ON CONTINUOUS-DOMAIN INVERSE PROBLEMS WITH SPARSE SUPERPOSITIONS OF DECAYING SINUSOIDS AS SOLUTIONS

Rahul Parhi and Robert D. Nowak

Department of Electrical and Computer Engineering, University of Wisconsin–Madison

ABSTRACT

We study a family of inverse problems in which a continuous-domain object is reconstructed from a finite number of noisy linear measurements. We study regularization methods for solving these problems in which the regularizers promote sparsity in the frequency domain. We show that sparse superpositions of decaying sinusoids are solutions to these continuous-domain linear inverse problems, where the number of terms in the superposition is upper bounded by the number of measurements. This results in new forms of regularization for sparse reconstruction that are different from classical techniques. We numerically illustrate the efficacy of these new regularization techniques in the problem of image reconstruction.

Index Terms— inverse problems, regularization, sparsity, Banach space, spectral Barron space

1. INTRODUCTION

Solving a linear inverse problem amounts to reconstructing a *continuous-domain* object from a finite number of possibly noisy *linear* measurements. The prototypical problem involves the reconstruction of the continuous-domain object $f : \mathbb{R}^d \to \mathbb{R}$ from a vector of measurements

$$\boldsymbol{y} = \mathrm{H}\{f\} + \boldsymbol{\varepsilon} \in \mathbb{R}^N,$$

where H symbolizes the linear measurement process, $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ denotes a perturbation or noise term, typically assumed to be a vector of i.i.d. zero-mean Gaussian random variables, $\boldsymbol{y} \in \mathbb{R}^N$ denotes the measurements, and N denotes the number of measurements (or data). This formulation captures many real-world problems, e.g., in magnetic resonance imaging, $\mathrm{H}\{f\} = \{F(\boldsymbol{\omega}_n)\}_{n=1}^N$ is a vector of samples of the Fourier transform F of f and in statistics and supervised machine learning, $\mathrm{H}\{f\} = \{f(\boldsymbol{x}_n)\}_{n=1}^N$ corresponds to a vector of samples of f. Clearly this problem is ill-posed since f is a continuous-domain object being reconstructed from a finite number of measurements.

This research was partially supported by NSF grant DMS-2134140, ONR MURI grant N00014-20-1-2787, AFOSR/AFRL grant FA9550-18-1-0166, and the NSF Graduate Research Fellowship Program under grant DGE-1747503.

A common way to solve this problem is to assume, *a priori*, that f can be written as a superposition of *atoms* from some *dictionary*. That is, assume that

$$f = \sum_{k \in \mathbb{Z}} \alpha_k \varphi_k,$$

where $\{\varphi_k\}_{k\in\mathbb{Z}}$ is a dictionary of atoms such that each atom φ_k maps $\mathbb{R}^d \to \mathbb{R}$, and $\{\alpha_k\}_{k\in\mathbb{Z}}$ are the expansion coefficients. The inverse problem is then solved by considering the following regularized least-squares problem

$$\min_{\boldsymbol{\alpha} \in \ell^p(\mathbb{Z})} \left\| \boldsymbol{y} - \mathrm{H}\left\{ \sum_{k \in \mathbb{Z}} \alpha_k \varphi_k \right\} \right\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_p^p, \tag{1}$$

where $\lambda > 0$ is an adjustable regularization parameter. While the choice of $p = 2$ has classically been the common choice, the last few decades have shown that *sparsity* ($p = 1$) plays a key role in signal reconstruction [4, 11]. In particular, in many real-world problems, assuming that the signal to be reconstructed is sparse in some dictionary results in better reconstruction error. Moreover, this idea plays a central role in compressed sensing and sparse dictionary learning [7, 9, 5].

Let $\widehat{\boldsymbol{\alpha}}$ be a solution to (1). Then, the object that generated the measurements is *estimated* by

$$\widehat{f} = \sum_{k \in \mathbb{Z}} \widehat{\alpha}_k \varphi_k. \tag{2}$$

The problem in (1) is the so-called *synthesis* formulation of the problem since the resulting solution is explicitly *synthesized* from the dictionary $\{\varphi_k\}_{k\in\mathbb{Z}}$. When $p = 1$ and $\{\varphi_k\}_{k\in\mathbb{Z}}$ is an orthogonal wavelet basis, the estimator in (2) is the well-known wavelet shrinkage estimator [10].

In this paper, we will adopt the so-called *analysis* (or variational) formulation of this problem in which the estimator is a solution to a *variational problem* of the form

$$\min_{f \in \mathcal{X}'} \|\boldsymbol{y} - \mathrm{H}\{f\}\|_2^2 + \lambda \|f\|_{\mathcal{X}'}^p, \tag{3}$$

where $\lambda > 0$ and $p \in [1, \infty)$ are adjustable hyperparameters, $\mathcal{X}'$ is an appropriate native space, and the regularization term $\|\cdot\|_{\mathcal{X}'}$ is typically a norm or seminorm that defines $\mathcal{X}'$. This formulation captures many techniques for estimating objects

from measurements including splines [8, 15, 25, 14] and neural networks [17, 18, 20, 19]. We write $\mathcal{X}'$ since it's convenient to view the native space as a dual space.

There are many works that study the problem of reconstructing infinite-dimensional objects from measurements [1, 6, 12], and this work provides a new perspective. In this paper, we study the problem in (3) where the regularization term corresponds to the *sparsity* of the first $s \geq 0$ (fractional) derivatives of f in the frequency domain. These regularization terms define function spaces known as the so-called *spectral Barron spaces* studied by a number of authors in the context of approximation theory with neural networks [2, 16, 22]. It turns out these spaces are rather interesting from a signal reconstruction perspective and differ from classical Fourier techniques due to their sparsity-promoting nature. Let $\mathscr{B}^s(\mathbb{R}^d)$ denote the spectral Barron space of order s, which is the space of functions $f : \mathbb{R}^d \to \mathbb{R}$ in $\mathscr{B}^s(\mathbb{R}^d)$ such that the quantity

$$\|f\|_{\mathscr{B}^s(\mathbb{R}^d)} = \|(1 + \|\cdot\|_2)^s F(\cdot)\|_{\mathcal{M}(\mathbb{R}^d)}$$

is finite, where F is the Fourier transform of f and the $\mathcal{M}$-norm is a "generalization" of the L^1-norm that can also be applied to distributions such as the Dirac impulse. Whenever the quantity $(1 + \|\cdot\|_2)^s F(\cdot)$ is a *bona fide* function (and not a distribution), we have that

$$\|f\|_{\mathscr{B}^s(\mathbb{R}^d)} = \int_{\mathbb{R}^d} (1 + \|\boldsymbol{\omega}\|_2)^s |F(\boldsymbol{\omega})| \, \mathrm{d}\boldsymbol{\omega}.$$

In this paper, we show that

1. $\mathscr{B}^s(\mathbb{R}^d)$ is a *Banach space*[1].
2. Under mild conditions on the measurement operator H, the solution set to the variational problem

$$\min_{f \in \mathscr{B}^s(\mathbb{R}^d)} \|\boldsymbol{y} - \mathrm{H}\{f\}\|_2^2 + \lambda \|f\|_{\mathscr{B}^s(\mathbb{R}^d)}, \qquad (4)$$

   is completely characterized by *sparse* superpositions of decaying sinusoids of the form

$$s(\boldsymbol{x}) = \sum_{k=1}^{K} \alpha_k (1 + \|\boldsymbol{\omega}_k\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\omega}_k^\mathsf{T} \boldsymbol{x}},$$

   where $\mathrm{j}^2 = -1$, $\boldsymbol{\omega}_k \in \mathbb{R}^d$, $k = 1, \ldots, K$, and $K \leq N$, where N is the number of measurements. These solutions are sparse in the sense that $\|s\|_{\mathscr{B}^s(\mathbb{R}^d)} = \|\boldsymbol{\alpha}\|_1$, and in the sense that the Fourier transform of s is an impulse train.
3. Many convolution operators satisfy the mild conditions on H, and therefore the problem becomes equivalent to deconvolution.
4. We numerically illustrate the efficacy of the regularization scheme in (4) for image reconstruction.

[1] In fact, it is a Banach space with a sparsity-promoting norm.

2. MAIN RESULTS

In order to prove our results outlined in Section 1, we require some definitions and notation from functional analysis [13]. Let $\mathscr{S}(\mathbb{R}^d)$ denote the Schwartz space of smooth and rapidly decaying test functions in $\mathbb{R}^d$ with continuous dual $\mathscr{S}'(\mathbb{R}^d)$, the space of tempered distributions on $\mathbb{R}^d$. Let $C_0(\mathbb{R}^d)$ denote the space of continuous functions vanishing at infinity with continuous dual $\mathcal{M}(\mathbb{R}^d)$, the space of finite Radon measures. The space $\mathcal{M}(\mathbb{R}^d)$ is a Banach space when equipped with the total variation norm in the sense of measures, which can be viewed as a "generalization" of the L^1-norm, but may also be applied to distributions such as the Dirac impulse[2]. Since $\mathscr{S}(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$, we may view $\mathcal{M}(\mathbb{R}^d)$ as a subspace of $\mathscr{S}'(\mathbb{R}^d)$. In particular, we may view it, formally, as the space of "absolutely integrable distributions". Moreover, the $\mathcal{M}$-norm is the continuous-domain analogue of the ℓ^1-norm. Working with these spaces allows us to rigorously work with distributional Fourier transforms and derivatives.

Given $f \in \mathscr{S}'(\mathbb{R}^d)$, consider the operator

$$\mathrm{B}_s : \mathscr{S}'(\mathbb{R}^d) \ni f \mapsto (1 + \|\cdot\|_2)^s F(\cdot) \in \mathscr{S}'(\mathbb{R}^d),$$

where F is the (distributional) Fourier transform[3] of f. Then, the spectral Barron space of order $s \geq 0$ is given by

$$\mathscr{B}^s(\mathbb{R}^d) \coloneqq \Big\{ f \in \mathscr{S}'(\mathbb{R}^d) : \ \|\mathrm{B}_s f\|_{\mathcal{M}(\mathbb{R}^d)} < \infty \Big\}. \qquad (5)$$

In this paper, we are interested in the solutions to the variational problem in (4). Notice that $\|\mathrm{B}_s f\|_{\mathcal{M}(\mathbb{R}^d)}$ corresponds to the integrability (really the total variation in the sense of measures) of the first s derivatives of f in the frequency domain and therefore measures *sparsity* in the frequency domain. Before characterizing the solution set to (4), we require some intermediary results summarized in the following lemma.

Lemma 1. *The space $\mathscr{B}^s(\mathbb{R}^d)$, $s \geq 0$, when equipped with the norm*

$$\|f\|_{\mathscr{B}^s(\mathbb{R}^d)} \coloneqq \|\mathrm{B}_s f\|_{\mathcal{M}(\mathbb{R}^d)} \qquad (6)$$

has the following properties:

1. *It is a Banach space.*
2. *There exists a Banach space $\mathcal{X}$ such that $\mathcal{X}'$, the continuous dual of $\mathcal{X}$, is $\mathscr{B}^s(\mathbb{R}^d)$.*
3. *The extreme points of the unit ball*

$$\Big\{ f \in \mathscr{B}^s(\mathbb{R}^d) : \ \|f\|_{\mathscr{B}^s(\mathbb{R}^d)} \leq 1 \Big\}$$

   are of the form $\boldsymbol{x} \mapsto \pm(1 + \|\boldsymbol{\omega}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\omega}^\mathsf{T} \boldsymbol{x}}$, $\boldsymbol{\omega} \in \mathbb{R}^d$.

[2] Indeed, for any $f \in L^1(\mathbb{R}^d)$ we have that $\|f\|_{L^1(\mathbb{R}^d)} = \|f\|_{\mathcal{M}(\mathbb{R}^d)}$. Moreover, the shifted Dirac impulse $\delta(\cdot - \boldsymbol{x}_0)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, is not in $L^1(\mathbb{R}^d)$ but is in $\mathcal{M}(\mathbb{R}^d)$ with the property that $\|\delta(\cdot - \boldsymbol{x}_0)\|_{\mathcal{M}(\mathbb{R}^d)} = 1$.

[3] If $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, then $F(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\boldsymbol{x}) e^{-\mathrm{j}2\pi \boldsymbol{\omega}^\mathsf{T} \boldsymbol{x}} \, \mathrm{d}\boldsymbol{x}$. Working with this normalization of the Fourier transform allows for the cleanest arguments, avoiding the necessity of keeping track of normalization constants.

Proof. Consider the operator

$$\mathrm{B}_s^{-1} : \varphi \mapsto \int_{\mathbb{R}^d} (1 + \|\boldsymbol{\xi}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T}(\cdot)} \varphi(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi},$$

where $\varphi \in \mathscr{S}(\mathbb{R}^d)$. Note that for $\varphi \in \mathscr{S}(\mathbb{R}^d)$,

$$\begin{aligned}
&(\mathrm{B}_s \, \mathrm{B}_s^{-1} \, \varphi)(\boldsymbol{\omega}) \\
&= (1 + \|\boldsymbol{\omega}\|_2)^s \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (1 + \|\boldsymbol{\xi}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}} \\
&\qquad\qquad \varphi(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} \, e^{-\mathrm{j}2\pi \boldsymbol{\omega}^\mathsf{T} \boldsymbol{x}} \,\mathrm{d}\boldsymbol{x} \\
&= (1 + \|\boldsymbol{\omega}\|_2)^s \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}} e^{-\mathrm{j}2\pi \boldsymbol{\omega}^\mathsf{T} \boldsymbol{x}} \,\mathrm{d}\boldsymbol{x} \\
&\qquad\qquad (1 + \|\boldsymbol{\xi}\|_2)^{-s} \varphi(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} \\
&\overset{(*)}{=} (1 + \|\boldsymbol{\omega}\|_2)^s \int_{\mathbb{R}^d} \delta(\boldsymbol{\xi} - \boldsymbol{\omega}) (1 + \|\boldsymbol{\xi}\|_2)^{-s} \varphi(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} \\
&= (1 + \|\boldsymbol{\omega}\|_2)^s (1 + \|\boldsymbol{\omega}\|_2)^{-s} \varphi(\boldsymbol{\omega}) = \varphi(\boldsymbol{\omega}),
\end{aligned}$$

where $(*)$ holds since the Fourier transform of $\boldsymbol{x} \mapsto e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}}$ is the Dirac impulse $\boldsymbol{\omega} \mapsto \delta(\boldsymbol{\xi} - \boldsymbol{\omega})$. Thus, B_s^{-1} is a *right-inverse* of B_s when restricted to $\mathscr{S}(\mathbb{R}^d)$. We can then extend B_s^{-1} to act on $\mathcal{M}(\mathbb{R}^d)$ with the same right-inverse property.

Next, note that for $f \in \mathscr{S}(\mathbb{R}^d)$,

$$\begin{aligned}
&(\mathrm{B}_s^{-1} \, \mathrm{B}_s \, f)(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} (1 + \|\boldsymbol{\xi}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}} \Big[(1 + \|\boldsymbol{\xi}\|_2)^s F(\boldsymbol{\xi}) \Big] \,\mathrm{d}\boldsymbol{\xi} \\
&= \int_{\mathbb{R}^d} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}} F(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} = f(\boldsymbol{x}),
\end{aligned}$$

where the last line holds by the Fourier inversion formula. Thus, B_s^{-1} is a *left-inverse* of B_s when restricted to $\mathscr{S}(\mathbb{R}^d)$. One can check that $\mathscr{S}(\mathbb{R}^d) \subset \mathscr{B}^s(\mathbb{R}^d) \subset \mathscr{S}'(\mathbb{R}^d)$. We can then extend B_s to act on $\mathscr{B}^s(\mathbb{R}^d)$ with the same left-inverse property.

Therefore, we have that $\mathrm{B}_s : \mathscr{B}^s(\mathbb{R}^d) \to \mathcal{M}(\mathbb{R}^d)$ is invertible with inverse given by $\mathrm{B}_s^{-1} : \mathcal{M}(\mathbb{R}^d) \to \mathscr{B}^s(\mathbb{R}^d)$. Moreover, these are isometric isomorphisms by definition of $\|\cdot\|_{\mathscr{B}^s(\mathbb{R}^d)}$ in (6). Since $\mathcal{M}(\mathbb{R}^d)$ is a Banach space, this says $\mathscr{B}^s(\mathbb{R}^d)$ is a Banach space when equipped with the norm in (6). Moreover, the predual $\mathcal{X}$ of $\mathscr{B}^s(\mathbb{R}^d)$ is given by the image of $C_0(\mathbb{R}^d)$ under the mapping B_s^*, the adjoint of B_s. This is summarized in the following diagram:

$$\begin{array}{ccc}
\mathscr{B}^s(\mathbb{R}^d) & \underset{\mathrm{B}_s^{-1}}{\overset{\mathrm{B}_s}{\rightleftarrows}} & \mathcal{M}(\mathbb{R}^d) \\
\Big\uparrow \text{dual} & & \Big\uparrow \text{dual} \\
\mathcal{X} & \underset{\mathrm{B}_s^{-1*}}{\overset{\mathrm{B}_s^*}{\rightleftarrows}} & C_0(\mathbb{R}^d)
\end{array} \tag{7}$$

Finally, it is well-known that the extreme points of the unit ball in $\mathcal{M}(\mathbb{R}^d)$ are of the form $\pm\delta(\cdot - \boldsymbol{\omega})$, $\boldsymbol{\omega} \in \mathbb{R}^d$ (see, e.g., [3, Proposition 4.1]). Since isometric isomorphisms map extreme points of unit balls to extreme points of unit balls, it follows that the extreme points of the unit ball in $\mathscr{B}^s(\mathbb{R}^d)$ are of the form $\mathrm{B}_s^{-1}\{\pm\delta(\cdot - \boldsymbol{\omega})\}(\boldsymbol{x}) = \pm(1 + \|\boldsymbol{\omega}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\omega}^\mathsf{T} \boldsymbol{x}}$. □

We can now characterize the solution set of (4).

Theorem 2. *Let $\mathcal{X}$ denote the predual of $\mathscr{B}^s(\mathbb{R}^d)$, i.e., $\mathcal{X}' = \mathscr{B}^s(\mathbb{R}^d)$. Let $\mathrm{H}\{f\} = (\langle h_1, f\rangle, \ldots, \langle h_N, f\rangle) \in \mathbb{R}^N$ denote the measurement process where $h_n \in \mathcal{X}$, $n = 1, \ldots, N$, i.e., the measurements are linear and weak* continuous. Then, for any fixed $\boldsymbol{y} \in \mathbb{R}^N$, the solution set*

$$\mathcal{V} \coloneqq \operatorname*{arg\,min}_{f \in \mathscr{B}^s(\mathbb{R}^d)} \|\boldsymbol{y} - \mathrm{H}\{f\}\|_2^2 + \lambda \|f\|_{\mathscr{B}^s(\mathbb{R}^d)}, \tag{8}$$

where $\lambda > 0$ is an adjustable regularization parameter, is nonempty, convex, and weak compact. Moreover, the extreme points of $\mathcal{V}$ are given by functions of the form*

$$s(\boldsymbol{x}) = \sum_{k=1}^{K} \alpha_k (1 + \|\boldsymbol{\omega}_k\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\omega}_k^\mathsf{T} \boldsymbol{x}}, \tag{9}$$

where $\boldsymbol{\omega}_k \in \mathbb{R}^d$, $k = 1, \ldots, K$, and $K \leq N$. The convex hull of these extreme points is the full solution set.

Proof. The proof follows by using Lemma 1 with [24, Theorems 2 and 3], which characterizes the solution sets of variational problems in terms of the extreme points of the unit ball of the regularization term. □

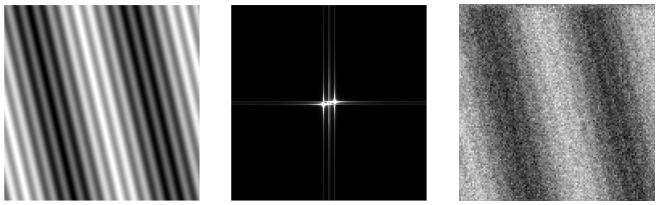
Corollary 3. *The measurements $h_n = \phi(\boldsymbol{x}_n - \cdot)$ given by the convolution kernel ϕ, i.e., $\langle h_n, f\rangle = (\phi * f)(\boldsymbol{x}_n)$, satisfies the hypotheses of Theorem 2 when its Fourier transform $\Phi(\boldsymbol{\omega})$ decays as $\|\boldsymbol{\omega}\|_2^{-t}$, where $t > -s$, as $\|\boldsymbol{\omega}\|_2 \to \infty$.*

Proof. We must show that $\phi(\boldsymbol{x}_0 - \cdot) \in \mathcal{X}$, for any $\boldsymbol{x}_0 \in \mathbb{R}^d$. From the diagram in (7), we see that this amounts to showing that $\mathrm{B}_s^{-1*}\{\phi(\boldsymbol{x}_0 - \cdot)\} \in C_0(\mathbb{R}^d)$. We have

$$\begin{aligned}
&\mathrm{B}_s^{-1*}\{\phi(\boldsymbol{x}_0 - \cdot)\}(\boldsymbol{\xi}) \\
&= (1 + \|\boldsymbol{\xi}\|_2)^{-s} \int_{\mathbb{R}^d} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}} \phi(\boldsymbol{x}_0 - \boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\
&= (1 + \|\boldsymbol{\xi}\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\xi}^\mathsf{T} \boldsymbol{x}_0} \Phi(\boldsymbol{\xi}).
\end{aligned}$$

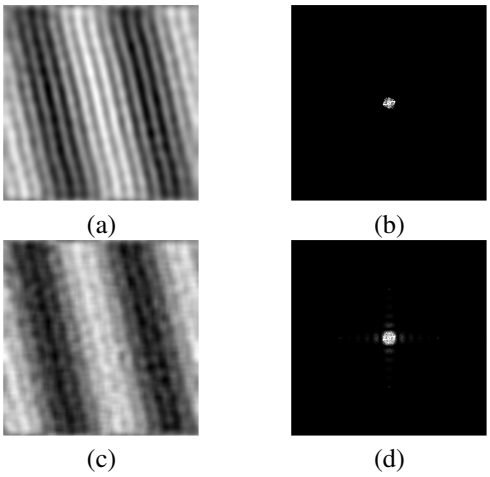
When Φ satisfies the hypothesis from the corollary statement, the above tends to 0 as $\|\boldsymbol{\xi}\|_2 \to \infty$, and therefore $\mathrm{B}_s^{-1*}\{\phi(\boldsymbol{x}_0 - \cdot)\} \in C_0(\mathbb{R}^d)$. □

The decay condition of Corollary 3 says that the convolved atoms in (9) must vanish at infinity as a function of the frequency parameter $\boldsymbol{\omega}$. When $s > 0$, the decay condition captures the ideal sampling setting by choosing $\phi = \delta$, the Dirac impulse, i.e., $\mathrm{H}\{f\} = \{f(\boldsymbol{x}_n)\}_{n=1}^N$ and so (8) reduces to a nonparametric least-squares problem. This decay conditions also captures the impulse response of many physical systems.



(a) (b) (c)

Fig. 1. In (a) we show the image to be reconstructed. In (b), we show the magnitude of the DFT of the image. In (c) we show the measurement of the image from (a) which is downsampled by a factor of 8 and then corrupted with noise.



(a) (b) (c) (d)

Fig. 2. In (a) we show the reconstruction from the procedure in (10) with $s = 1$. In (b) we show the magnitude of the DFT of (a). In (c) we show the reconstruction from the procedure in (11). In (d) we show the magnitude of the DFT of (c).

Remarkably, if we compute the $\mathscr{B}^s(\mathbb{R}^d)$-norm of the solution in (9), we find $\|s\|_{\mathscr{B}^s(\mathbb{R}^d)} = \|\boldsymbol{\alpha}\|_1$. Moreover, when $s = 0$ and the input dimension $d = 1$, the problem in (8) reduces to the problem of off-the-grid compressed sensing [23]. We also remark that when the measurement operator is as in Corollary 3, we have $\langle h_n, s\rangle = (\phi * s)(\boldsymbol{x}_n)$. Using the Fourier transform to compute the convolution, we find

$$(\phi * s)(\boldsymbol{x}_n) = \sum_{k=1}^{K} \alpha_k \Phi(\boldsymbol{\omega}_k)(1 + \|\boldsymbol{\omega}_k\|_2)^{-s} e^{\mathrm{j}2\pi \boldsymbol{\omega}_k^{\mathsf{T}} \boldsymbol{x}_n},$$

where Φ is the Fourier transform of ϕ.

3. NUMERICAL EXPERIMENTS

In this section we experimentally show the efficacy of the solutions to the variational problem in Theorem 2 in the problem of image reconstruction from noisy measurements. We discretize the problem in (8) by working with digital images and noting that every $N \times N$ digital image $\boldsymbol{f}$ admits a representation in terms of its discrete Fourier transform (DFT) via

$$\boldsymbol{f}[n_1, n_2] = \frac{1}{N^2} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \boldsymbol{F}[k_1, k_2]\, e^{\mathrm{j}2\pi \frac{k_1}{N} n_1} e^{\mathrm{j}2\pi \frac{k_2}{N} n_2},$$

where $\boldsymbol{F}$ is the DFT of $\boldsymbol{f}$. We can consider the following discretization of the problem in (8):

$$\min_{\boldsymbol{f}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{f}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \tag{10}$$

where $\mathbf{H}$ is a discrete approximation of H,

$$\boldsymbol{\alpha}[k_1, k_2] = \left(1 + \sqrt{\left|\frac{k_1}{N}\right|^2 + \left|\frac{k_2}{N}\right|^2}\right)^s \boldsymbol{F}[k_1, k_2]$$

denotes the discrete analogue of the coefficients α_k in (9), and $\|\boldsymbol{\alpha}\|_1$ denotes the ℓ^1-norm of $\boldsymbol{\alpha}$ viewed as a vector, which is the discrete analogue of $\|f\|_{\mathscr{B}^s(\mathbb{R}^2)}$.

Consider reconstructing an image from the observations

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{f} + \boldsymbol{\varepsilon},$$

where $\mathbf{H}$ is a convolution followed by downsampling and $\boldsymbol{\varepsilon}$ is additive white Gaussian noise. In Fig. 1 we show an example image to be reconstructed, the magnitude of its DFT, and measurements of the image given by a convolved, downsampled, and noisy version of the original image. This image was chosen due to its sparse DFT. In Fig. 2 we show how the image is reconstructed by the procedure in (10) with $s = 1$. We compare this procedure with the classical Tikhonov regularization procedure

$$\min_{\boldsymbol{f}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{f}\|_2^2 + \lambda \|\boldsymbol{f}\|_2^2, \tag{11}$$

where $\|\boldsymbol{f}\|_2$ denotes the ℓ^2-norm of $\boldsymbol{f}$ viewed as a vector. The problems in (10) and (11) were solved using proximal gradient methods [21]. We see in Fig. 2(b) and Fig. 2(d) that solutions to (10) have sparser DFTs than solutions to (11). Moreover, the reconstruction in Fig. 2(a) is much better than the reconstruction in Fig. 2(c).

4. CONCLUSION

In this paper we studied a family of continuous-domain linear inverse problems with sparse superpositions of decaying sinusoids as solutions. We studied new forms of regularization for sparse reconstruction which promote solutions with sparse Fourier transforms. We numerically illustrated the efficacy of these new regularization techniques over classical techniques for image reconstruction. Future work will be directed towards quantifying the reconstruction error from these regularization techniques and an exhaustive comparison of this new approach to to previously studied techniques from sparse reconstruction, particularly with real-world signals and images.

5. REFERENCES

[1] B. Adcock and A. C. Hansen, "Generalized sampling and infinite-dimensional compressed sensing," *Foundations of Computational Mathematics*, vol. 16, no. 5, pp. 1263–1323, 2016.

[2] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information theory*, vol. 39, no. 3, pp. 930–945, 1993.

[3] K. Bredies and M. Carioni, "Sparsity of solutions for variational inverse problems with finite-dimensional data," *Calculus of Variations and Partial Differential Equations*, vol. 59, no. 1, p. 14, 2020.

[4] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[5] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007.

[6] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on pure and applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[7] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[8] C. de Boor and R. E. Lynch, "On splines and their minimum properties," *Journal of Mathematics and Mechanics*, vol. 15, no. 6, pp. 953–969, 1966.

[9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[10] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *The Annals of Statistics*, vol. 26, no. 3, pp. 879–921, 1998.

[11] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer Science & Business Media, 2010.

[12] Y. C. Eldar and T. Werther, "General framework for consistent sampling in hilbert spaces," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 3, no. 04, pp. 497–509, 2005.

[13] G. B. Folland, *Real analysis: modern techniques and their applications*, 2nd ed. New York: John Wiley & Sons, 1999.

[14] H. Gupta, J. Fageot, and M. Unser, "Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4670–4684, 2018.

[15] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of mathematical analysis and applications*, vol. 33, no. 1, pp. 82–95, 1971.

[16] J. M. Klusowski and A. R. Barron, "Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls," *IEEE Transactions on Information Theory*, vol. 64, no. 12, pp. 7649–7656, 2018.

[17] R. Parhi and R. D. Nowak, "The role of neural network activation functions," *IEEE Signal Processing Letters*, vol. 27, pp. 1779–1783, 2020.

[18] R. Parhi and R. D. Nowak, "Banach space representer theorems for neural networks and ridge splines," *Journal of Machine Learning Research*, vol. 22, no. 43, pp. 1–40, 2021.

[19] R. Parhi and R. D. Nowak, "Near-minimax optimal estimation with shallow ReLU neural networks," *arXiv preprint arXiv:2109.08844*, 2021.

[20] R. Parhi and R. D. Nowak, "What kinds of functions do deep neural networks learn? Insights from variational spline theory," *arXiv preprint arXiv:2105.03361*, 2021.

[21] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 2015.

[22] J. W. Siegel and J. Xu, "Characterization of the variation spaces corresponding to shallow neural networks," *arXiv preprint arXiv:2106.15002*, 2021.

[23] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[24] M. Unser, "A unifying representer theorem for inverse problems and machine learning," *Foundations of Computational Mathematics*, pp. 1–20, 2020.

[25] M. Unser, J. Fageot, and J. P. Ward, "Splines are universal solutions of linear inverse problems with generalized TV regularization," *SIAM Review*, vol. 59, no. 4, pp. 769–793, 2017.