutyniok

March 13, 2018

Abstract

In this paper we develop a general theory of compressed sensing for analog signals, in close similarity to prior results for vectors in finite dimensional spaces that are sparse in a given orthonormal basis. The signals are modeled by functions in a reproducing kernel Hilbert space. Sparsity is defined as the minimal number of terms in expansions based on the kernel functions. Minimizing this number is under certain conditions equivalent to minimizing an atomic norm, the pre-dual of the supremum norm for functions in the Hilbert space. The norm minimizer is shown to exist based on a compactness argument. Recovery based on minimizing the atomic norm is robust and stable, so it provides controllable accuracy for recovery when the signal is only approximately sparse and the measurement is corrupted by noise. As applications of the theory, we include results on the recovery of sparse bandlimited functions and functions that have a sparse short-time Fourier transform.

1 Introduction

Sparse regularization techniques are a nowadays common and successful approach to solve ill-posed inverse problems. Due to constraints imposed by an application one often faces the problem of having access to only very few measured quantities, sometimes far less than the ambient dimension of the space in which the solution to the problem resides. The linear version of this problem – solving an underdetermined linear problem for a solution being sparse in a prescribed representation system – has been extensively studied in the area of what is called compressed sensing or – seemingly more general, but often used interchangeably – sparse recovery [14, 10, 17].

The sparse recovery problem has so far been predominantly treated in discrete settings, following the tacit assumption that it is inherent in the digital domain only. However, real-world signals are foremost of analog nature, calling for an adapted algorithmic framework to this structure. Let us exemplarily consider the problem of determining the dominant frequencies of a musical score. If we assume that the range of possible frequencies is a discrete set – e.g., that the instruments are perfectly tuned according to the well-tempered scale and that the musicians playing the instruments are not making the slightest of mistakes – we can without problem convert this into a discrete setting. As soon as the frequencies deviate slightly from the fixed grid – i.e., if the instruments are not perfectly tuned – it becomes impossible to precisely represent the scores as discrete vectors. This is not merely

an academic issue: As an illustrative example, consider the vector

$$v = (e^{\frac{i\theta k}{m}})_{k=-m}^m \in \mathbb{C}^{2m+1}. \quad (1)$$

Despite its apparent sparsity in the frequency domain, its discrete Fourier transform \hat{v} has many nonzero entries as soon as θ is not an integer multiple of 2π . This phenomenon occurs much more widely, and is commonly called the *basis mismatch problem* [13].

In this paper, we will develop a sparse recovery approach suitable for analog signals and provide precise recovery guarantees, even in the situation of the signals being impacted by noise. The key ideas of our framework are to consider reproducing kernel Hilbert spaces as ambient spaces with the kernel functions being the elements of the representation system, and to solve a total-variation minimization problem for sparse atomic measures as an extension of classical atomic measures.

1.1 Desiderata for Analog Signals

A classical sparse recovery problem aims to recover a vector $x \in \mathbb{R}^n$ from linear, non-adaptive, and underdetermined measurements $b := Ax, A \in \mathbb{R}^{m \times n}$ with n being much larger than m . The key constraint imposed on x is *sparsity*, namely that either

$$\|x\|_0 := \#\{i : x_i \neq 0, i = 1, \dots, n\}$$

is small or that there exists a representation system $\Phi \in \mathbb{R}^{n \times k}$ – often coined *dictionary* and assumed to be an orthonormal basis or more generally a frame – such that $x = \Phi c$ and $\|c\|_0$ is small. One of the most popular and well-explored recovery approaches is to solve a convex optimization problem, more precisely, to find a minimizer for the ℓ_1 -norm that achieves

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_1 \text{ subject to } A\tilde{x} = b.$$

Under certain conditions [21], such a minimizer is indeed the desired solution.

Aiming for a true analog setting, the following four desiderata can be identified:

- (D1) *Hilbert space model.* A classical mathematical model for analog signals consists of reproducing kernel Hilbert spaces, one particularly prominent example being the reproducing kernel Hilbert space of bandlimited functions.
- (D2) *General measurements.* In contrast to the classical setting, in which measured quantities are indexed by a finite or countable, discrete set, we wish to include measurements whose index set is equipped with a more general metric structure. This includes the case of a measurement whose range is a reproducing kernel Hilbert space.
- (D3) *Reconstruction in infinite dimensions.* The reconstruction should not impose a discretization, but remain in the continuum setting. Solving an appropriate convex optimization problem based on the measurement should allow robust and stable recovery of the analog signal with respect to the relevant norm.
- (D4) *Analog sparsity.* The notion of sparsity itself should not be linked to a specific discrete set, but in fact – following the analog spirit – allow for sparse representations within an uncountably infinite dictionary.

Besides developing an appropriate model setting and recovery strategy, the main goal will consist in providing precise recovery guarantees. In particular, we will not address the problem of numerically solving the infinite dimensional optimization problems we will investigate. It should be noted that in some special cases such as when the linear measurements are (related to) Fourier samples (see, e.g., [8, 32, 19, 25]) or piecewise linear measurement functions [23], the infinite dimensional problem can in fact be solved exactly through finite-dimensional problems. Deriving such statements for our more general situation is beyond the scope of this paper and will be studied in future work.

1.2 Related Work

The discrete paradigm of sparse recovery was challenged in the ground-breaking paper [12], where *atomic norms* with respect to infinite dictionaries were introduced. Although the concept of atomic norms was certainly not new in itself, the authors argued that many of the norms used for structured regularization, i.e., the nuclear norm, could be viewed as the atomic norms of certain dictionaries. We will exploit and extend the idea of atomic norms in order to realize our envisaged properties (D1)–(D4). One deficiency of the previously investigated atomic norms is that the norms still relied on discrete representations; in the particular example of the nuclear norm of a matrix, it is equal to the sum of its finitely many singular values.

A different, and more truly “non-discrete” type of sparse recovery problem was already treated in [18], before much of the literature on compressed sensing. The idea in this work was to consider the recovery of a *signed measure* μ of the form

$$\mu = \sum_{i=1}^s c_i \delta_{x_i} \quad (2)$$

for $c \in \mathbb{R}^s$ and $x_i \in \Omega$, rather than a discretized vector. Hence, the theory does address (D4). The other desiderata were not considered in that work; the article only treats the problem of recovering a measure from a part of its Fourier transform $\widehat{\mu}$,

$$\widehat{\mu}(\xi) = \int_0^1 \exp(2\pi i \xi x) d\mu(x), \quad (3)$$

thus not using a more general type of measurement as in (D2). Moreover, rather than proposing a concrete recovery guarantee, the article provides a condition on the support of the measure ensuring the *possibility* of stable recovery of it in form of an oracle inequality, so that (D3) is not fulfilled. The set-up also does not use explicit signal models as in (D1).

The analogue of the ℓ_1 -norm in this setting is the total variation norm, or simply *TV*-norm. Having the success of the ℓ_1 -minimization from the discrete realm in mind, it seems plausible that, given linear measurements $b = A\mu$, a *TV*-norm minimization of the form

$$\min \|\mu\|_{TV} \text{ subject to } A\mu = b \quad (4)$$

should be successful at recovering a measure of the form (2). Notice that the model of a measure resolves the basis mismatch problem described before by viewing v from (1) as samples of the “perfectly sparse” measure δ_θ .

One of the first publications giving a concrete treatise of this optimization problem, in particular, giving recovery guarantees, is [15]. In this work, the authors consider measurements of the form

$$\int_{\Omega} m_k(x) d\mu(x),$$

where $(m_k)_{k=1}^m$ is a so-called *M-system*, meaning that, for each K , any function of the form

$$\sum_{k=1}^K c_k m_k$$

possesses at most K zeroes in Ω . Notice that various relevant measurement families fall into this category such as (trigonometric) polynomials. The authors provide guarantees for the exact recovery of both positive and more general measures, the latter obeying a minimal separation condition.

This setting was treated further in the two contributions [8] and [32]. The authors of [8] assumed that all coefficients $(A\mu)_k$ for $|k| \leq m$ are known; whereas in [32] it is assumed that only a random subsampling is given. In both papers it is then proven that under a condition on the minimal separation of the x_i 's, depending on m , the solution of (4) equals (2). In comparison to [15], the number of measurements needed for resolving measures with a certain separation is much smaller. It should also be mentioned that the authors provide a method for numerically resolving the infinite dimensional *TV*-minimization problem exactly using a finite-dimensional *SDP*.

Relating again to our list of desiderata, in these three articles neither our model assumption (D1) is satisfied, nor are infinite indexed measurement sets allowed. But within the model setting considered in [8] and [32], (D3) and (D4) are indeed fulfilled.

Since the appearance of [8] and [32], research progress on problems of the form (4) has in fact been quite limited. A few articles on purely theoretical considerations, such as existence and structure of the solutions, on the behavior of discretized versions of (4) or on questions of identifiability (e.g., if a measurement operator A is injective on the set of sparse atomic measures) have been published [33, 20, 23, 22, 7]. But concerning the question of concrete recovery guarantees, the literature is quite scarce.

There are a few notable exceptions. We would like to mention [25] and [19], in which two-dimensional Fourier measurements are considered. The first paper analyzes the situation of sampling on a rectangular grid, whereas the second paper studies sampling along radial lines. Although the articles without doubt contain important and substantial new ideas, the specific Fourier nature of the measurements allow their arguments to make heavy use of the results from [8, 32]. Hence, they fail to meet the desideratum (D2).

During the final preparation of this manuscript, we also became aware of [2]. This paper treats a somehow different type of measurements, namely short time Fourier measurements of measures on \mathbb{R} (or the torus). The authors furthermore allow for the measures to have countably many support points, which by itself imposes many delicate problems not present in the model of finitely many support points considered here. In [2], all of the desiderata are met with the exception of (D2), since the authors only treat the case of short time Fourier

measurements.

Let us finally comment on another line of work which constitutes a generalization of compressed sensing to infinite dimensions (see, e.g., [1, 24]). In these papers, it is assumed that the signal f has a sparse representation in a countable representation system $(\psi_k)_{k \in \mathbb{Z}}$, e.g.,

$$f = \sum_{k \in \mathbb{Z}} d_k \psi_k$$

with d being (approximately) sparse (see, for example, [1, 24]). This model is quite general, since it includes cases such as assuming f to be piecewise smooth and choosing $(\psi_k)_{k \in \mathbb{Z}}$ to be equal to a wavelet system, which then leads to d being approximately sparse. It however cannot capture sparsity in the sense of (D4).

The main ideas of this line of work is to carefully sample the function f using filters $(s_n)_n$, resulting in a countable compressed sensing problem of the form

$$\langle s_\ell, f \rangle = \sum_{k \in \mathbb{Z}} d_k \langle s_\ell, \psi_k \rangle, \ell \in \mathbb{Z}.$$

This problem is subsequently, again with great care, approximated with finite dimensional compressed sensing problems. These can be solved with methodologies such as Basis Pursuit, leading to a provably stable recovery procedure. This philosophy is quite different from ours: whereas this approach first discretize and then optimize, we will analyze an infinite dimensional optimization problem directly. Hence, these approaches do also not fulfill (D3). We believe that these types of theories should be considered to be entirely separate from, rather than competing with, the one we are treating, since it is designed for and applies to an entirely different class of signals.

1.3 Our Contribution

Let us for intuition purposes first focus on the case of Fourier measurements (3). It is then clear that the recovery of μ from $M\mu$ is equivalent to the task of recovering the parameters (c_i, x_i) from samples

$$\sum_{\ell=1}^s c_i e^{2\pi i k x_i}, \quad k = -m, \dots, m$$

of the sum of complex exponentials $P(\xi) = \sum_{\ell=1}^s c_i e^{2\pi i \xi x_i}$. Historically, this problem was in fact already studied by Prony in the late 19th century. The algorithm he proposed is of algebraic nature and consists of re-encoding the samples into a polynomial, which then can be proven to have zeroes exactly in the points x_i . In recent years, it has experienced quite a renaissance, for instance, in the context of signals with finite rate of innovation [34]. It is commonly referred to as Prony's algorithm [29]. This algorithm is however very specifically tailored to the special case of Fourier measurements on a uniform grid, and is furthermore quite sensitive to noise.

A different interpretation of the problem – which is of inevitable importance for us – is that the sum of complex exponentials P can be seen as a signal having a sparse representation in the *dictionary of complex exponentials*

$$\Phi = (\exp(2\pi ix(\cdot)))_{x \in [0,1]}.$$

With the methodology of [12] in mind, it seems natural to minimize the atomic norm over all elements Q (in some appropriately chosen ambient space) having the same – for now discrete – measurements $(Q(k))_{k=-m}^m$ as P , with the goal to find the parameters (c_i, x_i) sparsely representing the element P .

In this article, we will follow this general strategy. Our ambient model spaces will be reproducing kernel Hilbert spaces \mathcal{H} , thereby satisfying (D1). The kernel functions $(K_x)_{x \in \mathcal{H}}$ of those Hilbert spaces will serve as dictionary functions, allowing for an analog version of sparsity (D4). The atomic norm with respect to such dictionaries will be defined through a *TV*-minimization problem. Hence problems of the form (4) – extended to countably infinite indexed measurement sets (D2) – can readily be interpreted as atomic norm minimizations.

This recovery strategy then also fulfills (D3). Considering this framework for sparse recovery for analog signals, we derive precise recovery guarantees in the noiseless (Theorem 1) and noise-impacted (Theorems 2 and 3) situations. The general ideas are the same as the line of work started by [8, 32] – however, formulating the problem in reproducing kernel Hilbert spaces sheds new light on it, and allows for unification. In particular, the work [8] can be seen as a special case of the theory. We will also be able to derive results for measurement operators which have not been considered before, related to bandlimited functions and functions with a sparse time-frequency representation.

1.4 Outline

This paper is organized as follows. In Section 2, we introduce and prove a few fundamental properties of atomic norms on reproducing kernel Hilbert spaces. Then, in Section 3, we present our recovery guarantees. Finally, in Section 4, we apply the framework of the previous sections to two concrete settings.

2 Reproducing Kernel Hilbert Spaces and Atomic Norms

To set up our model situation according to (D1), we start by recalling the key definitions and notation related to reproducing kernel Hilbert spaces. For the sparse recovery approach, we in addition require a suitable norm on this space. The so-called atomic norm unified various sparse recovery settings in Euclidean space, and it seems conceivable that such a type of norm will also be highly beneficial for our general setting. This atomic norm was originally defined in [12] in a slightly different, more geometric manner. In the sequel, we will introduce a version specifically adapted to the reproducing kernel Hilbert space setting.

2.1 Reproducing Kernel Hilbert Spaces Revisited

Let us start with recalling the definition of a reproducing kernel Hilbert space.

Definition 1. A *reproducing kernel Hilbert space* \mathcal{H} over a set \mathcal{X} is a set of functions on \mathcal{X} forming a Hilbert space such that for each $x \in \mathcal{X}$, there is a function $K_x \in \mathcal{H}$ that produces the point evaluation of any $f \in \mathcal{H}$ at x via the inner product with K_x , i.e.,

$$f(x) = \langle f, K_x \rangle. \quad (5)$$

The *reproducing kernel* is the function $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$ given by $K(x, y) = K_x(y)$, $x, y \in \mathcal{X}$. We say that \mathcal{H} has a *unit-norm reproducing kernel*, if $K(x, x) = 1$ for all $x \in \mathcal{X}$.

The inner product of \mathcal{H} induces the usual norm and the topology with respect to which \mathcal{H} is complete. From the fact that each K_x furnishes the point evaluation, one can deduce that the span of $\{K_x\}_{x \in \mathcal{X}}$ is dense in \mathcal{H} . If the kernel K is not unit-norm, then one can switch to a normalized one, $\tilde{K}(x, y) = K(x, y)/(K(x, x)K(y, y))^{1/2}$ and pass the weight onto the functions. This defines a Hilbert space isomorphism $\tilde{\cdot} : \mathcal{H} \rightarrow \tilde{\mathcal{H}}$, $\tilde{f}(x) = \sqrt{K(x, x)}f(x)$, $x \in \mathcal{X}$ to the Hilbert space $\tilde{\mathcal{H}}$ with unit-norm reproducing kernel \tilde{K} . This change is useful to show that each function has a norm-convergent expansion in an at most countable number of kernel functions. Assuming a unit-norm kernel function, then the weak greedy algorithm described by Temlyakov [31] provides a norm-convergent expansion for each $\tilde{f} \in \tilde{\mathcal{H}}$, and by unitarity, also for each $f \in \mathcal{H}$ in terms of an at most countable number of kernel functions $\{K_x\}_{x \in \mathcal{X}}$.

Although each function f can be represented as a countable norm-convergent sequence, it will turn out to be convenient to allow also expansions of the form

$$K * \mu \equiv \int_{\mathcal{X}} K_x d\mu(x), \quad (6)$$

where μ is an element of $\mathcal{M}(\mathcal{X})$, the space of regular complex Borel measures of bounded total variation on a locally compact Hausdorff space \mathcal{X} . For the purposes of this paper, we understand the integral of the vector-valued function $x \mapsto K_x$ with respect to the complex measure μ in the sense of Bochner [35], for which it is sufficient to assume that \mathcal{H} is separable, K is a unit-norm kernel, and each $f \in \mathcal{H}$ is Borel measurable [28, 30]. A simple condition that ensures measurability is to have $x \mapsto K_x$ continuous from \mathcal{X} to \mathcal{H} , which implies through the Cauchy-Schwarz inequality and the reproducing kernel identity (5) that each $f \in \mathcal{H}$ is continuous. All of the examples we consider have this property.

2.2 Atomic Norms on Reproducing Kernel Hilbert Spaces

The space $\mathcal{M}(\mathcal{X})$ is naturally equipped with the total-variation norm, which assigns to $\mu \in \mathcal{M}(\mathcal{X})$ the value

$$\|\mu\|_{TV} \equiv |\mu|(\mathcal{X}) = \sup_{\substack{\bigcup_{i=1}^n E_i = \mathcal{X} \\ E_i \text{ disjoint}}} \sum_{i=1}^n |\mu(E_i)|,$$

where the elements of the partitions of \mathcal{X} are Borel measurable sets. The Riesz-Markov-Kakutani theorem identifies this normed space with the dual of the space $\mathcal{C}_0(\mathcal{X})$ of continuous functions vanishing at infinity. In the following, we will frequently view $\mathcal{M}(\mathcal{X})$ as a

locally convex vector space, equipped with the weak- $*$ -topology that it obtains as the dual of $\mathcal{C}_0(\mathcal{X})$.

We may now define the atomic norm in \mathcal{H} .

Definition 2. Let \mathcal{X} be a locally compact Hausdorff space and $\mathcal{M}(\mathcal{X})$ the space of regular complex Borel measures with bounded total variation. Let \mathcal{H} be a separable reproducing kernel Hilbert space of Borel measurable functions on \mathcal{X} with a unit-norm reproducing kernel K . The *atomic norm* of \mathcal{H} associates with each $f \in \mathcal{H}$ the value

$$\|f\|_A = \inf\{\|\mu\|_{TV} : \mu \in \mathcal{M}(\mathcal{X}), K * \mu = f\}. \quad (\mathcal{P})$$

By the triangle inequality for the norm, the Radon-Nikodym property [28, 30], and the normalization of each K_x ,

$$\|K * \mu\| = \left\| \int_{\mathcal{X}} K_x d\mu(x) \right\| \leq \int_{\mathcal{X}} \|K_x\| d|\mu|(x) = |\mu|(\mathcal{X}),$$

the map $D : \mu \mapsto K * \mu$ is (strongly) continuous. This allows us to prove the following lemma.

Lemma 1. In addition to the assumptions on \mathcal{X} and \mathcal{H} , let \mathcal{X} be second countable and $\mathcal{H} \cap \mathcal{C}_0(\mathcal{X})$ be dense in \mathcal{H} , then for $f \in \mathcal{H}$ with $\|f\|_A < \infty$, there exists a $\mu \in \mathcal{M}(\mathcal{X})$ with $f = K * \mu$ and $\|\mu\|_{TV} = \|f\|_A$.

Proof. What we need to prove is that (\mathcal{P}) has a minimizer. By \mathcal{X} being second countable, $\mathcal{C}_0(\mathcal{X})$ is separable. Take a number $R \in \mathbb{R}$ with $R > \|f\|_A$. Using Helly's version of Banach-Alaoglu for dual spaces of separable Banach spaces, the closed ball $\{\mu \in \mathcal{M}(\mathcal{X}) : \|\mu\|_{TV} \leq R\}$ is weak- $*$ sequentially compact. Moreover, $D^{-1}(\{f\})$ is weak- $*$ closed because it is the intersection of the weak- $*$ closed sets $M_g = \{\mu : \langle K * \mu, g \rangle = \int_{\mathcal{X}} \bar{g} d\mu = \langle f, g \rangle\}$ over all $g \in \mathcal{H} \cap \mathcal{C}_0(\mathcal{X})$. Hence, any minimizing sequence for (\mathcal{P}) has a weak- $*$ convergent subsequence. The limit μ of this subsequence then satisfies $K * \mu = f$ and $\|\mu\|_{TV} = \|f\|_A$. \square

Although the atomic norm does not appear to be discussed in the context of Hilbert spaces with reproducing kernel, its dual is more commonly known.

Proposition 1. Given $g \in \mathcal{H}$, then the dual of the atomic norm is

$$\|g\|_A^* = \sup_{x \in \mathcal{X}} |g(x)| = \|g\|_{\infty}.$$

Proof. Using Lemma 1 and the definition of the dual norm, we see that

$$\|g\|_A^* = \sup_{\|f\|_A \leq 1} |\langle g, f \rangle| \leq \sup_{\|\mu\|_{TV} \leq 1} |\langle g, K * \mu \rangle| = \sup_{\|\mu\|_{TV} \leq 1} \left| \int_{\mathcal{X}} g(x) d\mu(x) \right| = \|g\|_{\infty}.$$

Hence, $\|g\|_A^* \leq \|g\|_{\infty}$. To see the opposite inequality, let $\epsilon > 0$ be arbitrary but fixed and x be such that $|g(x)| \geq (1 - \epsilon) \|g\|_{\infty}$. We then have

$$(1 - \epsilon) \|g\|_{\infty} \leq |\langle g, K_x \rangle| \leq \|g\|_A^* \|K_x\|_A \leq \|g\|_A^*,$$

since $\|K_x\|_A \leq 1$ due to $K_x = K * \delta_x$. Since $\epsilon > 0$ was arbitrary, this proves $\|g\|_{\infty} \leq \|g\|_A^*$. \square

Let us now discuss the connection to the original atomic norm definition from [12]. As said before, this notion has a more geometric nature: If Φ is a symmetric subset of a real finite dimensional vector space, the authors defined the atomic norm $\|\cdot\|_{A'}$ as

$$\|x\|_{A'} = \inf \{t > 0 \mid tx \in \overline{\text{conv}}(\Phi)\}, \quad (7)$$

where $\overline{\text{conv}}(\Phi)$ is the closed convex hull of the set Φ . When \mathcal{X} is a second countable locally compact metric space, and we take $\Phi = (\omega \cdot K_x)_{|\omega|=1, x \in \mathcal{X}}$, then the previous definition presented in (7) is equivalent to our TV -optimization based definition also in the infinite dimensional setting. We prepare this statement with an approximation lemma that concentrates the support of elements in $\mathcal{M}(\mathcal{X})$.

Lemma 2. Let \mathcal{X} be a second countable locally compact metric space, then the set of finitely supported complex measures

$$\Delta = \left\{ \sum_{i \in P} c_i \delta_{x_i} \mid |P| < \infty, x_i \in \mathcal{X}, c_i \in \mathbb{C}, i \in P \right\}$$

is weak- $*$ -dense in $\mathcal{M}(\mathcal{X})$. More specifically, for each $\mu \in \mathcal{M}(\mathcal{X})$, there exists a sequence $(\nu_n)_{n \in \mathbb{N}}$ in Δ such that $\nu_n \xrightarrow{*} \mu$ and for each $n \in \mathbb{N}$, $\|\nu_n\|_{TV} \leq \|\mu\|_{TV}$.

Proof. The fact that Δ is dense in the weak- $*$ topology is by the double-perp theorem equivalent to the pre-annihilator of Δ being the trivial subspace $\{0\}$. The Dirac measures are included in Δ , so if $f \in C_0(\mathcal{X})$ is annihilated by each element in Δ then it vanishes.

For the approximation result, let $\mu \in \mathcal{M}(\mathcal{X})$ be given. For each $n \in \mathbb{N}$, choose a compact set C_n such that $|\mu|(C_n) > \|\mu\|_{TV} - 1/n$. Next, by the definition of the total variation norm, let $\{A_k^{(n)}\}_{k=1}^{m(n)}$ be such that $\sum_{k=1}^{m(n)} |\mu(A_k^{(n)})| > |\mu|(C_n) - 1/n$. Using a countable basis for the topology consisting of open balls, for any $\delta > 0$ we refine the partition to obtain a Borel measurable partition $\{A_k^{(n,\delta)}\}_{k=1}^{m(n,\delta)}$ whose sets have a diameter of at most δ .

We now specify the sequence by choosing a sequence of diameters, $\delta_n = 1/n$; we select a point $x_k^{(n)} \in A_k^{(n,1/n)}$ for each $k \leq m(n, 1/n)$, and let

$$\nu_n = \sum_{k=1}^{m(n,n^{-1})} \mu(A_k^{(n,n^{-1})}) \delta_{x_k^{(n)}}.$$

By definition, $\|\nu_n\|_{TV} \leq \|\mu\|_{TV}$. We wish to prove $\nu_n \xrightarrow{*} \mu$. For this, let $f \in C_0(\mathcal{X})$. Since f is vanishing at infinity, f is uniformly continuous, and hence given $\epsilon > 0$, $|f(x) - f(x_k^{(n)})| \leq \epsilon$ for all $x \in A_k^{(n,1/n)}$ provided n is large enough.

This implies that

$$\begin{aligned} \left| \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f d\nu_n \right| &\leq \left| \int_{C_n} \left(f - \sum_{k=1}^{m(n,n^{-1})} f(x_k^{(n)}) \mathbb{1}_{A_k^{(n,n^{-1})}} \right) d\mu \right| + \|f\|_{\infty} |\mu|(\mathcal{X} \setminus C_n) \\ &\leq \|\mu\|_{TV} \epsilon + \|f\|_{\infty} / n \leq C\epsilon \end{aligned}$$

for n large enough, where the constant C only depends on μ and f . Hence, $\nu_n \xrightarrow{*} \mu$. \square

This now allows us to discuss to which extent both definitions are equivalent.

Proposition 2. Let \mathcal{X} be a second countable locally compact metric space, \mathcal{H} be a separable reproducing kernel Hilbert space over \mathcal{X} with unit-norm kernel K such that $\mathcal{C}_0(\mathcal{X}) \cap \mathcal{H}$ is dense in \mathcal{H} . Let $\Phi = (\omega \cdot K_x)_{|\omega|=1, x \in \mathcal{X}}$, then $\|\cdot\|_{A'}$ defined in (7) is equal to $\|\cdot\|_A$.

Proof. Let $f \in \mathcal{H}$ be arbitrary and $t > \|f\|_{A'}$. Then there exists a $g \in \overline{\text{conv}} \Phi$ such that $f = tg$. By assumption, the function g can be expressed as a limit $\lim_{n \rightarrow \infty} g_n$ with

$$g_n = \sum_{i \in M_n} \theta_i^n \omega_i^n K_{x_i^n} = K * \left(\sum_{i \in M_n} \theta_i^n \omega_i^n \delta_{x_i^n} \right),$$

for sets $|M_n| < \infty$, real, positive, scalars $(\theta_i^n)_{i \in M_n}$ with $\sum_{i \in M_n} \theta_i^n \leq 1$ and unimodular complex numbers $(\omega_i^n)_{i \in M_n}$. The corresponding sequence of measures $\mu_n = \sum_{i \in M_n} \theta_i^n \omega_i^n \delta_{x_i^n}$ have norms bounded above by one and therefore, due to sequential weak-* compactness of the closed unit ball in $\mathcal{M}(\mathcal{X})$, there is a weak-* convergent subsequence $\mu_{n'} \xrightarrow{*} \mu_*$, $\|\mu_*\|_{TV} \leq 1$. Due to continuity of D as a map from $\mathcal{M}(\mathcal{X})$ to \mathcal{H} equipped with the weak topology, $f = tg = \text{w-lim } tg_n = \text{w-lim } tD\mu_n = tD\mu_*$, which proves that $\|f\|_A \leq \|t\mu_*\|_{TV} \leq t$. Since $t > \|f\|_{A'}$ was arbitrary, we conclude $\|f\|_A \leq \|f\|_{A'}$.

If on the other hand $\|f\|_A = t$, there exists by Lemma 1 some μ with $f = D\mu$ and $\|\mu\|_{TV} = t$. Lemma 2 shows that there exists a sequence of finitely supported measures $\mu_n = \sum_{i \in M_n} c_i^n \delta_{x_i^n}$ weak-* converging to μ , with $\|\mu_n\| \leq \|\mu\|$. Continuity of D as above implies that $f = \text{w-lim } D\mu_n = \text{w-lim}_{n \rightarrow \infty} \sum_{i \in M_n} c_i^n K_{x_i^n}$. By decomposing $c_i^n = t\theta_i^n \omega_i^n$ with $|\omega_i^n| = 1$, $\theta_i^n \geq 0$, and $\sum_{i \in M_n} \theta_i^n \leq 1$, we see that f is in the weak closure of the convex hull of $t\Phi$, which coincides with the strong closure, so $f \in t\overline{\text{conv}} \Phi$. This completes the proof. \square

3 Sparse Recovery by Atomic Norm Minimization

Having prepared and introduced our general model situation following (D1), we now aim for developing a theory for sparse recovery based on atomic norm minimization satisfying our desiderata (D2)–(D4).

3.1 Sparsity in Reproducing Kernel Hilbert Spaces

We start by formalizing the notion of sparsity we aim to exploit. Notice that this definition is indeed in line with (D4).

Definition 3. A function $f \in \mathcal{H}$ is called *s-sparse*, if there exist sequences $\{c_n\}_{n=1}^s$ in \mathbb{C}^s and $\{x_n\}_{n=1}^s$ in \mathcal{X}^s , such that

$$f = \sum_{n=1}^s c_n K_{x_n}$$

and for any other expansion $f = \sum_{n=1}^{s'} c'_n K_{x'_n}$ of f we have

$$s \leq s'.$$

It would be useful if any linear combination of s kernel functions gave an s -sparse function. In order to guarantee this, we need to rule out linear dependencies of the kernel functions. This property is present in the function spaces we consider here, but proving it can be challenging, as seen in the conjecture by Heil, Ramanathan and Topiwala [26]. The key notion for ensuring this property is made precise by the following definition.

Definition 4. A Hilbert space with reproducing kernel has the s -HRT property, if for any set of s points $\{x_n\}_{n=1}^s$ in \mathcal{X} the corresponding kernel functions form a linearly independent set $\{K_{x_n}\}_{n=1}^s$.

The next result shows that indeed the s -HRT property leads to the required injectivity when a complex measure with support of size at most s is mapped the corresponding expansion in terms of kernel functions.

Proposition 3. If \mathcal{H} has the $2s$ -HRT property, then any $f \in \mathcal{H}$ of the form $f = \sum_{n=1}^s c_n K_{x_n}$ with $c \in \mathbb{C}^s$ and $x \in \mathcal{X}^s$ is s -sparse and the choice of $c \in \mathbb{C}^s$ and $x \in \mathcal{X}^s$ is uniquely determined by f .

Proof. If this were not the case, then there would be two non-identical linear combinations of at most s terms giving f . Equating them shows that the zero vector can be written as a non-trivial linear combination of at most $2s$ terms, which contradicts the $2s$ -HRT property. \square

The next two examples shall show that the s -HRT property is by far not artificial and in fact occurring in many well-known situations.

Example 1. Let \mathcal{X} be the torus, identified with $[0, 1)$ and the metric distance $d(x, y) = \min\{|x - y|, |x - y + 1|, |x - y - 1|\}$ between $x, y \in [0, 1)$. For $m \in \mathbb{N}$, let the Dirichlet kernel be given by $D_m(x) = \frac{1}{2m+1} \sum_{k=-m}^m e^{2\pi i k x}$, and let $K_y(x) = D_m(x - y)$. Then the space of trigonometric polynomials \mathcal{T}_m of degree at most m , equipped with the inner product

$$\langle p, q \rangle = (2m + 1) \int_0^1 p(x) \overline{q(x)} dx$$

between p and q in \mathcal{T}_m , is a reproducing kernel space. By the interpolation property of \mathcal{T}_m , any set of at most $2m + 1$ kernel functions is linearly independent, hence \mathcal{T}_m has the $(2m + 1)$ -HRT property.

Another example we study further below is a space of entire functions with significance in time-frequency analysis.

Example 2. Consider the space \mathcal{F} containing each entire function $f : \mathbb{C} \rightarrow \mathbb{C}$ obtained from a series

$$f(z) = \sum_{n=0}^{\infty} c_n \frac{z^n}{\sqrt{n!}}$$

with $c \in \ell^2$. The map from c to $f \in \mathcal{F}$ becomes an isometry when the norm of f is given by an L^2 -norm with a Gaussian weight,

$$\|f\|^2 = \int_{\mathbb{C}} |f(x + iy)|^2 e^{-(x^2 + y^2)} \frac{dx dy}{\pi}.$$

The sequence of monomials $(e_n)_{n=0}^\infty$ with $e_n(z) = z^n/\sqrt{n!}$ is then an orthonormal basis for \mathcal{F} . When approximating any entire function f whose weighted L^2 -norm is finite by $f_m = \sum_{n=m}^\infty \langle f, e_n \rangle e_n$, its orthogonal projection onto the subspace \mathcal{F}_m of polynomials of degree m , then it can be shown with the mean value property of entire functions that $(f_m)_{m \in \mathbb{N}}$ converges uniformly on compact sets in \mathbb{C} , so each such f is in \mathcal{F} . This space is called Bargmann space [4].

From the resolution of the identity with $(e_n)_{n \in \mathbb{N}}$ and summability,

$$f(w) = \sum_{n=0}^{\infty} \langle f, e_n \rangle w^n / \sqrt{n!}$$

which shows that $f(w) = \langle f, K_w \rangle$ with the reproducing kernel $K_w(z) = \sum_{n=0}^{\infty} \frac{z^n \bar{w}^n}{n!} = e^{z\bar{w}}$. This kernel is not unit norm, but as described we can pass to the space $\tilde{\mathcal{F}}$ with kernel $\tilde{K}_w(z) = e^{-(|z|^2 + |w|^2)/2} e^{z\bar{w}}$ whose inner product is given by the Lebesgue measure on $\mathbb{C} \simeq \mathbb{R}^2$. Because of the interpolation property of polynomials, \mathcal{F} and $\tilde{\mathcal{F}}$ and the m -dimensional subspaces \mathcal{F}_m and $\tilde{\mathcal{F}}_m$ associated with polynomials of maximal degree m have the m -HRT property.

3.2 Main Sparse Recovery Result

In this section, we present our main recovery result. The setting is as follows: We are given linear measurements $(\langle f, M_i \rangle)_{i \in I}$ of a signal $f \in \mathcal{H}$, indexed by an at most countable set I and corresponding to a family of vectors $(M_i)_{i \in I}$ in \mathcal{H} , thereby ensuring satisfaction of (D2). Our standard assumption is that $(M_i)_{i \in I}$ is a Bessel family with Bessel bound 1, or equivalently, that the measurement map

$$M : \mathcal{H} \rightarrow \ell^2(I), (Mf)_i = \langle f, M_i \rangle$$

has operator norm $\|M\| = 1$. Whenever convenient, we will also replace $\ell^2(I)$ by a general separable Hilbert space \mathcal{H}' and think of the measurement as a contraction $M : \mathcal{H} \rightarrow \mathcal{H}'$. Note that the latter interpretation allows for inclusion of even uncountable sampling sets: Observing a function Mf in a reproducible kernel Hilbert space \mathcal{H}' on an uncountable domain Ω amounts to taking uncountably many samples $Mf(x), x \in \Omega$ of the function Mf .

We will assume that $f = \sum_{n=1}^s c_n K_{x_n}$, for some coefficient vector $c \in \mathbb{C}^s$. The first main result of this section is dedicated to finding conditions that ensure $\|c\|_1 = \|f\|_A$, that is, the measure $\sum_{n=1}^s c_n \delta_{x_n}$ is a minimizer of the program (\mathcal{P}) . Since it will be very convenient in the sequel, let us introduce a term for such minimizers:

Definition 5. Let \mathcal{X} and \mathcal{H} be as stated in Lemma 1, and $f \in \mathcal{H}$ with $\|f\|_A < \infty$. A minimizer μ_* of (\mathcal{P}) is called an *atomic decomposition* of f .

Note that atomic decompositions do not necessarily need to be unique – but they always exist due to Lemma 1. When there is no risk for confusion, we will also very liberally speak of $D\mu_* = \int_{\mathcal{X}} K_x d\mu_*(x)$ as being an atomic decomposition of f .

Inspired by e.g. [12], we propose the following program for recovering f from the measurement with $(M_i)_{i \in I}$:

$$\min \|g\|_A \text{ subject to } \langle f, M_i \rangle = \langle g, M_i \rangle, i \in I. \quad (\mathcal{P}_A)$$

We stress that this also following our philosophy for recovering analog signals in the sense of (D3).

We now state and prove the first main result on sparse recovery in reproducing kernel Hilbert spaces. It generalizes results by Candès and Fernandez-Granda [8] as well as Tang et al. [32], based on the use of dual certificates as in [11].

Theorem 1. Let $f \in \mathcal{H}$ be given by the linear combination

$$f = \sum_{n=1}^s c_n K_{x_n}$$

with $c \in \mathbb{C}^s$ and $x \in \mathcal{X}^s$. Assume that the closed linear span of the measurement vectors $\overline{\text{span}} \{M_i, i \in I\} = \mathcal{H}'$ contains a continuous function $\psi \in \mathcal{H} \cap \mathcal{C}_0(\mathcal{X})$ with the properties

- (i) $\|\psi\|_\infty = 1$,
- (ii) $\psi(x_j) = c_j/|c_j|$, $j = 1, \dots, m$,
- (iii) $|\psi(x)| < 1$ for $x \notin \{x_j, j = 1, \dots, m\}$,

and additionally that $(K_{x_n})_{n=1}^m$ is linearly independent. Then f is the unique solution of the program (\mathcal{P}_A) and $\mu_0 = \sum_{n=1}^s c_n \delta_{x_n}$ is an atomic decomposition of f .

Proof. Towards a contradiction, assume that there exists a $g \in \mathcal{H}$ with $\langle g, M_i \rangle = \langle f, M_i \rangle$ for all $i \in I$ but $\|g\|_A \leq \|f\|_A \leq \|c\|_1$. Let μ be an atomic decomposition of g . By continuity, $\langle f, M_i \rangle = \langle g, M_i \rangle$ for all i implies that $\langle g, \psi \rangle = \langle f, \psi \rangle$. This has the consequence

$$\begin{aligned} 0 &= \langle f - g, \psi \rangle = \langle K * (\mu_0 - \mu), \psi \rangle = \int_{\mathcal{X}} \langle K_x, \psi \rangle d(\mu_0 - \mu) = \int_{\mathcal{X}} \bar{\psi} d\mu_0 - \int_{\mathcal{X}} \bar{\psi} d\mu \\ &= \sum_{n=1}^m |c_n| - \int_{\mathcal{X}} \bar{\psi} d\mu = \|c\|_1 - \int_{\text{supp } \mu} \bar{\psi} d\mu. \end{aligned}$$

Now if there were $x \in \text{supp } \mu$ with $|\psi(x)| < 1$, there would exist a neighborhood U of x with $\mu(U) > 0$ and $|\psi(y)| < 1$ for all $y \in U$. This would have the consequence that $\|g\|_A \leq \|c\|_1 = \left| \int_{\text{supp } \mu} \bar{\psi} d\mu \right| < \|\mu\|_{TV}$, which is a contradiction to μ being an atomic decomposition of g . Hence, $\text{supp } \mu \subseteq \{x_n\}_{n=1}^m$, and the linear independence property implies that $\mu = \mu^0$, in particular $f = g$. \square

The problem addressed in this work is whether a signal in a space of trigonometric polynomials that satisfies a sparsity pattern requirement can be recovered from projections onto lower-dimensional subspaces. The subspaces are chosen to consist of low-frequency content so that the measurement observes a low-pass version of the signal, and the recovery restores the signal to its original level of resolution.

Example 3. Consider the space of trigonometric functions presented in Example 1. We wish to recover a sufficiently sparse signal $f \in \mathcal{F}_N$ from a set of measurements with $\{M_j\}_{j=1}^{2m+1}$ with $M_j(t) = e^{2\pi i j t}$ and $m \leq N$.

As shown by [8, Proposition 2.1], if $m \geq 128$, and $T = \{t_1, t_2, \dots, t_s\}$ is a sequence in $[0, 1)$ such that $\Delta(T) = \min_{i \neq j} d(t_i, t_j) \geq 2/m$, then for any $w \in \mathbb{T}^s$, there exists $\psi \in \mathcal{F}_m$ such that $\psi(t_j) = w_j$ for each $j \in \{1, 2, \dots, s\}$ and $|h(t)| < 1$ if $t \notin T$. Consequently, by the preceding theorem, given any $f \in \mathcal{F}_N$ such that

$$f = \sum_{n=1}^s c_n K_{t_n}$$

and $T = \{t_j\}_{j=1}^s$ satisfies the separation condition for $\Delta(T)$, then $c \in \mathbb{T}^s$, $t \in [0, 1)^s$ and f are uniquely characterized by the projection of f onto \mathcal{F}_m and by the identity

$$\|f\|_A = \sum_{n=1}^s |c_n|.$$

3.3 Stability and Robustness

In this section we study the effect of noisy measurements and the question whether accurate recovery can be achieved even if the model assumption of sparsity is only approximately satisfied.

Let M be the measurement operator on \mathcal{H} , associated with the Bessel family $(M_i)_{i \in I}$ as described in the preceding section. Suppose that we are given *contaminated* measurements $b = Mf + n$, of an element $f \in \mathcal{H}$ with an *approximately* sparse atomic decomposition

$$f = \sum_{n=1}^s c_n K_{x_n} + K * \mu_c,$$

in $(K_x)_{x \in \mathcal{X}}$. The complex measure μ_c is hereby arbitrary in $\mathcal{M}(\mathcal{X})$, and intuitively should be thought of as a small ‘non-sparse’ part of f . We wish to find conditions under which the following problem approximately recovers f :

$$\min \|g\|_A \quad \text{subject to} \quad \|Mg - b\| \leq \epsilon. \quad (\mathcal{P}_A^\epsilon)$$

For ℓ_1 -minimization for recovery of sparse signals $v_0 \in \mathbb{R}^n$, theorems like Theorem 1 often generalize to statements about approximate recovery by programs like (\mathcal{P}_A^ϵ) . A typical statement would be that the solution x^* of a regularized version ℓ_1 -minimization

$$\min \|x\|_1 \quad \text{subject to} \quad \|Mx - b\|_2 < \epsilon$$

obeys

$$\|x^* - x_0\|_1 \leq C_1 \cdot \left(\min_{c \text{ } s\text{-sparse}} \|x_0 - c\|_1 \right) + C_2 \epsilon,$$

for constants $C_1, C_2 \in \mathbb{R}$ (see, for instance, [21, 6.12]). The direct analogue of that result in our setting would be obtained by replacing the ℓ_1 -norm with the total variation norm, x^*

with μ_* , x_0 with μ_0 and c with μ , where the minimization is over μ whose support is of size s . It is unfortunately not possible to prove a statement like this in our setting. To see this, note that often the map $x \mapsto K_x$ is continuous. (In the two cases considered in Section 4, this is the case.) Therefore, $\|MK_{x'} - MK_x\|_2$ is small for x close to x' . Hence, for certain contaminations n , a multiple of $K_{x'}$ will be a solution of \mathcal{P}_A^ϵ – but $\|\delta_x - \delta_{x'}\|_{TV} = 2$ for all $x \neq x'$.

In [20], the author perform a very precise analysis of the structure of the solution of the *Beurling LASSO*, given by

$$\min_{\mu \in \mathcal{M}(\mathbb{T})} \frac{1}{2} \|\Phi\mu - b\|_2^2 + \lambda \|\mu\|_{TV},$$

for the case of $\Phi : \mathcal{M}(\mathbb{T}) \rightarrow L^2([0, 2\pi])$ being a smooth convolution operator, i.e.,

$$\Phi\mu(t) = \int_0^{2\pi} \chi(t-x)d\mu(x) \quad \text{with } \chi \in \mathcal{C}^2(\mathbb{T}).$$

\mathbb{T} is thereby the torus $\mathbb{T} \simeq \mathbb{R}/(2\pi\mathbb{Z})$. The authors prove that for small noise levels ϵ and small values of λ , under certain dual certificate conditions partly resembling the ones we have discussed above, the solutions μ_* of the Beurling LASSO are concentrated on sets close to the support of the ground truth solution μ_0 . In the particular case of a train of s δ -spikes as ground truth measure, the minimizer of the Beurling Lasso will also be a train of s δ -spikes, whose positions and amplitudes are close to the ones of μ_0 .

To prove this is certainly an achievement, but the special form of the measurement operator is explicitly used several times in the arguments. In particular, some of their conditions involve both the first and second derivative of the dual certificate (the function ψ in Theorem 1), which does not need to exist for more general measurement operators.

We will prove a result which is of similar flavor, although by far not as precise, as the ones of [20] for (\mathcal{P}_A^ϵ) . Intuitively, it states that for ground truth signals having an atomic decomposition with small, finite support, most of the mass of the atomic decompositions of the minimizer of \mathcal{P}_A^ϵ will be concentrated in a set close to the support of an atomic measure.

We introduce some notation to make the notion of concentration quantitative. For a sequence $(x_i)_{i \in \mathbb{N}}$ in the metric space \mathcal{X} and $\delta > 0$, we define $S_\delta = \cup_{i \in \mathbb{N}} U_\delta(x_i)$ where $U_\delta(x)$ is the open ball of radius δ centered at x . Given a reproducing kernel Hilbert space \mathcal{H} and a contraction $M : \mathcal{H} \rightarrow \mathcal{H}'$ to another Hilbert space \mathcal{H}' , for a sequence $(\omega_j)_{j \in \mathbb{N}}$ and $\lambda, \delta > 0$, we define the following set of interpolating measurements:

$$\Theta_{x,\omega,\lambda,\delta} = \{\nu \in \mathcal{H}' : M^*\nu \in \mathcal{H} \cap \mathcal{C}_0(\mathcal{X}) \tag{8}$$

$$M^*\nu(x_j) = \omega_j, j \in \mathbb{N} \tag{9}$$

$$\|M^*\nu\|_\infty \leq 1, \tag{10}$$

$$|M^*\nu(x)| \leq \lambda \text{ for } x \notin S_\delta \}. \tag{11}$$

This allows us to derive a stability result of our sparse recovery approach for analog signals when those signals are affected by noise.

Theorem 2. Let \mathcal{X} and \mathcal{H} be as in Proposition 2 and $M : \mathcal{H} \rightarrow \mathcal{H}'$ a contraction. Let $f \in \mathcal{H}$ have the form

$$f = \sum_{i \in \mathbb{N}} c_i K_{x_i} + K * \mu_c$$

and let f_* be a solution of the program,¹ given by

$$\min \|g\|_A \text{ subject to } \|b - Mg\| \leq \epsilon, \quad (\mathcal{P}_\epsilon)$$

with $\|b - Mf\| \leq \epsilon$. Further, define μ_* through $f_* = D\mu_*$, $\|f_*\|_A = \|\mu_*\|_{TV}$, and for $\delta, \lambda > 0$, set

$$C(\lambda, \delta) := \sup_{|\omega_j|=1} \inf\{\|\nu\| : \nu \in \Theta_{x, \omega, \lambda, \delta}\}.$$

Then we have

$$|\mu_*|(S_\delta) \geq \|f\|_A - \frac{2C(\lambda, \delta)\epsilon + \|\mu_c\|_{TV}}{1 - \lambda}. \quad (12)$$

Proof. We first note that f is included in the set over which the atomic norm is minimized, so the minimizer μ_* satisfies $\|\mu_*\|_{TV} = \|f_*\|_A \leq \|f\|_A$. Furthermore, combining this with the triangle inequality gives

$$\|\mu_*\|_{TV} \leq \|f\|_A \leq \|c\|_1 + \|\mu_c\|_{TV}.$$

Denote $\mu_0 = \sum_{i \in \mathbb{N}} c_i \delta_{x_i}$ and pick $\omega_j = \frac{c_j}{|c_j|}$. If there exists no ν fulfilling the constraint, there is nothing to prove. So let ν fulfill (8)-(11) and $\psi = M^*\nu$. Due to (9) and the triangle inequality, $\|f\|_A \leq \langle K * \mu_0, \psi \rangle + \|\mu_c\|_{TV}$. Also, $\langle K * \mu_0, \psi \rangle = \langle K * \mu_0, M^*\nu \rangle = \langle Mf, \nu \rangle$. This implies

$$\begin{aligned} \|f\|_A &\leq \langle K * \mu_0, \psi \rangle + \|\mu_c\|_{TV} = |\langle Mf - Mf_*, \nu \rangle| + |\langle K * \mu_*, \psi \rangle| + \|\mu_c\|_{TV} \\ &\leq 2\|\nu\| \epsilon + \left| \int_{\mathcal{X}} \overline{\psi(x)} d\mu_*(x) \right| + \|\mu_c\|_{TV}, \end{aligned}$$

where the last inequality follows from the constraint of \mathcal{P}_ϵ , as well as the fact that $\psi \in \mathcal{H} \cap \mathcal{C}_0(\mathcal{X})$. Now let us take a closer look at the integral. Splitting the domain of integration according to

$$\int_{\mathcal{X}} \overline{\psi} d\mu_* = \int_{S_\delta} \overline{\psi} d\mu_* + \int_{\mathcal{X} \setminus S_\delta} \overline{\psi} d\mu_*$$

and estimating ψ by (10) and (11) gives

$$\begin{aligned} \left| \int_{\mathcal{X}} \overline{\psi} d\mu_* \right| &\leq |\mu_*|(S_\delta) + \lambda |\mu_*|(\mathcal{X} \setminus S_\delta) \\ &\leq \lambda \|f_*\|_A + (1 - \lambda) |\mu_*|(S_\delta). \end{aligned}$$

¹Such a solution always exists: the set $\{\mu \mid \|Mf - DM\mu\| \leq \epsilon, \|\mu\|_{TV} \leq 2\|f\|_A\}$ is convex, bounded and closed, thus weak-*compact.

Next, we estimate $\|f_*\|_A \leq \|f\|_A$. Inserted in the preceding estimate for $\|f\|_A$, this yields

$$\|f\|_A \leq 2\|\nu\|\epsilon + \lambda\|f\|_A + (1-\lambda)|\mu_*|(S_\delta) + \|\mu_c\|_{TV}.$$

Rearranging, we arrive at

$$|\mu_*|(S_\delta) \geq \|f\|_A - \frac{2\|\nu\|\epsilon + \|\mu_c\|_{TV}}{1-\lambda}$$

The definition of $C(\lambda, \delta)$ yields the claim. \square

Looking at the proof again, we see that it actually proves that if μ is a measure satisfying $\|\mu\|_{TV} \leq \|f\|_A$ and $\|b - MD\mu\| \leq \epsilon$, the existence of the dual certificate ψ implies the estimate (12). This fact allows us to prove the following ‘‘interpolation result’’ (which is arguably more practically relevant than the above.)

Theorem 3. Let \mathcal{X} , K and \mathcal{H} be as in Proposition 2. Assume that the map $\mathcal{X} \rightarrow \mathcal{H}$, $x \mapsto K_x$ is continuous. Let

$$f = \sum_{i=1}^s c_i K_{x_i} + K * \mu_c$$

and let $(\mu^n) \subseteq \Delta$ (Δ still denotes the set of finitely supported measures) be a minimizing sequence of the program

$$\inf_{\mu \in \Delta} \|\mu\|_{TV} \text{ subject to } \|b - MK * \mu\| \leq \epsilon, \quad (\mathcal{P}_\Delta)$$

with $\|b - Mf\| \leq \epsilon$. Then there exists a $\delta_0 > 0$ so that, for all $\delta \leq \delta_0$ satisfying the same assumptions as in Theorem 2 and for all sufficiently large n , the following error bound holds:

$$\begin{aligned} \|D\mu^n - f\| &\leq 2C(\lambda, \delta)\epsilon + (C(\lambda, \delta) + 1) \left(\epsilon\|f\|_A + \|\mu_c\|_{TV} + \frac{2C(\lambda, \delta)\epsilon + \|\mu_c\|_{TV}}{1-\lambda} \right) \\ &= C_1\epsilon + C_2\|\mu_c\|_{TV}. \end{aligned}$$

Proof. We begin by proving that the optimal value of (\mathcal{P}_Δ) is smaller than $\|f\|_A$: Due to Lemma 2, there exists a sequence of measures v_c^n with $v_c^n \xrightarrow{*} \mu_c$ and $\|v_c^n\|_{TV} \leq \|\mu_c\|_{TV}$. The continuity of D and M then implies that for $\eta \in (0, 1)$ sufficiently close to one,

$$\eta \sum_{k=1}^s c_k \delta_{x_k} + v_c^n \quad (13)$$

will eventually be feasible for (\mathcal{P}_Δ) . Hence, by $\eta < 1$ the infimum of that program is strictly smaller than $\|f\|_A$. Since (μ^n) is a minimizing sequence, μ^n will therefore eventually satisfy $\|\mu^n\|_{TV} \leq \|f\|_A$ and $\|b - MD\mu\| \leq \epsilon$. Theorem 2 together with the remark directly prior to the current theorem implies that

$$|\mu^\circ|(S_\delta) \geq \|f\|_A - \frac{2C(\lambda, \delta)\epsilon + \|\mu_c\|_{TV}}{1-\lambda} =: \|f\|_A - \epsilon',$$

where μ° denotes μ^n with n large enough. This implies

$$|\mu^\circ|(\mathcal{X} \setminus S_\delta) = \|\mu^\circ\|_{TV} - |\mu^\circ|(S_\delta) \leq \|f\|_A - (\|f\|_A - \epsilon') = \epsilon'. \quad (14)$$

Next, we define a sequence of complex measures $\bar{\mu}^\circ$ of support $\{x_1, x_2, \dots, x_s\}$ through

$$\bar{\mu}^\circ = \sum_{k=1}^s \mu^\circ(U_\delta(x_k)) \delta_{x_k}.$$

Assuming that δ is sufficiently small so that $\{U_\delta(x_k)\}_{k=1}^s$ is a sequence of disjoint sets, we then have

$$\left\| \int_{S_\delta} K_x d\mu^\circ(x) - \int_{S_\delta} K_x d\bar{\mu}^\circ(x) \right\| \leq \sum_{i=1}^s \int_{U_\delta(x_i)} \|K_{x_i} - K_x\| d\mu^\circ$$

with $S_\delta = \cup_{i=1}^s U_\delta(x_i)$. Since $x \mapsto K_x$ is continuous, there exists a δ_0 such that for all $\delta \leq \delta_0$, $\|K_{x_i} - K_x\| \leq \epsilon$ for all i and $x \in U_\delta(x_i)$. This, and the fact that $\|\nu_c^n\|_{TV} \leq \|\mu_c\|_{TV}$ for all n , implies that the above expression can be estimated by

$$\sum_{i=1}^s \int_{U_\delta(x_i)} \|K_{x_i} - K_x\| d\mu^\circ \leq \epsilon \mu^\circ(S_\delta). \quad (15)$$

Inequalities (14) and (15) together imply

$$\begin{aligned} \|K * (\mu^\circ - \bar{\mu}^\circ)\| &\leq \left\| \int_{S_\delta} K_x d\mu^\circ(x) - \int_{S_\delta} K_x d\bar{\mu}^\circ(x) \right\| + \left\| \int_{\mathcal{X} \setminus S_\delta} K_x d\mu^\circ(x) \right\| \\ &\leq \epsilon \|f\|_A + |\mu^\circ|(\mathcal{X} \setminus S_\delta) \leq \epsilon \|f\|_A + \epsilon'. \end{aligned} \quad (16)$$

Next, let ψ be a function obeying the constraints (8) - (11) with $\omega_j \in \mathbb{C}$, $|\omega_j| = 1$ such that $\bar{\omega}_j(\mu^\circ(U_\delta(x_j)) - c_j) = |\mu^\circ(U_\delta(x_j)) - c_j|$. We recall that for μ° feasible,

$$\|Mf - MK * \mu^\circ\| \leq 2\epsilon$$

so by $Mf = MK * (\mu_s + \mu_c)$, this choice of ψ gives

$$\begin{aligned} \|\bar{\mu}^\circ - \mu_s\|_{TV} &= \sum_{i=1}^s |\mu^\circ(U_\delta(x_j)) - c_j| = \langle K * (\bar{\mu}^\circ - \mu_s), \psi \rangle = \langle MK * (\bar{\mu}^\circ - \mu_s), \nu \rangle \\ &\leq \|\nu\| (\|MK * (\bar{\mu}^\circ - \mu^\circ)\| + \|MK * \mu^\circ - M(f - K * \mu_c)\|) \\ &\leq \|\nu\| (\|K * (\bar{\mu}^\circ - \mu^\circ)\| + \|MK * \mu^\circ - Mf\| + \|\mu_c\|_{TV}) \\ &\leq \|\nu\| (\epsilon \|f\|_A + \epsilon' + 2\epsilon + \|\mu_c\|_{TV}). \end{aligned} \quad (17)$$

This finally implies, together with estimate (16),

$$\begin{aligned} \|f - K * \mu^\circ\| &\leq \|f - K * \mu_s\| + \|K * \mu_s - K * \bar{\mu}^\circ\| + \|K * (\bar{\mu}^\circ - \mu^\circ)\| \\ &\leq \|\mu_c\|_{TV} + \|\bar{\mu}^\circ - \mu_s\|_{TV} + \|K * (\bar{\mu}^\circ - \mu^\circ)\| \\ &\leq \|\mu_c\|_{TV} + \|\nu\| (\epsilon \|f\|_A + \epsilon' + 2\epsilon + \|\mu_c\|) + \epsilon \|f\|_A + \epsilon' \\ &\leq 2\|\nu\| \epsilon + (\|\nu\| + 1)(\epsilon \|f\|_A + \|\mu_c\|_{TV} + \epsilon'). \end{aligned}$$

The definitions of ϵ' and $C(\lambda, \delta)$ give the claim. \square

4 Applications

We will now discuss two applications of our methodology, namely sparse recovery in the situations of bandlimited functions – as stated in the introduction the “classical” example for analog signals – and of functions with sparse expansions in the short-time Fourier transforms.

4.1 Sparse recovery for bandlimited functions

Definition 6. The space of bandlimited functions on \mathbb{R} with bandlimit $1/2$ is also known as the Paley-Wiener space $PW_{1/2}$. It is the space consisting of square integrable functions f whose Fourier transforms \hat{f} have essential support in $[-1/2, 1/2]$. In particular, they are obtained from their Fourier transform by

$$f(t) = \int_{-1/2}^{1/2} \hat{f}(\omega) e^{2\pi i \omega t} d\omega.$$

The norm on $PW_{1/2}$ is the L^2 -norm.

From the inclusion $L^2([-1/2, 1/2]) \subset L^1([-1/2, 1/2])$, it can be inferred that each $f \in PW_{1/2}$ is a continuous function. The space $PW_{1/2}$ is indeed a reproducing kernel space on $\mathcal{X} = \mathbb{R}$, with reproducing kernel

$$K_t(s) = \int_{-1/2}^{1/2} e^{2\pi i \omega(t-s)} d\omega = \frac{\sin(\pi(t-s))}{\pi(t-s)} \quad \text{for } s \neq t, \quad s, t \in \mathbb{R},$$

and $K_t(t) = 1$ for $t \in \mathbb{R}$.

Now assume that a function $f \in PW_{1/2}$ is given through $f = \sum_{j=1}^s c_j K_{t_j} + K * \mu_c$, with $t_j \in [-L/2, L/2]$ and μ_c a complex Borel measure of bounded total variation. Intuitively, f is a linear combination of a finitely supported complex measure plus a noise part μ_c , run through a low pass filter; see also Figure 4.1. We wish to recover f approximately, assuming the set of points $\{t_j\}$ is sufficiently separated. We note that if $\{t_j\} \subset [-L/2, L/2]$, then the Fourier transform of f satisfies

$$\hat{f}(\omega) = \sum_{j=1}^s c_j e^{2\pi i t_j \omega} + \hat{\mu}_c(\omega)$$

on $\omega \in (-1/2, 1/2)$.

Motivated by a result concerning recovery of trigonometric polynomials [8], it seems natural to use sampling in the frequency domain as measurements. However, point evaluations of the Fourier transform are not bounded linear functionals on $PW_{1/2}$. We will circumvent this fact by instead using “mollified” Fourier measurements: Assuming that $(2m+1)/L < 1$, then integrating the Fourier transform \hat{f} with a moving window

$$\langle f, M_k \rangle = \left(\frac{L}{2\rho} \right)^{1/2} \int_{(k-\rho)/L}^{(k+\rho)/L} \hat{f}(\omega) d\omega$$

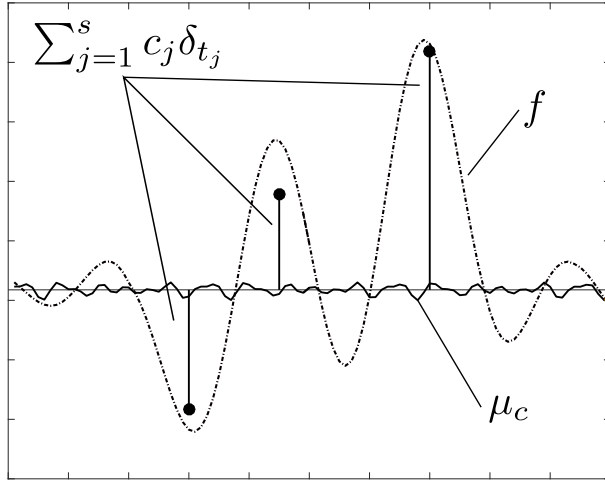


Figure 1: The function $f \in PW_{1/2}$ (dashed) is a low-pass filtered version of the sum of the finitely supported measure $\sum_{i=1}^s c_i \delta_{t_i}$ and a non-sparse part μ_c .

for $k \in \{-m, -m+1, \dots, m-1, m\}$, gives perturbed measurements of the Fourier transform of $\mu_0 = \sum_{j=1}^s c_j \delta_{t_j}$ according to

$$\langle f, M_k \rangle = \hat{f} * \hat{h}_\rho(k/L)$$

with $h_\rho(x) = \left(\frac{L}{2\rho}\right)^{1/2} \sin(\pi\rho x/L)/(\pi x)$, $\rho \leq 1/2$. We note that with this choice $\{M_k\}_{k=-m}^m$ is an orthonormal system.

It is convenient to re-encode these measured values in a trigonometric polynomial according to

$$\begin{aligned} p(y) &= \sum_{k=-m}^m \langle f, M_k \rangle e^{2\pi i k y} \\ &= \sum_{k=-m}^m \hat{f} * \hat{h}_\rho(k/L) e^{2\pi i k y} \\ &= \int_{\mathbb{R}} \sum_{k=-m}^m \exp(2\pi i k(y - x/L)) f(x) h_\rho(x) dx \\ &= (2m+1) \int_{\mathbb{R}} D_m(y - x/L) h_\rho(x) d(\mu_0 + \mu_c)(x) \end{aligned}$$

where we used the notation from Example 1. The last identity follows from the fact that $x \mapsto D_m(y - x/L) h_\rho(x)$ is in $PW_{1/2} \cap \mathcal{C}_0(\mathbb{R})$. Hence, M maps functions in $PW_{1/2}$ to the space $\mathcal{TC}_{\leq m}[x]$ of trigonometric polynomials on $[-1/2, 1/2]$ of degree at most m .

In the following, we will investigate to which extent Theorem 2 is applicable in this context. The result of the application of Theorems 2 and also 3 to this setting will be summarized in Theorem 4.

Let us start by noting that the range of the operator M^* is the space $\{x \mapsto q(x/L)h_\rho(x), q$ a trigonometric polynomial (periodically extended on \mathbb{R}) of degree at most $m\}$. This can be confirmed from the form of p and from the fact that the translates of the Dirichlet kernel D_m span the space of trigonometric polynomials of degree m . This fact will enable us to prove the following lemma.

Lemma 3. If $\Delta(T) \geq \frac{5}{m}$ and $\rho > 0$ is small enough, there exists a trigonometric polynomial q^ρ such that $g = q^\rho h_\rho$ satisfies

$$g(x_k) = 1, x_k \in T, \tag{18}$$

$$|g(x)| \leq 1 - 0.34 m^2 (x - x_k)^2 \text{ if } |x - x_k| \leq 0.16749/m, \tag{18}$$

$$|g(x)| \leq 1 - 0.34 \cdot 0.16749^2 \text{ else.} \tag{19}$$

Proof. Considering the form of the range of the measurement operator M , it is necessary to construct a trigonometric polynomial q with the properties

$$q(x_k) = \omega_k h_\rho(x_k)^{-1}, \tag{20}$$

$$q'(x_k) = -\frac{h'_\rho(x_k)}{h_\rho(x_k)^2}, \tag{21}$$

then the function $q \cdot h$ namely obeys $q(x_k)h(x_k) = \omega_k$, $\partial_x(q(x_k)h(x_k)) = 0$. To do this, we will use an ansatz of the form

$$q = \sum_{j=1}^s \alpha_j \kappa(\cdot - x_j) + \beta_j \kappa'(\cdot - x_j),$$

where $\kappa(x) = \left(\frac{\sin(\pi(m/2+1)x)}{\sin(\pi x)}\right)^4$, just as in the paper [8]. The interpolation problem then reduces to the following system of linear equations:

$$\sum_{j=1}^s \alpha_j \kappa(x_k - x_j) + \beta_j \kappa'(x_k - x_j) = \omega_k h_\rho(x_k)^{-1},$$

$$\sum_{j=1}^s \alpha_j \kappa'(x_k - x_j) + \beta_j \kappa''(x_k - x_j) = -\frac{h'_\rho(x_k)}{h_\rho(x_k)^2}.$$

The proof of Lemma 2.2 in [8] shows that the matrix

$$\kappa = \begin{bmatrix} [\kappa(x_k - x_j)]_{k,j=1}^s & [\kappa'(x_k - x_j)]_{k,j=1}^s \\ [\kappa'(x_k - x_j)]_{k,j=1}^s & [\kappa''(x_k - x_j)]_{k,j=1}^s \end{bmatrix}$$

is invertible if the minimum separation $\Delta(T)$ obeys $\Delta(T) \geq \frac{2}{m}$. Under this assumption, it hence exists for each $\rho > 0$ a trigonometric polynomial q^ρ satisfying (20).

By choosing ω_k instead of $\omega_k h_\rho(x_k)^{-1}$, and 0 instead of $-\frac{h'_\rho(x_k)}{h_\rho(x_k)^2}$, respectively, one obtains a different trigonometric polynomial. Let us call this polynomial \tilde{q} . Assuming $\Delta(T) \geq \frac{5}{m}$, then for \tilde{q} and for each $x_k \in T$, [8, Lemma 2.5] implies

$$|\tilde{q}(x)| \leq 1 - .3354m^2(x - x_k)^2 \text{ for } |x - x_k| < 0.16749/m.$$

If $|x - x_k| \geq 0.16749/m$ for all k , we further have $|\tilde{q}(x)| < 1 - 0.3354m^2 \cdot 0.16749^2$.

If we prove that $q^\rho h_\rho \rightarrow \tilde{q}$ in $\mathcal{C}^2([-\pi, \pi])$ for $\rho \rightarrow 0$, then the inequalities (18)–(19) will eventually be induced by the corresponding ones for \tilde{q} . This is clear for (19), and as for (18), a Taylor expansion around x_k implies

$$\begin{aligned} |\tilde{q}(x) - q^\rho h_\rho(x)| &\leq |\tilde{q}(x_k) + \tilde{q}'(x_k)(x - x_k) - q^\rho(x_k)h_\rho(x_k) - \partial_x(q^\rho h_\rho)(x_k)| \\ &\quad + \sup_{y \in [x_k, x]} |\tilde{q}''(y) - \partial_x(q^\rho h_\rho)(y)| \frac{(x - x_k)^2}{2} \\ &\leq \|\tilde{q} - \partial_x(q^\rho h_\rho)\|_{\mathcal{C}^2} \frac{(x - x_k)^2}{2} \end{aligned}$$

First, we have $h_\rho(x) = \sqrt{\rho/2L} \operatorname{sinc}(\rho x/L)$. Consequently, $h'_\rho(x) = \sqrt{\rho/2L} \cdot \rho/L\pi \operatorname{sinc}'(\rho x/L)$ and $h''_\rho(x) = \sqrt{\rho/2L} \cdot \rho^2/L^2\pi^2 \operatorname{sinc}''(\rho x/L)$. Therefore,

$$\begin{aligned} \sqrt{2L/\rho} h_\rho &\rightarrow 1 \\ \sqrt{2L/\rho} \cdot L/\rho h'_\rho(x) &\rightarrow 0 \\ \sqrt{2L/\rho} \cdot L^2/\rho^2 h''_\rho(x) &\rightarrow \frac{\pi^2}{6}, \end{aligned}$$

where the convergence is uniform on any compact interval. This actually proves that $\sqrt{2L/\rho} h_\rho \rightarrow 1$ in $\mathcal{C}^2([-R, R])$ – the extra ρ 's in front prevents the second derivative to converge to zero – removing them makes it do that.

Now we notice that the trigonometric polynomial $q_\Delta = (\rho/2L)^{1/2} q^\rho - \tilde{q}$ solves the interpolation problem

$$\begin{aligned} q_\Delta(x_k) &= \omega_k((\rho/2L)^{1/2} h_\rho^{-1}(x_k) - 1) \\ q_\Delta(x_k) &= -\frac{\rho^{1/2} h'_\rho(x_k)}{(2L)^{1/2} h_\rho(x_k)^2}. \end{aligned}$$

Due to construction, q_Δ is linearly dependent on the finitely many values $w = (q_\Delta(x_k), q'_\Delta(x_k))$. Hence, if we prove that $w \rightarrow 0$ as $\rho \rightarrow 0$, $q_\Delta \rightarrow 0$. The latter convergence is furthermore true in any norm on $\mathcal{TC}_{\leq m}[x]$, in particular the $\mathcal{C}^2[-1, 1]$ -norm, since all norms on finite dimensional spaces are equivalent. But $((\rho/2L)^{1/2} h_\rho^{-1}(x_k) - 1) \rightarrow 0$ is a consequence of $(2L/\rho)^{1/2} h_\rho \rightarrow 1$, and

$$\lim_{\rho \rightarrow 0} \sqrt{\frac{\rho}{2L}} \frac{h'_\rho(x_k)}{h_\rho^2(x_k)} = \lim_{\rho \rightarrow 0} \sqrt{\frac{\rho}{2L}} \frac{\sqrt{\rho/2L} \pi \rho/L \operatorname{sinc}'(\rho x_k/L)}{\rho/2L \cdot \operatorname{sinc}^2(\rho x_k/L)} = \lim_{\rho \rightarrow 0} \frac{\pi \rho \operatorname{sinc}'(\rho x_k/L)}{L \operatorname{sinc}^2(\rho x_k/L)} = 0.$$

We conclude that $(\rho/2L)^{1/2} q^\rho \rightarrow \tilde{q}$ and $(2L/\rho)^{1/2} h_\rho \rightarrow 1$ in $\mathcal{C}^2([-1, 1])$. Consequently, $q^\rho h_\rho \rightarrow \tilde{q}$ in the same sense, and the claim has been proven. \square

We now assume that we observe a noisy version of Mf , i.e. $Mf + v$, $\|v\|_2 \leq \epsilon$ (this is in particular the case if we assume that $v(y) = \sum_{k=-m}^m \lambda_k \exp(ik\pi y)$ with $\|\lambda\|_2^2 \leq \epsilon$, i.e. that we build Mf via noisy Fourier samples $\langle f, \exp(ik\pi(\cdot)) \rangle + \lambda_k$. Using the previous lemma, we can derive the following result.

Theorem 4. Assume that $\Delta(T) \geq \frac{5}{m}$. Then there exists a $\rho_0 > 0$ so that for every $\rho \leq \rho_0$, the following is true for the solution μ_* of \mathcal{P}_A

$$|\mu_*| (S_{\min(\delta, .16749/m)}) \geq \|f\|_A - \frac{2\sqrt{\frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2} \epsilon + \|\mu_c\|_{TV}}}{1 - .34 \cdot \min(\delta, .16749)^2}, \quad (22)$$

where S_δ is as in Theorem 2. Furthermore, a minimizing sequence (f_n) for the TV -minimization problem of the form $\sum_{t \in \Delta_n} c_t K_t$ will eventually satisfy

$$\begin{aligned} \|f_n - f\| &\leq \sqrt{\frac{L}{2\rho \operatorname{sinc}(\rho/2L)^2} \epsilon} \\ &+ \left(\sqrt{\frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2} + 1} \right) \left(\|f\|_A \epsilon + \|\mu_c\|_{TV} + \frac{\sqrt{\frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2} \epsilon + \|\mu_c\|_{TV}}}{1 - \min(\delta, .16749)^2} \right). \end{aligned}$$

Proof. By construction, $\psi = q^\rho(\cdot/L)h_\rho$, where q is the trigonometric polynomial constructed in Lemma 3 can be used as ψ in the stability theorem. Since for $\nu \in \mathcal{TC}_{\leq m}[x]$

$$M^* f(x) = h_\rho(x) \int_{-1/2}^{1/2} D_m(y - x/L) \nu(y) dy = h_\rho(x) \nu(x/L),$$

we see that the ν -function in the stability theorem can be identified with q^ρ .

Since $\|q^\rho h_\rho\|_\infty \leq 1$, we can conclude

$$\|\nu\|_2^2 = \|q^\rho\|_2^2 \leq \int_{-1/2}^{1/2} h_\rho^{-2}(x) dx = \frac{2L}{\rho} \int_{-1/2}^{1/2} \operatorname{sinc}(\rho x/L)^{-2} dx \leq \frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2}$$

Lemma 3 (in particular the bounds (18) and (19)) now implies that the constant $C(\lambda, \delta)$ in Theorem 2 satisfies

$$\begin{aligned} C(0.34 \cdot 0.16749^2, L \cdot 0.16749/m) &\leq \sqrt{\frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2}}, \\ C(0.34 \cdot m^2 \delta^2, L \cdot 0.16749/m) &\leq \sqrt{\frac{2L}{\rho \operatorname{sinc}(\rho/2L)^2}}, \quad \delta \leq 0.3298/m. \end{aligned}$$

(Note that Lemma 3 is to be applied for the points x_k/L !) Applying Theorems 2 and 3 now immediately yields the statements of the theorem. \square

Remark 1. Note that the constant in front of ϵ in (22) tends to infinity as $\rho \rightarrow 0$. On the other hand, $2/(1 - 0.34 \cdot 0.16749^2) \leq 1.00963$, so the constant in front of $\|\mu_c\|_{TV}$ is quite moderate.

4.2 Gaussian Window Radar Measurements.

Next, we consider another model for sparsity, which is motivated by an application to Doppler radar. Very rudimentarily explained, a radar device functions as follows: A waveform h , say in $L^2(\mathbb{R})$, is sent out from the radar antenna. This waveform is reflected by different objects, and is then measured by a receiving antenna as it returns. The transmission and reflection cause the signal to lose power and shift phase, and since the reflection happens at a distance from the device, it returns with a delay. In addition, if the reflecting object is moving, the received signal is shifted in frequency. Assuming that the reflection takes place at distinct point-like objects, the signal r acquired by the receiving antenna is a linear combination of time-frequency shifted copies of the signal h ,

$$r(t) = \sum_{i=k}^s r_k h(t - \tau_k) e^{2\pi i \omega_k t}.$$

$r_k \in \mathbb{R}$ are reflection coefficients modelling the loss of power, $\tau_k \in \mathbb{R}$ are the delay and ω_k the frequency shifts.

This problem was treated in [25]. By assuming that h is a trigonometric polynomial, the authors were able to rewrite the problem as a super resolution task from 2D Fourier measurements. This strategy allow the authors to apply techniques from [8, 32] to come up with a recovery guarantee.

Here, we will instead assume that the *original waveform is a Gaussian*, i.e. $h(t) = e^{-\Lambda t^2}$, $\Lambda > 0$, and hence fully leave the Fourier regime. This filter is certainly not non-realistic, but it should be noted that its exact form will be essential for our arguments. The received signal can then be rewritten as follows:

$$r(t) = \sum_{k=1}^s r_k e^{-\Lambda(t-\tau_k)^2} e^{2\pi i \omega_k t}$$

The problem at hand is now to reconstruct the parameters r_k , τ_k and ω_k from the knowledge of the function r . In the following, we will show that we can map r onto a signal \tilde{r} which is a short linear combination of the kernel functions, indexed by (τ_k, w_k) , of the so called *Bargmann space*. This will allow us to apply the framework we have developed to recover (τ_k, w_k) by atomic norm minimization. Towards this goal, let us begin to describe the route from the radar measurements r to the signal in the Bargmann space \tilde{r} .

Definition 7. Let $\{T_a\}_{a \in \mathbb{R}}$ be the family of translation operators, acting on $f \in L^2(\mathbb{R})$ by $T_a f(x) = f(x - a)$ for Lebesgue-almost every $x \in \mathbb{R}$. The *short-time Fourier transform* of a function $f \in L^2(\mathbb{R})$ with a window $g \in L^2(\mathbb{R})$ is given by the function $V_g f$, defined on the time-frequency plane \mathbb{R}^2 . The transform $V_g f$ has the value

$$V_g f(t, \omega) = (f T_t \bar{g})^\wedge(\omega) = \int_{\mathbb{R}} \overline{g(s-t)} e^{-2\pi i \omega s} f(s) ds, \quad (t, \omega) \in \mathbb{R}^2.$$

Henceforth, we exclusively use L^2 -normalized Gaussian windows, $g_\Lambda(s) = (\frac{2}{\pi})^{1/4} \Lambda^{1/2} e^{-\Lambda^2 s^2}$, and replace the subscript g_Λ with the parameter $\Lambda > 0$,

$$V_\Lambda f(t, \omega) = (\frac{2}{\pi})^{1/4} \Lambda^{1/2} \int_{\mathbb{R}} e^{-\Lambda^2(s-t)^2 - 2\pi i \omega s} f(s) ds, \quad (t, \omega) \in \mathbb{R}^2.$$

Using the isometry of the Fourier transform together with the normalization assumption on g , one easily proves that V_Λ is an isometry and consequently maps $L^2(\mathbb{R})$ to a closed subspace of $L^2(\mathbb{R}^2)$. This space is a reproducing kernel Hilbert space, as the following lemma shows.

Lemma 4. The range of V_Λ is a reproducing kernel Hilbert space with a unit-norm reproducing kernel given by

$$K_{t,\omega}(t',\omega') = e^{-\Lambda^2(t-t')^2/2 - \pi^2(\omega-\omega')^2/2\Lambda^2} e^{i\pi(t+t')(\omega-\omega')}.$$

Proof. By definition of V_Λ , $V_\Lambda f(t,\omega) = \langle f, M_\omega T_t g_\Lambda \rangle$ hence with the isometry property, $K_{t,\omega} = V_\Lambda M_\omega T_t g_\Lambda$ is the claimed reproducing kernel. Moreover, by the normalization of g_Λ , it is a unit-norm kernel. The explicit expression for the kernel follows from $K_{t,\omega}(t',\omega') = \langle M_\omega T_t g_\Lambda, M_{\omega'} T_{t'} g_\Lambda \rangle$. \square

We also define a transform A_Λ , which maps f to an analytic function, by

$$A_\Lambda f(z) = \left(\frac{2}{\pi}\right)^{1/4} \Lambda^{1/2} e^{z^2/2} \int_{\mathbb{R}} e^{-(z-\Lambda s)^2} f(s) ds, z \in \mathbb{C}.$$

To establish the relation between V_Λ and A_Λ , we map a point (t,ω) in the time-frequency plane to $z = \Lambda t - i\pi\omega/\Lambda$ and verify

$$A_\Lambda f(z) = e^{iz_1 z_2} e^{(z_1^2 + z_2^2)/2} V_\Lambda f(z_1/\Lambda, z_2\Lambda/\pi).$$

Hence, we can interpret A_Λ as a short-time Fourier transform in complex coordinates with a change of measure. The range of this transform is Bargmann space [3, 4] from Example 2.

Lemma 5. The map A_Λ is an isometry with the space \mathcal{F} as its range.

Proof. The map A_Λ is seen to be an isometry from the relationship between A_Λ and V_Λ , the change of measure $\frac{dz_1 dz_2}{\pi} = dt d\omega$ and from the fact that V_Λ is an isometry. The fact that the space \mathcal{F} is the range of A_Λ follows from the fact that each monomial is in the range of A_Λ . Monomials can be obtained by repeatedly differentiating the kernel function K_w with respect to w and then setting $w = 0$. This is admissible because the vector-valued function $w \mapsto K_w = \sum_{n=0}^{\infty} \frac{\bar{w}^n}{\sqrt{n!}} e_n$ is arbitrarily often differentiable. \square

Lemma 6. If $h = g_\Lambda$ and r is a finite linear combination as described, then $A_\Lambda r$ is a linear combination of kernel functions with indices located at $z_k = \Lambda\tau_k - i\pi\omega_k/\Lambda$.

Proof. On the space \mathcal{F} , the time-frequency translations act unitarily by

$$A_\Lambda T_t f(z) = e^{-\frac{\Lambda}{2}t^2} e^{tz} f(z - \Lambda t)$$

and

$$A_\Lambda M_\omega f(z) = e^{-\frac{\pi^2}{2\Lambda^2}\omega^2} e^{i\pi\omega z/\Lambda} f(z - i\pi\omega/\Lambda)$$

\square

By Lemma 4, $V_\Lambda r$ is an element of the reproducing kernel Hilbert space

$$\mathcal{H} = \{g : g(z) = e^{-|z|^2/2} f(z), f \in \mathcal{F}\}$$

with unit-norm reproducing kernel $K(z, w) = e^{-(|z|^2+|w|^2)/2} e^{z\bar{w}}$. Hence, by choosing the measurement operator M equal to the identity, we are in the setting of our main results with $\mathcal{H}' = \mathcal{H}$. We write $\{\eta_w\}_{w \in \mathbb{C}}$ for the normalized family of kernel functions,

$$\eta_w(z) = e^{-\frac{1}{2}|w|^2} e^{\bar{w}z},$$

so $\langle f, \eta_w \rangle = e^{-\frac{1}{2}|w|^2} f(w)$.

Remark 2. Note that although it may seem that our setting is related to [2], the two problems are in fact quite different: In [2], the problem is to reconstruct a measure defined on \mathbb{R} from short time Fourier measurements. We on the other hand are essentially assuming that we are given the *adjoint* of the short time Fourier transform applied to a measure defined on \mathbb{R}^2 .

Let us remark that \mathcal{H} is more regular than a standard reproducing kernel Hilbert space, in the sense that not only the evaluation of a function itself, but also its derivatives, is continuous on \mathcal{H} . The following proposition is a consequence of standard calculations:

Lemma 7. (i) For $w, z \in \mathbb{C}$, consider the function

$$d\eta_w(z) = e^{-\frac{|w|^2+|z|^2}{2}} e^{\bar{w}z} (z - w) \in \mathcal{H}$$

Then for $g \in \mathcal{H}$, we have

$$\langle d\eta_w, g \rangle = g'(w)$$

(ii) Define the function $d^2\eta_w \in \mathcal{H}$ through

$$d^2\eta_w(z) = e^{-\frac{|w|^2+|z|^2}{2}} e^{\bar{w}z} (z - w)^2$$

Then, for $g \in \mathcal{H}$, the Hessian H_w (in Wirtinger notation— see for instance²) [27] of the function $z \mapsto |g(z)|^2$ is given by

$$H_w(|g|^2) = \begin{bmatrix} |\langle d\eta_w, g \rangle|^2 - |\langle K_w, g \rangle|^2 & \overline{\langle K_w, g \rangle} \langle d^2\eta_w, g \rangle \\ \overline{\langle d^2\eta_w, g \rangle} \langle K_w, g \rangle & |\langle d\eta_w, g \rangle|^2 - |\langle K_w, g \rangle|^2 \end{bmatrix}$$

Now we are in a position to prove the main theorem of this section. The technique will be similar to above: towards applying Theorem 1, we will construct a function $g \in \text{ran } V_\Lambda = \mathcal{H}$ with $g(w_k) = u_k$ and $|g(w)| < 1$ for $w \in \mathbb{C} \setminus \{w_k, k = 1, \dots, s\}$, for unimodular values $(u_k)_{k=1}^s$.

²The notation used here differs slightly from the form given in [27], since the complex gradient is defined slightly different there. We choose this form to not avoid unnecessary complications.

If such a function exists, it certainly satisfies $\frac{d}{dw} |g(w_k)|^2 = 0$ for $k = 1, \dots, s$. Our strategy will therefore be to try and find a function $g \in \mathcal{H}$ which satisfies

$$\begin{cases} g(w_k) & = u_k \\ \frac{d}{dw} |g(w_k)|^2 & = 0 \end{cases}, \quad k = 1, \dots, s \quad (23)$$

and subsequently check under which conditions this function really obeys $\|g\|_\infty = 1$. (Compare [8] and [20, Section 4].) The idea of the proof will be similar to the mentioned sources, i.e. to make an ansatz, conclude that the conditions (23) is equivalent to a system of linear equations, prove that that is solvable, and then finally check that $|g| \leq 1$ for the resulting function. Under a few conditions on the set $W = \{w_k \mid k = 1, \dots, s\}$, this strategy will be successful.

The formal theorem and proof are as follows:

Theorem 5. Let $R \geq 1$, $\Omega = B_R(0) \subseteq \mathbb{C}$, $W = \{w_k \mid k = 1, \dots, s\} \subseteq \mathbb{C}$ be a set of time-frequency shifts and $f_0 \in \mathcal{F}$ a signal of the form

$$f_0 = \sum_{w_k \in W} c_k K_{w_k}.$$

Define the quantities

$$\begin{aligned} \bar{\sigma}(W, z) &= \sum_{j=1}^s e^{-\frac{1}{2}|z-w_j|^2} (1 + |z-w_j| + |z-z_j|^2 + |z-w_j|^3) \\ \bar{\sigma}_\downarrow(W) &= \sup_i \bar{\sigma}(W, w_i), \quad \bar{\sigma}_\uparrow(W) = \sup_{z \in \mathbb{C}} \bar{\sigma}(W, z). \end{aligned}$$

Suppose that there exists a $\tau > 0$ for which

$$\sup \left\{ \sum_{j=1}^s e^{-\frac{1}{2}|z-w_j|^2} \mid \text{dist}(z, W) > \frac{\sqrt{5}-1}{4\bar{\sigma}(W)_\uparrow} \right\} < 1 - \tau \quad (24)$$

Then, the measure $\sum_{w_k \in W} c_k \delta_{w_k}$ (and therefore also f_0) can be recovered by solving the problem

$$\min_{\mu \in \mathcal{M}(B_R(0))} \|\mu\|_{TV} \text{ subject to } \langle K * \mu, m_n \rangle = \langle f_0, M_n \rangle, n = 1, \dots, N.$$

Proof. Let $\epsilon \in (0, \frac{1}{4})$ be a parameter for which we will specify the value later. Using the notation introduced above, consider the ansatz

$$g = \sum_{w_k \in W} \alpha_k \eta_{w_k} + \sum_{w_k \in W} \beta_k d\eta_{w_k}.$$

Lemma 7 implies that the conditions $g(w_k) = \text{sgn}(c_k)$ and $g'(w_k) = 0$ for each k then corresponds to the system of equations

$$\begin{bmatrix} \langle \eta_{w_j}, \eta_{w_k} \rangle_{j,k=1}^s & \langle \eta_{w_j}, d\eta_{w_k} \rangle_{k,j=1}^s \\ \langle d\eta_{w_j}, d\eta_{w_k} \rangle_{k,j=1}^s & \langle d\eta_{w_j}, \eta_{w_k} \rangle_{k,j=1}^s \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} (\text{sgn } c_k)_{k=1}^s \\ 0 \end{bmatrix}. \quad (25)$$

Let us define the matrix in the above equation as $\Phi(W)$. Noticing that $\langle \eta_w, \eta_w \rangle = \langle d\eta_w, d\eta_w \rangle = 1$, and that the off-diagonal entries are under control, we see that $\Phi(W)$ is close to the identity. Concretely, we have, using the explicit forms of the functions η_w and $d\eta_w$,

$$\begin{aligned} \|\Phi(W) - \text{id}_{2s}\|_\infty &\leq \sup_i \sup_{k=0,1} \sum_{j \neq i} \left| \left\langle d^{(k)} K_{w_j}, K_{w_i} \right\rangle \right| + \left| \left\langle d^{(k)} K_{w_j}, dK_{w_i} \right\rangle \right| \\ &\leq \sup_i \sum_{j \neq i} e^{-\frac{|w_i - w_j|^2}{2}} (1 + |w_i - w_j| + |w_i - w_j|^2) \leq \bar{\sigma}_\downarrow(W) - 1 \leq \epsilon, \end{aligned}$$

where we invoked the assumption that $\bar{\sigma}_\downarrow(W) - 1$ is small. Now by a Neumann series argument (compare [8]), we see that there will exist α and β for which (23) is satisfied, and additionally,

$$|\alpha_k - 1|, |\beta_k| \leq \frac{2\epsilon}{1 - 2\epsilon} \quad (26)$$

for each k .

Now we will check if the choice of the function g which is induced by the above choice of α and β satisfies $|g|^2 < 1$ for all $z \notin W$. We will proceed in two steps: First, we will prove that for z with $\text{dist}(z, W) \leq \tilde{\delta} H_z(|g|^2)$ will be negative definite. Then we will prove that for z with $\text{dist}(z, W) \geq \tilde{\delta}$, $|\langle g, K_z \rangle| < 1$. $\tilde{\delta}$ is thereby another parameter whose value we will specify later.

Let us begin by investigating $H_z(|g|^2)$ for z close to a w_j . Towards this goal, let us first note that $\langle K_{w_j}, g \rangle = \text{sgn } c_j$, $\langle dK_{w_j}, g \rangle = 0$ implies that

$$\begin{aligned} |\langle d\eta_z, g \rangle|^2 - |\langle \eta_z, g \rangle|^2 &= |\langle d\eta_z - d\eta_{w_j}, g \rangle|^2 - |\langle \eta_z - \eta_{w_j}, g \rangle|^2 \\ &\quad - 2 \text{Re}(\overline{\text{sgn } c_j} \langle \eta_z - \eta_{w_j}, g \rangle) - 1 \end{aligned} \quad (27)$$

$$\overline{\langle \eta_z, g \rangle} \langle d^2 \eta_z, g \rangle = \overline{\text{sgn } c_j} \langle d^2 \eta_z, g \rangle + \overline{\langle \eta_z - \eta_{w_j}, g \rangle} \langle d^2 \eta_z, g \rangle. \quad (28)$$

Hence, if we can prove that $|\langle d^{(k)} \eta_z - d^{(k)} \eta_{w_j}, g \rangle|$ for $k = 0, 1$ as well as $\langle d^2 \eta_z, g \rangle$ is close to zero for z close to w_j , Lemma 7 proves that dH_z is close to $-\text{id}_{2s}$ for z close to w_j , and hence negative definite.

Let us first note that for w, z arbitrary and $\ell = 0, 1, 2$, we have

$$\left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, g \right\rangle \right| \leq \sum_{j=1}^s |\alpha_j| \left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, \eta_{w_j} \right\rangle \right| + \sum_{j=1}^s |\beta_j| \left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, d\eta_{w_j} \right\rangle \right|.$$

Remembering (26), we obtain

$$\begin{aligned} \left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, g \right\rangle \right| &\leq \frac{1}{1 - 2\epsilon} \sum_{j=1}^s \left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, \eta_{w_j} \right\rangle \right| \\ &\quad + \sum_{j=1}^s \frac{2\epsilon}{1 - 2\epsilon} \left| \left\langle d^{(\ell)} \eta_w - d^{(\ell)} \eta_z, d\eta_{w_j} \right\rangle \right| + \epsilon, \end{aligned}$$

where it should not be forgotten that we have made R -related terms small by arguing that N is large. By a continuity argument, we can further bound $\frac{1}{1-2\epsilon} \sum_{j=1}^s |\langle d^{(\ell)}\eta_w - d^{(\ell)}\eta_z, d\eta_{w_j} \rangle|$ uniformly for $z, w \in \mathbb{C}$ by a constant K .

It is now only left to estimate $\langle d^{(\ell)}\eta_w - d^{(\ell)}\eta_z, \eta_{w_j} \rangle$. (Note that due to $\langle d^2\eta_{w_j}, \eta_{w_j} \rangle = (\bar{z} - \bar{w})^2 e^{-\frac{|w|^2 + |z|^2}{2}} e^{\bar{w}z}|_{z=w_j} = 0$, this also produces a bound for $|\langle d^2K_w, K_{w_j} \rangle|$). This is however not hard: The mean value theorem implies

$$\left| \langle d^{(\ell)}\eta_w - d^{(\ell)}\eta_z, \eta_{w_j} \rangle \right| \leq \sup_{\xi \in [z, w]} \left| \eta_{w_j}^{(\ell+1)}(\xi) \right| |z - w|,$$

so that

$$\left| \langle d^{(\ell)}\eta_w - d^{(\ell)}\eta_z, \eta_{w_j} \rangle \right| \leq \left(\begin{array}{l} \sup_{\xi \in [z, w]} e^{-\frac{|\xi - w_j|^2}{2}} |\xi - w_j|, \quad \ell = 0 \\ \sup_{\xi \in [z, w]} e^{-\frac{|\xi - w_j|^2}{2}} (1 + |\xi - w_j|^2), \quad \ell = 1 \\ \sup_{\xi \in [z, w]} e^{-\frac{|\xi - w_j|^2}{2}} |\xi - w_j| (1 + |\xi - w_j|^2), \quad \ell = 2 \end{array} \right) \cdot |z - w|$$

Summing up, and letting ξ run through the entirety of \mathbb{C} , we obtain

$$\begin{aligned} \left\| \langle d^{(\ell)}\eta_w - d^{(\ell)}\eta_z, g \rangle \right\| &\leq \frac{\bar{\sigma}_\uparrow(W)}{(1-2\epsilon)} |z - w| + (2K + 1)\epsilon, \\ \implies \sup_{|z - w_j| \leq \delta} \left\| H_z(|g|^2) - (-\text{id}_2) \right\| &\leq \frac{2\bar{\sigma}_\uparrow(W)}{1-2\epsilon} \delta + 2 \left(\frac{\bar{\sigma}_\uparrow(W)}{1-2\epsilon} \delta \right)^2 + C\bar{\sigma}_\uparrow^2(W)\epsilon \end{aligned}$$

where applied Lemma 7,(27)-(28), collected a number of constants to form C , and used the evident bound $\bar{\sigma}_\uparrow(W) \geq \bar{\sigma}(W, w_i) \geq 1$. Here, the norm can be chosen to be the spectral norm: The same estimate for the ℓ_1 and ℓ_∞ -operator norms can be proven with the same technique used as for bounding $\text{id}_{2s} - \Phi(W)$. Then, an interpolation argument does the job. If we now choose

$$\bar{\delta} = \frac{1}{\bar{\sigma}_\uparrow} \left(\sqrt{\frac{3}{4} - \frac{C\bar{\sigma}_\uparrow^2\epsilon}{2}} - \frac{1}{2} \right) (1 - 2\epsilon),$$

$H_z(|g|^2)$ will be negative definite for each z with $\text{dist}(z, W) \leq \bar{\delta}$. Also note that for ϵ small enough, we then have $\bar{\delta} \geq \frac{\sqrt{3}-1}{4\bar{\sigma}(W)_\uparrow}$.

It is by now easily estimate $|\langle K_z, g \rangle|$ for z with $\text{dist}(z, W) \geq \bar{\delta}$ directly. We have,

$$\begin{aligned} |\langle \eta_z, g \rangle| &\leq \sum_{j=1}^s |\alpha_j| |\langle \eta_z, \eta_{w_j} \rangle| + |\beta_j| |\langle K_z, dK_{w_j} \rangle| \\ &\leq \frac{1}{1-2\epsilon} \sum_{j=1}^s e^{-\frac{|z - w_j|^2}{2}} + \frac{2\epsilon}{1-2\epsilon} \sum_{j=1}^s e^{-\frac{|z - w_j|^2}{2}} |z - w_j| \end{aligned}$$

where we again invoked (26). Now we take the supremum over all z with $\text{dist}(z, W) \geq \bar{\delta}$. These automatically obey the estimate $\text{dist}(z, W) \geq \frac{\sqrt{3}-1}{4\bar{\sigma}(W)\uparrow}$, so that

$$|\langle \eta_z, g \rangle| \leq \frac{1}{1-2\epsilon}(1-\tau) + \frac{2\epsilon}{1-2\epsilon} \bar{\sigma}_\uparrow(W) < 1,$$

for $\epsilon > 0$ small enough. The proof is finished. \square

Remark 3. In practice, we of course have to subsample $A_\Lambda r$ somehow. We can for instance do it as follows: Let $(M_n)_{n \in \mathbb{N}} := (e^{-|z|^2/2} \frac{1}{\sqrt{n!}} z^n)_{n \in \mathbb{N}}$ denote the orthonormal basis of monomials for \mathcal{H} . We then measure

$$\langle f, M_n \rangle, \quad n = 0 \dots N+1.$$

Note that we can only need to have access to r in order to calculate these values: $\langle f, M_n \rangle = \langle A_\Lambda r, e_n \rangle = \langle r, A_\Lambda^* e_n \rangle$.

The range of these measurements is given by $\text{span}(M_n)_{n=0}^{N+1}$. By using the following approximations of η_w and $d\eta_w$:

$$p_{w_k}^N = e^{-\frac{|w_k|^2 + |z|^2}{2}} \sum_{n=0}^N \frac{\overline{w_k}^n}{\sqrt{n!}} m_n, \quad q_{w_k}^N = (z - w_k) e^{-\frac{|w_k|^2 + |z|^2}{2}} \sum_{n=0}^N \frac{\overline{w_k}^n}{\sqrt{n!}} m_n.$$

instead of η_{w_k} and $d\eta_{w_k}$ for our ansatz, we can apply a similar argument to above to deduce that if N is large enough, exact recovery from these measurements. Due to the similarity to the above argument, we choose to not include the details here. These will instead be included in the upcoming PhD thesis of one of the authors.

Acknowledgement

B. G. Bodmann's research was partially supported by NSF grant DMS-1715735. A. Flinth acknowledges support from the Berlin Mathematical School, as well as DFG Grant KU 1446/18-1. G. Kutyniok acknowledges partial support by the Einstein Foundation Berlin, the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics" and the European Commission Project DEDALE (contract no. 665044) within the H2020 Framework Program, and DFG-SPP 1798 Grants KU 1446/21 and KU 1446/23. She also thanks the Department of Mathematics at the University of Houston for its hospitality and support during her visit.

References

- [1] Adcock, B. and Hansen, A. C., Generalized sampling and infinite-dimensional compressed sensing, *Found. Comp. Math.*, 16(5):1263–1323, 2016.
- [2] Aubel, C., Stotz, D., and Bölcskei, H., A Theory of Super-Resolution from Short-Time Fourier Transform Measurements, *J. Four. Anal. Appl.*, 24(1):45–107, 2018.

- [3] Bargmann, V., On a Hilbert space of analytic functions and an associated integral transform, Part I. *Commun. Pure Appl. Math.*, 14(3):187–214,1961.
- [4] Bargmann, V., On a Hilbert space of analytic functions and an associated integral transform. part II: A family of related function spaces application to distribution theory. *Commun. Pure Appl. Math.*, 20(1):1–101, 1967.
- [5] Beurling, A., Interpolation for an interval on R^1 . A density theorem, Mittag-Leffler Lectures on Harmonic Analysis 1977-78. From *Collected Works of Arne Beurling: Volume 2 Harmonic Analysis*, p. 351-359, Birkhäuser, 1989.
- [6] Borwein, J.M., and Lewis, A.S. Partially finite convex programming, Part I: Quasi relative interiors and duality theory. *Math. Program.*, 57(15):15–48, 1992.
- [7] Bredies, K., and Pikkarainen, H. K., Inverse problems in spaces of measures, *ESAIM Contr. Op. Ca. Va.*, 19(1):190–218, 2013.
- [8] Candès, E. J., and Fernandez-Granda, C., Towards a Mathematical Theory of Super-resolution. *Commun. Pure Appl. Math.*, 67(6):906–956, 2014.
- [9] Candès, E. J., and Plan, Y., A Probabilistic and RIPless Theory of Compressed Sensing, *IEEE Trans. Inf. Theory*, 57(11): 7235–7254, 2011.
- [10] Candès, E. J., Romberg, J. K., and Tao, T., Stable signal recovery from incomplete and inaccurate measurements *Commun. Pure Appl. Math.*, 59(8):1207–1223,2006.
- [11] Candès, E. J., and Tao, T., Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.
- [12] Chandrasekaran, V., Recht, B., Parrilo, P., and Willsky, A. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6): 805–549, 2012.
- [13] Chi, Y, Scharf, L. L, Pezeshki, A., and Calderbank, R. A., Sensitivity to basis mismatch in compressed sensing, *IEEE Trans. Sign. Proc.*, 59(5):2182–2195, 2011.
- [14] Davenport, M., Duarte, M., Eldar, Y., and G. Kutyniok. Introduction to Compressed Sensing. *Compressed Sensing: Theory and Applications*, 1-64, Cambridge University Press, 2012.
- [15] De Castro, Y. and Gamboa, F., Exact reconstruction using Beurling minimal extrapolation, *J. Math. Anal Appl.*, 395(1):336–354, 2012.
- [16] De Castro, Y., Gamboa, F., Henrion, D. and Lasserre, J-B. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inf. Theory*, 63(1):621–630,2017
- [17] Donoho, D. L, Compressed sensing, *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.
- [18] Donoho, D. L, Superresolution via sparsity constraints, *SIAM J. Math. Anal.*, 23(5):1309–1331, 1992.

- [19] Dossal, C., Duval, V., and Poon, C., Sampling the Fourier transform along radial lines, *SIAM J. Numer. Anal.*, 55(6):2540–2564, 2017.
- [20] Duval, V., and Peyrè, G., Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.*, 15(5):1314–1355, 2015.
- [21] Foucart, S., and Rauhut, H. A mathematical introduction to Compressed Sensing, Birkhäuser, 2013.
- [22] Gribonval, R., Blanchard, G., Keriven, N. and Traonmilin, Y., Compressive Statistical Learning with Random Feature Moments, *arXiv preprint arXiv: 1706.07180*, 2017
- [23] Flinth, A., and Weiss, P., Exact solutions of infinite dimensional total-variation regularized problems, *arXiv preprint arXiv: 1708.02157*, 2017
- [24] Eldar, Y. C., Compressed sensing of analog signals in shift-invariant spaces, *IEEE Trans. Sig. Process.*, 57(8):2986–2997, 2009
- [25] Heckel, R., and Morgenshtern, V. I. and Soltanolkotabi, M., Super-resolution radar, *Information and Inference*, 5(1):22–75, 2016.
- [26] Heil, C., Ramanathan, J., and Topiwala, P., Linear independence of time-frequency translates, *Proc. Amer. Math. Soc.*, 124(9):2787–2795, 1996.
- [27] Kreutz-Delgado, K., The complex gradient operator and the CR-calculus, *arXiv preprint arXiv:0906.4835*, 2009.
- [28] Pettis, B. J., On integration in vector spaces, *Trans. Amer. Math. Soc.*, 44, 277–304, 1938.
- [29] Plonka, G. and Tasche, M., Prony methods for recovery of structured functions, *GAMM-Mitteilungen*, 37(2):239–258, 2014.
- [30] Talagrand, M., Pettis Integral and Measure Theory, *Memoirs of the Amer. Math. Soc.* 307, 1984.
- [31] Temlyakov, V.N., Weak greedy algorithms. *Adv. Comp. Math.*, 12(2,3):213–227, 2000.
- [32] Tang, G., Bhaskar, B. N., Shah, P. and Recht, B., Compressed Sensing off the Grid *IEEE Trans. Inf. Theory*, 59(11):7465–7490, 2013
- [33] Unser, M., Fageot, J and Ward, J. P., Splines Are Universal Solutions of Linear Inverse Problems with Generalized TV Regularization, *SIAM Review*, 59(4):769–793, 2017.
- [34] Vetterli, M., Marziliano, P., and Blu, T., Sampling signals with finite rate of innovation, *IEEE Trans. Sign. Proc.*, 50(6):1417–1428, 2002.
- [35] Yosida, K., Functional Analysis, *Die Grundlehren der mathematischen Wissenschaften*, vol. 123, 6th edition, Springer-Verlag, Berlin/Heidelberg/New York, 1995.