This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

Continuous-Domain Solutions of Linear Inverse Problems with Tikhonov vs. Generalized TV Regularization

Harshit Gupta, Julien Fageot, and Michael Unser

***Abstract*—We consider 1D linear inverse problems that are formulated in the continuous domain. The object of recovery is a function that is assumed to minimize a convex objective functional. The solutions are constrained by imposing a continuous-domain regularization. We derive the parametric form of the solution (representer theorems) for Tikhonov (quadratic) and generalized total-variation (gTV) regularizations. We show that, in both cases, the solutions are splines that are intimately related to the regularization operator. In the Tikhonov case, the solution is smooth and constrained to live in a fixed subspace that depends on the measurement operator. By contrast, the gTV regularization results in a sparse solution composed of only a few dictionary elements that are upper-bounded by the number of measurements and independent of the measurement operator. Our findings for the gTV regularization resonates with the minimization of the ℓ_1 norm, which is its discrete counterpart and also produces sparse solutions. Finally, we find the experimental solutions for some measurement models in one dimension. We discuss the special case when the gTV regularization results in multiple solutions and devise an algorithm to find an extreme point of the solution set which is guaranteed to be sparse.**

***Index Terms*—Linear inverse problem, representer theorem, regularization, spline, total variation, L_2, quadratic regularization.**

I. Introduction

In a linear inverse problem, the task is to recover an unknown signal from a finite set of noisy linear measurements. To solve it, one needs a forward model that describes how these measurements are acquired. Generally, this model is stated as the continuous-domain transform of a continuous-domain signal. For example, MRI data is modeled as the samples of the Fourier transform of a continuous-domain signal. The traditional approach to state this inverse problem is to choose an arbitrary but suitable basis $\{\varphi_n\}$ and to write that the reconstructed signal is

$$f(x) = \sum_{n=1}^{N} \mathrm{f}_n \varphi_n(x), \tag{1}$$

where $\mathbf{f} = (\mathrm{f}_1, \dots, \mathrm{f}_N) \in \mathbb{R}^N$. Given the measurements $\mathbf{z} \in \mathbb{R}^M$, the task then is to find the expansion coefficients $\mathbf{f}$ by minimizing

$$\mathbf{f}^* = \arg\min_{\mathbf{f}\in\mathbb{R}^N} \left(\underbrace{\|\mathbf{z} - \mathbf{H}\mathbf{f}\|_2^2}_{\mathrm{I}} + \lambda \underbrace{\|\mathbf{L}\mathbf{f}\|_2^2}_{\mathrm{II}} \right), \tag{2}$$

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne 1015, Switzerland. This project has been funded by H2020-ERC, Grant agreement No. 692726-GlobalBioIm.

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ has elements $[\mathbf{H}]_{m,n} = \langle h_m, \varphi_n \rangle$. The analysis functions $\{h_m\}_{m=1}^M$ specify the forward model which encodes the physics of the measurement process. Term I in (2) is the data fidelity. It ensures that the recovered signal is close to the measurements. Term II is the regularization, which encodes the prior knowledge about the signal. The regularization is imposed on some transformed version of the signal coefficients using the matrix $\mathbf{L}$. Linear reconstruction algorithms [1], [2] can be used to solve Problem (2). In recent years, the notion that the real-world signals are sparse in some basis (e.g., wavelets) has become popular. This prior is imposed by using the sparsity-promoting ℓ_1-regularization norm [3], [4] and results in the minimization problem

$$\mathbf{f}^* = \arg\min_{\mathbf{f}\in\mathbb{R}^N} \left(\|\mathbf{z} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{L}\mathbf{f}\|_1 \right) \tag{3}$$

which can be efficiently solved using iterative algorithms [5], [6]. The solutions to (2), (3), and their variants with generalized data-fidelity terms are well known [7], [8], [9], [10].

While those discretization paradigms are well studied and used successfully in practice, it remains that the use of a prescribed basis $\{\varphi_n\}$, as in (1), is somewhat arbitrary.

In this paper, we propose to bypass this limitation by reformulating and solving the linear inverse problem directly in the continuous domain. To that end, we impose the regularization in the continuous domain, too, and restate the reconstruction task as a functional minimization. We show that this new formulation leads to the identification of an optimal basis for the solution which then suggests a natural way to discretize the problem.

Our contributions are two folds and are summarized as follows:

a) *Theoretical.*

- Given $\mathbf{z} \in \mathbb{R}^M$, we formalize 1D inverse problem in the continuous domain as

$$f_{\mathrm{R}}^* = \arg\min_{f\in\mathcal{X}} \underbrace{\left(\|\mathbf{z} - \mathrm{H}\{f\}\|_2^2 + \lambda \mathrm{R}(f) \right)}_{\mathrm{J}_{\mathrm{R}}(f|\mathbf{z})}, \tag{4}$$

  where f is a function that belongs to a suitable function space $\mathcal{X}$. Similarly to the discrete regularization terms $\|\mathbf{L}\mathbf{f}\|_{\ell_2}^2$ and $\|\mathbf{L}\mathbf{f}\|_{\ell_1}$ in (2) and (3), we focus on their continuous-domain counterparts $\mathrm{R}(f) = \|\mathrm{L}f\|_{L_2}^2$ and $\mathrm{R}(f) = \|\mathrm{L}f\|_{\mathcal{M}}$, respectively. There, L and H are the continuous-domain versions of $\mathbf{L}$ and $\mathbf{H}$, while $\|\mathrm{L}f\|_{\mathcal{M}}$ is the proper continuous-domain counterpart of the discrete

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

ℓ_1 norm. We show that the effect of these regularizations is similar to the one of their discrete counterparts.
- We provide the parametric form of the solution (representer theorem) that minimizes $\mathrm{J_R}(f|\mathbf{z})$ in (4) for the Tikhonov regularization $\mathrm{R}(f) = \|\mathrm{L}f\|_{L_2}^2$ and the generalized total-variation (gTV) regularization $\mathrm{R}(f) = \|\mathrm{L}f\|_{\mathcal{M}}$. Our results underline how the discrete regularization resonates with the continuous-domain one. The optimal solution for the Tikhonov case is smooth, while it is sparse for the gTV case. The optimal bases in the two cases are intimately connected to the operators L and H.
- We present theoretical results that are valid for any convex, coercive, and lower-semicontinuous data-fidelity term which is proper in the range of H. This includes the case when the data-fidelity term is $\|\mathbf{z} - \mathrm{H}\{f\}\|_2^2$. In this sense, for the gTV case our work extends the results in [11] which only deals with indicator function over a feasible convex-compact set as a data-fidelity term.

b) *Algorithmic.*
- We propose a discretization scheme to minimize $\mathrm{J_R}(f|\mathbf{z})$ in the continuous domain. Even though the minimization of $\mathrm{J_R}(f|\mathbf{z})$ is an infinite-dimensional problem, the knowledge of the optimal basis of the solution makes the problem finite-dimensional: it boils down to the search for a set of optimal expansion coefficients.
- We devise an algorithm to find a sparse solution when the gTV solution is non-unique. For this case, the optimization problem turns out to be a LASSO [8] minimization with non-unique solution. We introduce a combination of FISTA [12] and the simplex algorithm to find a sparse solution which we prove to be an extreme point of the solution set.

The paper is organized as follows: In Sections 2 and 3, we present the formulation and the theoretical results of the inverse problem for the two regularization cases. In Section 4, we compare the solutions of the two cases. We present our numerical algorithm in Section 5 and illustrate its behavior with various examples in Section 6. The mathematical proofs of the main theorems are given in the appendices and the supplementary material.

A. Related Work

The use of $\mathrm{R}(f) = \|\mathrm{L}f\|_{L_2}^2$ goes back to Tikhonov's theory of regularization [1] and to kernel methods in machine learning [13]. In the learning community, representer theorems (RT) as in [14], [15] use the theory of reproducing-kernel Hilbert spaces (RKHS) to state the solution of the problem for the restricted case where the measurements are samples of the function. For the generalized-measurement case, there are also tight connections between these techniques and variational splines and radial-basis functions [16], [17], [18]. These representer theorems, however, either have restrictions on the empirical risk functional or on the class of measurement operators.

Specific spline-based methods with quadratic regularization have been developed for inverse problems. In particular, [19], [20] used variational calculus. Here, we strengthen these results by proving the uniqueness and existence of the solution of (4) for $\mathrm{R}(f) = \|\mathrm{L}f\|_{L_2}^2$. We revisit the derivation of the result using the theory of RKHS.

Among more recent non-quadratic techniques, the most popular ones rely on (TV) regularization which was introduced as a noise-removal technique in [21] and is widely used in computational imaging and compressed sensing, although always in discrete settings. Splines as solutions of TV problems for restricted scenarios have been discussed in [22]. More recently, a RT for the continuous-domain $\mathrm{R}(f) = \|\mathrm{L}f\|_{\mathcal{M}}$ in a general setting has been established in [11], extending the seminal work of Fisher and Jerome [23]. The solution has been shown to be composed of splines that are directly linked to the differential operator L. Other recent contributions on inverse problems in the space of measures include [24]–[28]. In particular, in this paper, we extend the result of [11] to an unconstrained version of the problem. The unconstrained formulation is useful in devising numerical algorithms which are one of the main contributions of our paper. In addition our results are valid for a much larger set of data-fidelity terms than [11]. This is useful in practical scenarios where one may use data-fidelity terms depending on factors like distribution of noise, *etc.*.

B. Notation

Scalar constants, variables, and functions are denoted by oblique letters. For ex. in $f(x) = ax$, a is a constant, x is a variable, and f a function. Vectors are denoted by lowercase bold letters for ex. $\mathbf{a}, \mathbf{z}$. Discrete domain linear operators (or Matrix) are denoted by uppercase bold letters for ex. $\mathbf{H}, \mathbf{L}$. Continuous domain operators are denoted by uppercase straight letters for ex. L. Linear and non-linear functionals are denoted by uppercase straight letters followed by $\{\cdot\}$ and $(\cdot)$, respectively. For ex., $\mathrm{H}\{f\}$ and $\mathrm{R}(f)$. Function spaces are typically denoted by uppercase calligraphic letters for ex. $\mathcal{X}$.

II. Formulation

In our formulation of a linear inverse problem, the signal f is a function of the continuous-domain variable $x \in \mathbb{R}$. The task is then to recover f from the vector of measurements $\mathbf{z} = \mathrm{H}\{f\} + \mathbf{n} \in \mathbb{R}^M$, where $\mathbf{n}$ is an unknown noise component that is typically assumed to be i.i.d. Gaussian.
In the customary discrete formulation, the basis of the recovered function is already chosen and, therefore, all that remains is to recover the expansion coefficients of the signal representation (1). In this scenario, one often includes matrices $\mathbf{H}$ and $\mathbf{L}$ that directly operate on these coefficients. However, for our continuous-domain formulation, the operations have to act directly on the function f. For this reason, we also need the continuous-domain counterparts of the measurement and regularization operators. The entities that enter our formulation are described next.

A. Measurement Operator

The system matrix $\mathbf{H}$ in (2) and (3) is henceforth replaced by the operator $\mathrm{H} : \mathcal{X} \to \mathbb{R}^M$ that maps the continuous-

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

domain functions living in a Banach space $\mathcal{X}$ to the linear measurements $\mathbf{z} \in \mathbb{R}^N$. This operator is described as

$$\mathrm{H}\{f\} = (\langle h_1, f\rangle, \ldots, \langle h_M, f\rangle) = (z_1, \ldots, z_M) = \mathbf{z}, \quad (5)$$

where $\langle h, g\rangle = \int_{\mathbb{R}} h(x)g(x)\,\mathrm{d}x$, which in the case of generalized functions should be interpreted as the duality product. Furthermore, the map $h_m : f \mapsto \langle h_m, f\rangle$ is assumed to be continuous in $\mathcal{X} \to \mathbb{R}$. For example, the components of the measurement operator that samples a function at the locations $x_1, \ldots, x_M$ are represented by $h_m = \delta(\cdot - x_m)$ such that $\langle \delta(\cdot - x_m), f\rangle = f(x_m)$. Similarly, Fourier measurements at pulsations $\omega_1, \ldots, \omega_M$ are obtained by taking $h_m = \mathrm{e}^{-\mathrm{j}\omega_m \cdot}$.

B. Data-Fidelity Term

As extension of the conventional quadratic data-fidelity term $\|\mathbf{z} - \mathbf{H}\mathbf{f}\|_2^2$, we consider a general cost functional $E(\mathbf{z}, \cdot) : \mathbb{R}^M \to \mathbb{R}^+ \cup \{\infty\}$ with some assumptions (see Assumption 2 in Section III), that measures the discrepancy between the given measurements $\mathbf{z}$ and the values $\mathrm{H}\{f\}$ predicted from the reconstruction. A relevant example is the weighted quadratic data-fidelity term, which is often used when the measurement noise is Gaussian with diagonal covariance. Similarly, we can use $\|\mathbf{z} - \mathrm{H}\{f\}\|_1$, for example, when the additive noise is Laplacian. Alternatively, when the measurements are noiseless, we use the indicator function

$$\mathcal{I}(\mathbf{z}, \mathrm{H}\{f\}) = \begin{cases} 0, & \mathbf{z} = \mathrm{H}\{f\} \\ \infty, & \mathbf{z} \neq \mathrm{H}\{f\}, \end{cases} \quad (6)$$

which imposes an exact fit.

C. Regularization Operator

Since the underlying signal is continuously defined, we need to replace the regularization matrix $\mathbf{L}$ in (2) and (3) by a regularization operator $\mathrm{L} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are appropriate (generalized) function spaces to be defined in Section II-E. The typical example that we have in mind is the derivative operator $\mathrm{L} = \mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}x}$. The continuous-domain regularization is then imposed on $\mathrm{L}f$. We assume that the operator L is admissible in the sense of definition 1.

Definition 1. *The operator* $\mathrm{L} : \mathcal{X} \to \mathcal{Y}$ *is called* **spline-admissible** *if*

- *it is linear and shift-invariant;*
- *its null space* $\mathcal{N}_{\mathrm{L}} = \{p \in \mathcal{X} : \mathrm{L}p = 0\}$ *is finite-dimensional;*
- *it admits the Green's function* $\rho_{\mathrm{L}} : \mathbb{R} \to \mathbb{R}$ *with the property that* $\mathrm{L}\rho_{\mathrm{L}} = \delta$.

Given that $\widehat{\mathrm{L}}$ is the frequency response of L, the Green's function can be calculated through the inverse Fourier transform $\rho_{\mathrm{L}} = \mathrm{F}^{-1}\left\{\frac{1}{\widehat{\mathrm{L}}}\right\}$. For example, if $\mathrm{L} = \mathrm{D}$, then $\rho_{\mathrm{D}}(x) = \frac{1}{2}\mathrm{sign}(x)$. Here the Fourier transform, $\mathrm{F} : f \mapsto \mathrm{F}f = \int_{\mathbb{R}} f(x)\mathrm{e}^{-\mathrm{j}x(\cdot)}\,\mathrm{d}x$, is defined when the function is integrable and can be extended in the usual manner to $f \in \mathcal{S}'(\mathbb{R})$ where $\mathcal{S}'(\mathbb{R})$ is 'Schwartz' space of tempered distributions. In cases such as $\rho_{\mathrm{L}} = \mathrm{F}^{-1}\left\{\frac{1}{\widehat{\mathrm{L}}}\right\}$ when the argument is non-integrable, the definition should be seen in terms of generalized Fourier Transform [18, Defintion 8.9] which treats the argument as a distribution.

D. Regularization Norms

Since the optimization is done in the continuous domain, we also have to specify the proper counterparts of the ℓ_2 and ℓ_1 norms, as well as the corresponding vector spaces.

i) Quadratic (or Tikhonov) regularization: $\mathrm{R}_2(f) = \|\mathrm{L}f\|_{L_2}^2$, where

$$\|w\|_{L_2}^2 = \int_{\mathbb{R}} |w(x)|^2\,\mathrm{d}x. \quad (7)$$

ii) Generalized total variation: $\mathrm{R}_1(f) = \|\mathrm{L}f\|_{\mathcal{M}}$, where

$$\|w\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}), \|\varphi\|_\infty = 1} \langle w, \varphi\rangle. \quad (8)$$

There $\mathcal{S}(\mathbb{R})$ is the 'Schwartz' space of smooth and rapidly decaying functions, which is also the dual of $\mathcal{S}'(\mathbb{R})$. Moreover, $\mathcal{M} = \{w \in \mathcal{S}'(\mathbb{R}) \mid \|w\|_{\mathcal{M}} < \infty\}$. In particular, when $w \in L_1 \subset \mathcal{M}$, we have that

$$\|w\|_{\mathcal{M}} = \int_{\mathbb{R}} |w(x)|\,\mathrm{d}x = \|w\|_{L_1}. \quad (9)$$

Yet, we note that $\mathcal{M}$ is slightly larger than L_1 since it also includes the Dirac distribution δ with $\|\delta\|_{\mathcal{M}} = 1$. The popular TV norm is recovered by taking $\|f\|_{\mathrm{TV}} = \|\mathrm{D}f\|_{\mathcal{M}}$ [11].

E. Search Space

The Euclidean search space $\mathbb{R}^N$ is replaced by spaces of functions, namely,

$$\mathcal{X}_2 = \{f : \mathbb{R} \to \mathbb{R} \mid \|\mathrm{L}f\|_{L_2} < +\infty\}, \quad (10)$$

$$\mathcal{X}_1 = \{f : \mathbb{R} \to \mathbb{R} \mid \|\mathrm{L}f\|_{\mathcal{M}} < +\infty\}. \quad (11)$$

In other words, our search (or native) space is the largest space over which the regularization is well defined. It turns out that $\mathcal{X}_2$ and $\mathcal{X}_1$ are Hilbert and Banach spaces, respectively. However, this is nontrivial to see since these spaces contain the null space which makes $\|\mathrm{L}f\|_{L_2}$ and $\|\mathrm{L}f\|_{\mathcal{M}}$ semi-norms. This null-space can be taken care off by using an appropriate inner-product $\langle\cdot,\cdot\rangle_{\mathcal{N}_{\mathrm{L}}}$ (norm $\|\cdot\|_{\mathcal{N}_{\mathrm{L}}}$, respectively) such that $\langle\cdot,\cdot\rangle_{\mathcal{X}_2} = \langle \mathrm{L}\cdot, \mathrm{L}\cdot\rangle + \langle\cdot,\cdot\rangle_{\mathcal{N}_{\mathrm{L}}}$ ($\|\cdot\|_{\mathcal{X}_1} = \|\mathrm{L}\cdot\|_{\mathcal{M}} + \|\cdot\|_{\mathcal{N}_{\mathrm{L}}}$, respectively) is the inner-product (norm, respectively) on $\mathcal{X}_2$ ($\mathcal{X}_1$, respectively). The structure of these spaces has been studied in [11] and is recalled in the supplementary material. As we shall see in Section III, the solution of (4) will be composed of splines; therefore, we also review the definition of splines.

Definition 2 (Nonuniform L-spline). *A function* $f : \mathbb{R} \to \mathbb{R}$ *is called a nonuniform* L*-spline with spline knots* $(x_1, \ldots, x_K)$ *and weights* $(a_1, \ldots, a_K)$ *if*

$$\mathrm{L}f = \sum_{k=1}^{K} a_k \delta(\cdot - x_k). \quad (12)$$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

By solving the differential equation in (12), we find that the generic form of the nonuniform spline f is

$$f = p_0 + \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(\cdot - x_k), \tag{13}$$

where $p_0 \in \mathcal{N}_{\mathrm{L}}$.

III. Theoretical Results

To state our theorems, we need some technical assumptions.

Assumption 1. *Let the search space $\mathcal{X}$ and the regularization space $\mathcal{Y}$ be Banach spaces such that the following holds.*

i) *The functionals h_m for $m \in \{1, \ldots, M\}$ are linear continuous over $\mathcal{X}$ and the vector-valued functional $\mathrm{H} : \mathcal{X} \to \mathbb{R}^M$ gives the linear measurements $f \mapsto \mathrm{H}\{f\} = (\langle h_1, f\rangle, \ldots, \langle h_M, f\rangle)$.*

ii) *The regularization operator $\mathrm{L} : \mathcal{X} \to \mathcal{Y}$ is* spline-admissible*. Its finite-dimensional null space $\mathcal{N}_{\mathrm{L}}$ has the basis $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$.*

iii) *The inverse problem is well posed over the null space. This means that, for any pair $p_1, p_2 \in \mathcal{N}_{\mathrm{L}}$, we have that*

$$\mathrm{H}\{p_1\} = \mathrm{H}\{p_2\} \Leftrightarrow p_1 = p_2. \tag{14}$$

In other words, different null-space functions result in different measurements.

In particular, Condition iii) is equivalent to $\mathcal{N}_{\mathrm{L}} \cap \mathcal{N}_{\mathrm{H}} = \{0\}$, where $\mathcal{N}_{\mathrm{H}}$ is the null space of the vector-valued measurement functional. This property prevents from having a nonzero $f_0 \in \mathcal{N}_{\mathrm{L}} \cap \mathcal{N}_{\mathrm{H}}$ whose addition to any $f \in \mathcal{X}$ can neither be detected by the data-fidelity term nor by the regularization term. This is essential in ensuring the boundedness of the set of the minimizers.

Assumption 2. *For a given $\mathbf{z} \in \mathbb{R}^M$, the functional $E(\mathbf{z}, \cdot) : \mathbb{R}^M \to \mathbb{R}^+ \cup \{\infty\}$ is convex, coercive, and lower semi-continuous on the whole $\mathbb{R}^M$, and is proper (has finite value for at least one input) in the range of* H.

Assumption 2'. *For a given $\mathbf{z} \in \mathbb{R}^M$, the functional $E(\mathbf{z}, \cdot)$ satisfies Assumption 2 as well as one of the following.*

i) *It is strictly convex; or*

ii) *it is an indicator function $I(\mathbf{z}, \cdot)$.*

As we shall see later, stronger results can be derived for the $E(\mathbf{z}, \cdot)$ that satisfy Assumption 2'.

Two remarks are in order. Firstly, the condition of being proper in the range of H implies that there exists an $f \in \mathcal{X}$ such that $E(\mathbf{z}, \mathrm{H}\{f\})$ is finite. Secondly, when $E(\mathbf{z}, \cdot)$ is strictly convex or is such that its range does not include ∞, it is redundant to ensure that it is proper in the range of H.

We now state our two main results. Their proofs are given in Appendix A and Appendix B, respectively.

A. Inverse Problem with Tikhonov/L_2 Regularization

Theorem 3. *Let Assumptions 1 and 2 hold with the search space $\mathcal{X} = \mathcal{X}_2$ and regularization space $\mathcal{Y} = L_2$. Then, the set*

$$\mathcal{V}_2 = \arg\min_{f \in \mathcal{X}_2} \left(E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda \|\mathrm{L}f\|_{L_2}^2 \right) \tag{15}$$

of minimizers is nonempty, convex, and such that any $f_2^* \in \mathcal{V}_2$ is of the form

$$f_2^*(x) = \sum_{m=1}^{M} a_m \varphi_m(x) + \sum_{n=1}^{N_0} b_n p_n(x), \tag{16}$$

where $\varphi_m = \mathrm{F}^{-1}\left\{ \frac{\widehat{h}_m}{|\widehat{\mathrm{L}}|^2} \right\}$, and $\mathbf{a} = (a_1, \ldots, a_M)$, and $\mathbf{b} = (b_1, \ldots, b_{N_0})$ are expansion coefficients such that

$$\sum_{m=1}^{M} a_m \langle h_m, p_n \rangle = 0 \tag{17}$$

for all $n \in \{1, \ldots, N_0\}$. Moreover, if $E(\mathbf{z}, \cdot)$ satisfies Assumption 2' then the minimizer is unique (the set $\mathcal{V}_2$ is singleton).

B. Inverse Problem with gTV Regularization

Theorem 4. *Let Assumptions 1 and 2 hold for the search space $\mathcal{X} = \mathcal{X}_1$ and regularization space $\mathcal{Y} = \mathcal{M}$. Moreover, assume that* H *is weak*-continuous (see Supplementary Material). Then, the set*

$$\mathcal{V}_1 = \arg\min_{f \in \mathcal{X}_1} \left(E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda \|\mathrm{L}f\|_{\mathcal{M}} \right) \tag{18}$$

of minimizers is nonempty, closed-convex, weak-compact, and its extreme points are nonuniform* L*-splines of the form*

$$f_1^*(x) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(x - x_k) + \sum_{n=1}^{N_0} b_n p_n(x) \tag{19}$$

for some $K \leq (M - N_0)$. The unknown knots $(x_1, \ldots, x_K)$, and the expansion coefficients $\mathbf{a} = (a_1, \ldots, a_K)$ and $\mathbf{b} = (b_1, \ldots, b_{N_0})$ are the parameters of the solution with $\|\mathrm{L}f_1^*\|_{\mathcal{M}} = \|\mathbf{a}\|_1$. The solution set $\mathcal{V}_1$ is the closed-convex hull of these extreme points. Moreover, if Assumption 2' is satisfied then all the solutions have the same measurement; i.e., $\mathbf{z}_{\mathcal{V}_1} = \mathrm{H}\{f\}$, $\forall\, f \in \mathcal{V}_1$.

A sufficient condition for weak*-continuity of h_m is $\int_{\mathbb{R}} |h_m(x)|(1 + |x|)^D \mathrm{d}x < \infty$ ([11, Theorem 6]), meaning that h_m should exhibit some minimal decay at infinity (see Section A of the supplementary material for more details). Here $D = \inf\{n \in \mathbb{N} : (\operatorname{ess\,sup}_{x \in \mathbb{R}} \rho_{\mathrm{L}}(1 + |x|)^n) < +\infty\}$. The ideal sampling is feasible as well, provided that the ρ_{L} is continuous; a detailed proof of the weak*-continuity of $\delta(\cdot - x_n)$ for the case $\mathrm{L} = \mathrm{D}^2$ can be found in [29, Proposition 6].

We remark that [11, Theorem 2] is a special case of Theorem 4. The former states the same result as Theorem 4 for the minimization problem

$$\mathcal{V}_1 = \arg\min_{\mathrm{H}\{f\} \in \mathcal{C}} \|\mathrm{L}f\|_{\mathcal{M}}, \tag{20}$$

where $\mathcal{C}$ is feasible, convex, and compact. Feasibility of $\mathcal{C}$ means that the set $\mathcal{C}_{\mathcal{X}_1} = \{f \in \mathcal{X}_1 : \mathrm{H}f \in \mathcal{C}\}$ is nonempty. In our setting, problem (20) can be obtained by using an indicator function over the feasible set $\mathcal{C}$ as the data-fidelity term. However, Theorem 4 covers other more useful cases of E; for example, $\|\mathbf{z} - \mathrm{H}\{f\}\|_1$ and $\|\mathbf{z} - \mathrm{H}\{f\}\|_2^2$. Moreover, as discussed earlier, when data-fidelity terms are strictly convex or do not have ∞ in their range, they are proper in the range of

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

H for any $\mathbf{z} \in \mathbb{R}^M$. This means that they do not require careful selection of $\mathcal{C}$ in order to satisfy the feasibility condition. This is helpful in directly devising and deploying algorithms to find the minimizers.

Also, fundamentally (20) only penalizes the regularization value, whereas Theorem 4 additionally penalizes a data-fidelity term that can recover more desirable solutions. In fact, Theorem 4 also covers cases such as

$$\mathcal{V}_1 = \arg \min_{\mathrm{H}\{f\} \in \mathcal{C}} E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda \|\mathrm{L}f\|_{\mathcal{M}}, \qquad (21)$$

which allow more control than (20) over the data-fidelity of the recovered solution.

C. Illustration with Ideal Sampling

Here, we discuss the regularized case where noisy data points $((x_1, z_1), \ldots, (x_M, z_M))$ are fitted by a function. The measurement functionals in this case are the shifted Dirac impulses $h_m = \delta(\cdot - x_m)$ whose Fourier transform is $\widehat{h}_m(\omega) = \mathrm{e}^{-\mathrm{j}\omega x_m}$. We choose $\mathrm{L} = \mathrm{D}^2$ and $E = \|\mathbf{z} - \mathrm{H}\{f\}\|_2^2$ which satisfies Assumption 2'.i). Here D^2 is the generalized second-order derivative. For the L_2 problem, we have that

$$f_2^* = \arg \min_{f \in \mathcal{X}_2} \left(\sum_{m=1}^{M} |z_m - f(x_m)|^2 + \lambda \|\mathrm{D}^2 f\|_{L_2}^2 \right). \qquad (22)$$

As given in Theorem 3, f_2^* is unique and has the basis function $\varphi_m(x) = \mathrm{F}^{-1}\left\{ \frac{\mathrm{e}^{-\mathrm{j}(\cdot)x_m}}{|-(\cdot)^2|^2} \right\}(x) = \frac{1}{12}|x - x_m|^3$. The resulting solution is piecewise linear. It can be expressed as

$$f_2^*(x) = b_1 + b_2 x + \sum_{m=1}^{M} \frac{1}{12} a_m |x - x_m|^3, \qquad (23)$$

where $b_1 + b_2 x \in \mathcal{N}_{\mathrm{D}^2}$ is a linear function.
We contrast (22) with the gTV version

$$f_1^* = \arg \min_{f \in \mathcal{X}_1} \left(\sum_{m=1}^{M} |z_m - f(x_m)|^2 + \lambda \underbrace{\|\mathrm{D}^2 f\|_{\mathcal{M}}}_{\|\mathrm{D}f\|_{\mathrm{TV}}} \right). \qquad (24)$$

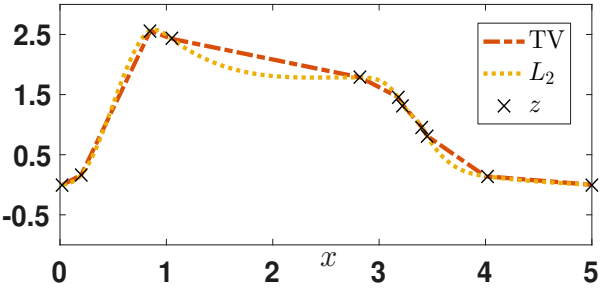
In this scenario, the term $\|\mathrm{D}^2 f\|_{\mathcal{M}}$ is the total variation of the function $\mathrm{D}f$. It penalizes solutions whose slope varies too much from one point to the next.

The Green's function in this case is $\rho_{\mathrm{D}^2} = \frac{|x|}{2}$. Based on Theorem 4, any extreme point of (24) is of the form

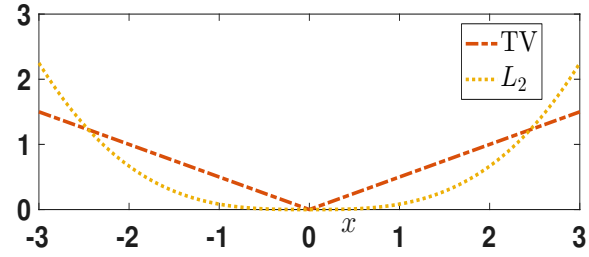
$$f_1^*(x) = b_1 + b_2 x + \frac{1}{2} \sum_{k=1}^{K} a_k' |x - \tau_k|, \qquad (25)$$

which is a piecewise linear function composed of a linear term $b_1 + b_2 x$ and $K \leq (M-1)$ basis functions, $\{|x - \tau_k|\}_{k=1}^K$. The knots (or locations) $\{\tau_k\}_{k=1}^K$ are not fixed a priori and usually differ from the measurement points $\{x_m\}_{m=1}^M$.

The two solutions and their basis functions are illustrated in Figure 1 for specific data. This example demonstrates that the mere replacement of the L_2 penalty with the gTV norm has a fundamental effect on the solution: piecewise-cubic functions having knots at the sampling locations are replaced by piecewise-linear functions with a lesser number



(a) $f_1^*(x)$ and $f_2^*(x)$.



(b) $\rho_{\mathrm{D}^2}(x)$ and $\rho_{\mathrm{D}^{2*}\mathrm{D}^2}(x)$.

Fig. 1: Reconstructions of a signal from nonuniform samples for $\mathrm{L} = \mathrm{D}^2$: (a) Tikhonov ($L_2$) vs. gTV solution, and (b) Corresponding basis functions ρ_{D^2} vs. $\rho_{\mathrm{D}^{2*}\mathrm{D}^2}$.

of adaptive knots. Moreover, in the gTV case, the regularization has been imposed on the generalized second-order derivative of the function $\left(\|\mathrm{D}^2 f\|_{\mathcal{M}}\right)$, which uncovers the innovations $\mathrm{D}^2 f_1^* = \sum_{k=1}^{K} a_k' \delta(\cdot - \tau_k)$. By contrast, when $\mathrm{R}_2(f) = \|\mathrm{D}^2 f\|_{L_2}^2 = \langle \mathrm{D}^{2*}\mathrm{D}^2 f, f \rangle$, the recovered solution is such that $\mathrm{D}^{2*}\mathrm{D}^2 f_2^* = \sum_{m=1}^{M} a_m \delta(\cdot - x_m)$, where $\mathrm{D}^{2*} = \mathrm{D}^2$ is the adjoint operator of D^2. Thus, in both cases, the recovered functions are composed of the Green's function of the corresponding active operators: D^2 vs. $\mathrm{D}^{2*}\mathrm{D}^2 = \mathrm{D}^4$.

IV. Comparison

We now discuss and contrast the results of Theorems 3 and 4. In either case, the solution is composed of a primary component and a null-space component whose regularization cost vanishes.

A. Nature of the Primary Component

1) Shape and Dependence on Measurement Functionals: The solutions for the gTV regularization are composed of atoms within the infinitely large dictionary $\{\rho_{\mathrm{L}}(\cdot - \tau)\}$, $\forall \tau \in \mathbb{R}$, whose shapes depend only on L. In contrast, the L_2 solutions are composed of fixed atoms $\{\varphi_m\}_{m=1}^M$ whose shapes depend on both L and H. As the shape of the atoms of the gTV solutions does not depend on H, this makes it easier to inject prior knowledge in that case.

2) Adaptivity: The weights and the location of the atoms of the gTV solution are adaptive and found through a data-dependent procedure which results in a sparse solution that turns out to be a nonuniform spline. By contrast, the L_2 solution lives in a fixed finite-dimensional space.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

B. Null-Space Component

The second component in either solution belongs to the null space of the operator L. As its contribution to regularization vanishes, the solutions tend to have large null-space components in both instances.

C. Oscillations

The modulus of the Fourier transform of the basis function of the gTV case, $\left|\left\{\frac{1}{\widehat{\mathrm{L}}}\right\}\right|$ typically decays faster than that of the L_2 case, $\left|\left\{\frac{\widehat{h}_m}{|\widehat{\mathrm{L}}|^2}\right\}\right|$. Therefore, the gTV solution exhibits weaker Gibbs oscillations at edges.

D. Uniqueness of the Solution

Our hypotheses guarantee existence. Moreover, the minimizer of the L_2 problem is unique when Assumption 2' is true. By contrast, even for this special category of $E(\mathbf{z},\cdot)$, the gTV problem can have infinitely many solutions, despite all having the same measurements. Remarkably, however, when the gTV solution is unique, it is guaranteed to be an L-spline.

E. Nature of the Regularized Function

One of the main differences between the reconstructions f_2^* and f_1^* is their sparsity. Indeed, $\mathrm{L}f_1^*$ uncovers Dirac impulses situated at $(M-1)$ locations for the gTV case, with $\mathrm{L}f_1^* = \sum_{m=1}^{M-1} a_m \delta(\cdot-\tau_m)$. In return, $\mathrm{L}f_2^*$ is a nonuniform L-spline convolved with the measurement functions, whose temporal support is not localized. This allows us to say that the gTV solution is sparser than the Tikhonov solution.

V. Discretization and Algorithms

We now lay down the discretization procedure that translates the continuous-domain optimization into a more tractable finite-dimensional problem. Theorems 3 and 4 imply that the infinite-dimensional solution lives in a finite-dimensional space that is characterized by the basis functions $\{\varphi_m\}_{m=1}^{M}$ for L_2 and $\{\rho_{\mathrm{L}}(\cdot-\tau_k)\}_{k=1}^{K}$ for gTV, in addition to $\{p_n\}_{n=1}^{N_0}$ as basis of the null space. Therefore, the solutions can be uniquely expressed with respect to the finite-dimensional parameter $\mathbf{a}\in\mathbb{R}^M$ or $\mathbf{a}\in\mathbb{R}^K$, respectively, and $\mathbf{b}\in\mathbb{R}^{N_0}$. Thus, the objective functional $\mathrm{J}_{\mathrm{R}_i}(f|\mathbf{z},\lambda)$, for a given $i\in\{1,2\}$, can be discretized to get the objective functional $\mathrm{J}_{\mathrm{R}_i}(\mathbf{a},\mathbf{b}|\mathbf{z},\lambda)$. Its minimization is done numerically, by expressing $\mathrm{H}\{f\}$ and $\|\mathrm{L}f\|_{L_2}^2$ or $\|\mathrm{L}f\|_{\mathcal{M}}$ in terms of $\mathbf{a}$ and $\mathbf{b}$. We discuss the strategy to achieve $\mathrm{J}_{\mathrm{R}_i}(\mathbf{a},\mathbf{b}|\mathbf{z},\lambda)$ and its minima for the two cases. From now onwards, we will use J_i for $\mathrm{J}_{\mathrm{R}_i}$ where $i\in\{1,2\}$.

A. Tikhonov Regularization

For the L_2 regularization, given $\lambda>0$, the solution

$$f_2^* = \arg\min_{f\in\mathcal{X}_2} \underbrace{\left(E(\mathbf{z},\mathrm{H}\{f\}) + \lambda\|\mathrm{L}f\|_{L_2}^2\right)}_{\mathrm{J}_2(f|\mathbf{z},\lambda)} \tag{26}$$

can be expressed as

$$f_2^* = \sum_{m=1}^{M} a_m\varphi_m + \sum_{n=1}^{N_0} b_n p_n. \tag{27}$$

Recall that $\mathrm{L}^*\mathrm{L}\varphi_m = h_m$, so that

$$\mathrm{L}^*\mathrm{L}f_2^* = \sum_{m=1}^{M} a_m h_m. \tag{28}$$

The corresponding $J_2(\mathbf{z}|\lambda,\mathbf{a},\mathbf{b})$ is then found by expressing $\mathrm{H}\{f_2^*\}$ and $\|\mathrm{L}f_2^*\|_{L_2}^2$ in terms of $\mathbf{a}$ and $\mathbf{b}$. Due to the linearity of the model,

$$\begin{aligned}\mathrm{H}\{f_2^*\} &= \sum_{m=1}^{M} a_m\mathrm{H}\{\varphi_m\} + \sum_{n=1}^{N_0} b_n\mathrm{H}\{p_n\}\\ &= \mathbf{V}\mathbf{a}+\mathbf{W}\mathbf{b},\end{aligned} \tag{29}$$

where $[\mathbf{V}]_{m,n} = \langle h_m,\varphi_n\rangle$ and $[\mathbf{W}]_{m,n} = \langle h_m, p_n\rangle$. Similarly,

$$\langle \mathrm{L}f_2^*, \mathrm{L}f_2^*\rangle = \langle \mathrm{L}^*\mathrm{L}f_2^*, f_2^*\rangle = \left\langle \sum_{m=1}^{M} a_m h_m, f_2^*\right\rangle \tag{30}$$

$$= \mathbf{a}^T\mathbf{V}\mathbf{a} + \mathbf{a}^T\mathbf{W}\mathbf{b} = \mathbf{a}^T\mathbf{V}\mathbf{a}, \tag{31}$$

where (30) uses (28) and where (31) uses the orthogonality property (17), which we can restate as $\mathbf{a}^T\mathbf{W}=\mathbf{0}$. By substituting these reduced forms in (26), the discretized problem becomes

$$f_2^* = \arg\min_{\mathbf{a},\mathbf{b}} \underbrace{\left(E(\mathbf{z},\mathbf{V}\mathbf{a}+\mathbf{W}\mathbf{b}) + \lambda\mathbf{a}^T\mathbf{V}\mathbf{a}\right)}_{\mathrm{J}_2(\mathbf{a},\mathbf{b}|\mathbf{z},\lambda)=\mathrm{J}_2(f_2^*|\mathbf{z},\lambda)}. \tag{32}$$

Due to Assumption 2, this problem is convex. If E is differentiable with respect to the parameters, the solution can be found by gradient descent.

When $E(\mathbf{z},\mathrm{H}\{f\}) = \|\mathbf{z}-\mathrm{H}\{f\}\|_2^2$, the problem is reduced to

$$\arg\min_{\mathbf{a},\mathbf{b}} \underbrace{\left(\|\mathbf{z}-(\mathbf{V}\mathbf{a}+\mathbf{W}\mathbf{b})\|_2^2 + \lambda\mathbf{a}^T\mathbf{V}\mathbf{a}\right)}_{\mathrm{J}_2(\mathbf{a},\mathbf{b}|\mathbf{z},\lambda)} \tag{33}$$

which is very similar to (2). This criterion is convex with respect to the coefficients $\mathbf{a}$ and $\mathbf{b}$. Enforcing that the gradient of J_2 vanishes with respect to $\mathbf{a}$ and $\mathbf{b}$ and setting the gradient to $\mathbf{0}$ then yields M linear equations with respect to the $M+N_0$ variables, while the orthogonality property (17) gives N_0 additional constraints. The combined equations correspond to the linear system

$$\begin{bmatrix}\mathbf{V}+\lambda\mathbf{I} & \mathbf{W}\\ \mathbf{W}^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} = \begin{bmatrix}\mathbf{z}\\ \mathbf{0}\end{bmatrix}. \tag{34}$$

The system matrix so obtained can be proven to be positive definite due to the property of Gram matrices generated in an RKHS and the admissibility condition of the measurement functional (Assumption 1). This ensures that the matrix is always invertible. The consequence is that the reconstructed signal can be obtained by solving a linear system of equation, for instance by QR decomposition or by simple matrix inversion. The derived solution is the same as the least-square solution in [20].

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

B. gTV Regularization

In the case of gTV regularization, the problem to solve is

$$f_1^* = \arg\min_{f\in\mathcal{X}_1} \underbrace{\left(E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda\|\mathrm{L}f\|_{\mathcal{M}}\right)}_{\mathrm{J}_1(f|\mathbf{z},\lambda)}. \tag{35}$$

According to Theorem 4, an extreme-point solution of (35) is

$$f_1^*(x) = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}(x-\tau_k) + \sum_{n=1}^{N_0} b_n p_n(x) \tag{36}$$

and satisfies

$$\mathrm{L}f_1^* = w_1 = \sum_{k=1}^{K} a_k \delta(\cdot - \tau_k) \tag{37}$$

with $K \leq (M - N_0)$. Theorem 4 implies that we only have to recover a_k, τ_k, and the null-space component p to recover f_1^*.

In our experiments, we shall consider the case of measurement functionals whose support is limited to $[0, T]$. We therefore only reconstruct the restriction of the signal in this interval. Since we usually know neither K nor τ_k beforehand, our solution is to quantize the x-axis and look for τ_k in the range $[0, T]$ on a grid with $N \gg K$ points. We control the quantization error with the grid step $\Delta = T/N$.

The discretized problem is then to find $\mathbf{a} \in \mathbb{R}^N$ with fewer than $(M - N_0)$ nonzero coefficients and $\mathbf{b} \in \mathbb{R}^{N_0}$ such that

$$f_{1,\Delta}^*(x) = \sum_{n=0}^{N-1} a_n \rho_{\mathrm{L}}(x - n\Delta) + \sum_{n=1}^{N_0} b_n p_n(x) \tag{38}$$

with $K \leq (M - N_0) \ll N$ nonzero coefficients a_n, satisfies a computationally feasible variant of (35). In other words, we solve the restricted version of (35)

$$\min_{f\in\mathcal{X}_{1,\Delta}} \underbrace{\left(E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda\|\mathrm{L}f\|_{\mathcal{M}}\right)}_{J_{1,\Delta}(\mathbf{z}|\lambda,f)}, \tag{39}$$

where

$$\mathcal{X}_{1,\Delta} = \left\{ \sum_{n=0}^{N-1} a_n \rho_{\mathrm{L}}(\cdot - n\Delta) + \sum_{n=1}^{N_0} b_n p_n \,\middle|\, (\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{N+N_0} \right\}.$$

Similarly to the L_2 case, $\mathrm{J}_{1,\Delta}(\mathbf{a}, \mathbf{b}|\mathbf{z}, \lambda)$ is found by expressing $\mathrm{H}\{f_{1,\Delta}^*\}$ and $\|\mathrm{L}f_{1,\Delta}^*\|_{\mathcal{M}}$ in terms of $\mathbf{a}$ and $\mathbf{b}$. For this, we use the properties that $\mathrm{L}\rho_{\mathrm{L}} = \delta$, $\|\delta\|_{\mathrm{TV}} = 1$, and $\mathrm{L}p_n = 0$ for $n \in [1 \dots N_0]$. This results in

$$\mathrm{H}\{f_{1,\Delta}^*\} = \mathbf{P}\mathbf{a} + \mathbf{Q}\mathbf{b}, \tag{40}$$

$$\|\mathrm{L}f_{1,\Delta}^*\|_{\mathcal{M}} = \|\mathbf{a}\|_1, \tag{41}$$

where $\mathbf{a} = (a_0, \dots, a_{N-1})$, $[\mathbf{P}]_{m,n} = \langle h_m, \rho_{\mathrm{L}}(\cdot - n\Delta)\rangle$ for $n \in [0 \dots N-1]$, $[\mathbf{Q}]_{m,n} = \langle h_m, p_n\rangle$ for $n \in [1 \dots N_0]$, $\|\mathbf{a}\|_1 = \sum_{n=1}^{N} |a_n|$, and where N is the initial number of Green's functions of our dictionary. The new discretized objective functional is

$$f_{1,\Delta}^* = \arg\min_{\mathbf{a},\mathbf{b}} \underbrace{\left(E(\mathbf{z}, (\mathbf{P}\mathbf{a} + \mathbf{Q}\mathbf{b})) + \lambda\|\mathbf{a}\|_1\right)}_{\mathrm{J}_{1,\Delta}(\mathbf{a},\mathbf{b}|\mathbf{z},\lambda) = \mathrm{J}_{1,\Delta}(f_{1,\Delta}^*|\mathbf{z},\lambda)}. \tag{42}$$

Note that (42) is the exact discretization of the infinite-dimensional problem (39). However, additional theories, such as $\Gamma-$convergence [30]–[32], are needed to show that the recovered signal $f_{1,\Delta}^*$ converges (in the weak sense) to one of the solution of the original problem (35) when the discretization step Δ goes to 0 (or when N is large enough). We leave this analysis for the future work.

When E is differentiable with respect to the parameters, a minimum can be found by using proximal algorithms where the slope of $\|\mathbf{a}\|_1$ is defined by a Prox operator. We discuss the two special cases when E is either an indicator function or a quadratic data-fidelity term.

1) Exact Fit with $E = \mathcal{I}(\mathbf{z}, \mathrm{H}\{f\})$: To perfectly recover the measurements, we impose an infinite penalty when the recovered measurements differ from the given ones. In view of (40) and (41), this corresponds to solving

$$(\mathbf{a}^*, \mathbf{b}^*) = \arg\min_{\mathbf{a},\mathbf{b}} \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathbf{P}\mathbf{a} + \mathbf{Q}\mathbf{b} = \mathbf{z}. \tag{43}$$

We then recast Problem (43) as the linear program

$$\begin{aligned}(\mathbf{a}^*, \mathbf{u}^*, \mathbf{b}^*) = \min_{\mathbf{a},\mathbf{u},\mathbf{b}} \sum_{n=1}^{N} u_n \text{ subject to } \mathbf{u} + \mathbf{a} \geq \mathbf{0},\\ \mathbf{u} - \mathbf{a} \geq \mathbf{0},\\ \mathbf{P}\mathbf{a} + \mathbf{Q}\mathbf{b} = \mathbf{z},\end{aligned} \tag{44}$$

where the inequality $\mathbf{x} \geq \mathbf{y}$ between any 2 vectors $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$ means that $x_n \geq y_n$ for $n \in [1 \dots N]$. This linear program can be solved by a conventional simplex or a dual-simplex approach [33], [34].

2) Least Squares Fit with $E = \|\mathbf{z} - \mathrm{H}\{f\}\|_2^2$: When E is a quadratic data-fidelity term, the problem becomes

$$(\mathbf{a}^*, \mathbf{b}^*) = \arg\min_{\mathbf{a},\mathbf{b}} \left(\|\mathbf{z} - (\mathbf{P}\mathbf{a} + \mathbf{Q}\mathbf{b})\|_2^2 + \lambda\|\mathbf{a}\|_1\right), \tag{45}$$

which is more suitable when the measurements are noisy. The discrete version (45) is similar to (3), the fundamental difference being in the nature of the underlying basis function.

The problem is converted into a LASSO formulation [8] by decoupling the computation of $\mathbf{a}^*$ and $\mathbf{b}^*$. Suppose that $\mathbf{a}^*$ is fixed, then $\mathbf{b}^*$ is found by differentiating (45) and equating the gradient to $\mathbf{0}$. This leads to

$$\mathbf{b}^* = \left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T(\mathbf{z} - \mathbf{P}\mathbf{a}^*). \tag{46}$$

Upon substitution in (45), we get that

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \left(\|\mathbf{Q}'\mathbf{z} - \mathbf{Q}'\mathbf{P}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1\right), \tag{47}$$

where $\mathbf{Q}' = \left(\mathbf{I} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right)$ and $\mathbf{I}$ is the $(M \times M)$ identity matrix. Problem (47) can be solved using a variety of optimization techniques such as interior-point methods or proximal-gradient methods, among others. We employ the popular iterative algorithm FISTA [12], which has an $\mathcal{O}(1/t^2)$ convergence rate with respect to its iteration number t. However, in our case, the system matrices are formed by the measurements of the shifted Green's function on a fine grid. This leads to high correlations among the columns and introduces two issues.

- If LASSO has multiple solutions, then FISTA can converge to a solution within the solution set, whose sparsity

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

index is greater than M. A similar type of limitation has been discussed in [35].

- If LASSO has a unique solution, then the convergence to the exact solution can be slow. The convergence rate is inversely proportional to the Lipschitz constant of the gradient of a quadratic loss function $\left(\max \text{Eig}\left(\mathbf{H}^T\mathbf{H}\right)\right)$, which is typically high for the system matrix obtained through our formulation.

We address these issues by using a combination of FISTA and simplex, governed by the following Lemma 5 and Theorem 6. The properties of the solution of the LASSO problem have been discussed in [36], [37], [38]. We quickly recall one of the main results from [36].

Lemma 5 ([36, Lemma 1 and 11])**.** *Let $\mathbf{z} \in \mathbb{R}^M$ and $\mathbf{H} \in \mathbb{R}^{M\times N}$, where $M < N$. Then, the solution set*

$$\alpha_\lambda = \left\{ \arg\min_{\mathbf{a}\in\mathbb{R}^N} \left(\|\mathbf{z}-\mathbf{H}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1\right) \right\} \tag{48}$$

has the same measurement $\mathbf{H}\mathbf{a}^* = \mathbf{z}_0$ for any $\mathbf{a}^* \in \alpha_\lambda$. Moreover, if the solution is not unique, then any two solutions $\mathbf{a}^{(1)}, \mathbf{a}^{(2)} \in \alpha_\lambda$ are such that their mth element satisfies $\left\{ \text{sign}\left(\mathbf{a}_m^{(1)}\right) \text{sign}\left(\mathbf{a}_m^{(2)}\right) \geq 0 \right\}$ for $m \in [1 \dots M]$. In other words, any two solutions have the same sign over their common support.

We use Lemma 5 to infer Theorem 6, whose proof is given in Appendix 6.

Theorem 6. *Let $\mathbf{z} \in \mathbb{R}^M$ and $\mathbf{H} \in \mathbb{R}^{M\times N}$, where $M < N$. Let $\mathbf{z}_{0,\lambda} = \mathbf{H}\mathbf{a}^*, \forall \mathbf{a}^* \in \alpha_\lambda$, be the measurement of the solution set α_λ of the LASSO formulation*

$$\mathbf{a}^* = \arg\min_{\mathbf{a}\in\mathbb{R}^N} \left(\|\mathbf{z}-\mathbf{H}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1\right). \tag{49}$$

Then, the solution $\mathbf{a}^*_{\text{SLP}}$ (obtained using the simplex algorithm) of the linear program corresponding to the problem

$$\mathbf{a}^*_{\text{SLP}} = \arg\min \|\mathbf{a}\|_1 \quad \textit{subject to} \quad \mathbf{H}\mathbf{a} = \mathbf{z}_{0,\lambda} \tag{50}$$

is an extreme point of α_λ. Moreover, $\|\mathbf{a}^*_{\text{SLP}}\|_0 \leq M$.

Theorem 6 helps us to find an extreme point of the solution set α_λ of a given LASSO problem in the case when its solution is non-unique. To that end, we first use FISTA to solve the LASSO problem until it converges to a solution $\mathbf{a}^*_{\text{F}}$. By setting $\mathbf{z}_{0,\lambda} = \mathbf{H}\mathbf{a}^*_{\text{F}}$, Lemma 5 then implies that $\mathbf{H}\mathbf{a} = \mathbf{z}_{0,\lambda}, \forall \mathbf{a} \in \alpha_\lambda$. We then run the simplex algorithm to find

$$\mathbf{a}^*_{\text{SLP}} = \arg\min \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathbf{H}\mathbf{a} = \mathbf{H}\mathbf{a}_{\text{F}},$$

which yields an extreme point of α_λ by Theorem 6.

An example where the LASSO problem has a non-unique solution is shown in Figure 2.b. In this case, FISTA converges to a non-sparse solution with $\|\mathbf{a}^*_{\text{F}}\| > M$, shown as solid stems. This implies that it is not an extreme point of the solution set. The simplex algorithm is then deployed to minimize the ℓ_1 norm such that the measurement $\mathbf{z}_0 = \mathbf{H}\mathbf{a}^*_{\text{F}}$ is preserved. The final solution shown as dashed stems is an extreme point with the desirable level of sparsity. The continuous-domain relation of this example is discussed later.

The solution of the continuous-domain formulation is a convex set whose extreme points are composed of at most M shifted Green's functions. To find the position of these Green's functions, we discretize the continuum into a fine grid and then run the proposed two-step algorithm. If the discretization is fine enough, then the continuous-domain function that corresponds to the extreme point of the LASSO formulation is a good proxy for the actual extreme point of the convex-set solution of the original continuous-domain problem. This makes the extreme-point solutions of the LASSO a natural choice among the solution set. For the case when there is a unique solution but the convergence is too slow owing to the high value of the Lipschitz constant of the gradient of the quadratic loss, the simplex algorithm is used after the FISTA iterations are stopped using an appropriate convergence criterion. For FISTA, the convergence behavior is ruled by the number of iterations t as

$$F(\mathbf{a}_t) - F(\mathbf{a}^*) \leq \frac{C}{(t+1)^2}, \tag{51}$$

where F is the LASSO functional and

$$C = 2\|\mathbf{a}_0 - \mathbf{a}^*\|_2^2 \max \text{Eig}\left(\mathbf{H}^T\mathbf{H}\right) \tag{52}$$

(see [12]). This implies that an ϵ neighborhood of the minima of the functional is obtained in at most $t = \sqrt{C/\epsilon}$ iterations. To ensure convergence, it is also advisable to rely on the modified version of FISTA proposed in [39].

However, there is no direct relation between the functional value and the sparsity index of the iterative solution. Using the simplex algorithm as the next step guarantees the upper bound M on the sparsity index of the solution. Also, $F(\mathbf{a}^*_{\text{SLP}}) \leq F(\mathbf{a}^*_{\text{F}})$. This implies that an ϵ-based convergence criterion, in addition to the sparsity-index-based criterion like $\mathbf{a}^*_{\text{F}} \leq M$, can be used to stop FISTA. Then, the simplex scheme is deployed to find an extreme point of the solution set with a reduced sparsity index.

Note that when $E(\mathbf{z}, \cdot)$ is not strictly convex, the solution set can have non-unique measurements. In that case, it is still possible to further sparsify a recovered solution by using the discussed Simplex approach.

C. Alternative Grid-free Techniques

Our proposed method relies on a grid based discretizatoin of the infinite-dimensional problem. For the sake of completeness, we discuss here alternative techniques for reconstructing continuous-domain sparse signals which employ grid-free optimization. Although elegant, these techniques have a more restricted range of applicability. The Taut-string algorithm (see [40]) can fit L-splines for $\text{L} = \text{D}^n$ but is devised for ideal sampling only. In [25], [26], [35], [41]–[43] the dual problem is considered for the optimization with an added emphasis on recovering the ground-truth signal. These methods, however, only deal with $\text{L} = \text{Id}$ and limited measurement operators.

Recently, in [28], motivated from [11], results for more general L and H have been derived. There the optimization is carried out in two steps; firstly, a finite dimensional dual problem involving two infinite-dimensional convex-constraints-sets is solved; secondly, the support of this solution is identified which is finally used to solve a finite-dimensional primal

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

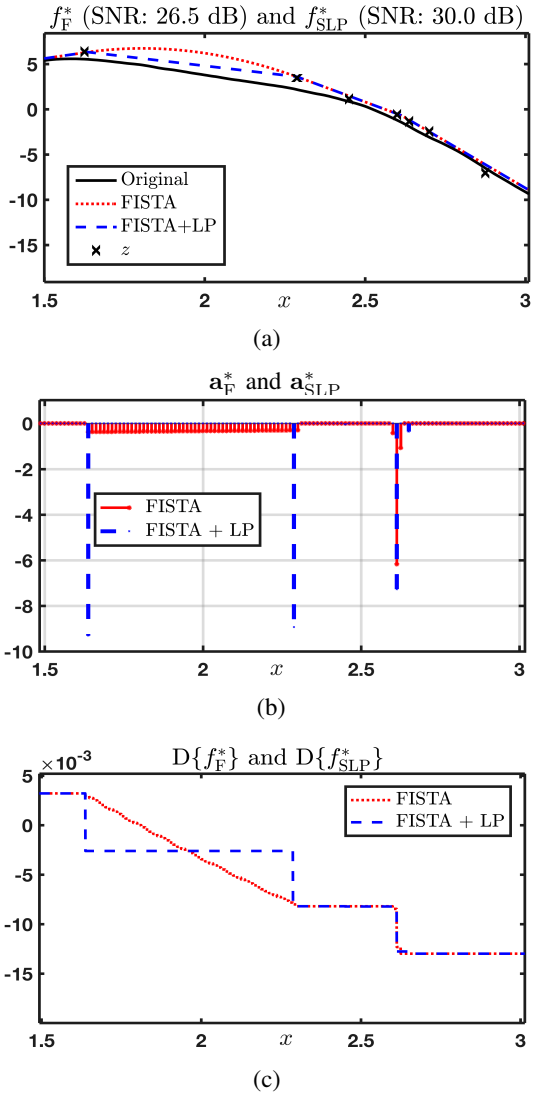


Fig. 2: Illustration of inability of FISTA to deliver a sparse solution : (a) comparison of solutions, f_{F}^* vs. f_{SLP}^* for continuous-domain gTV problem, (b) signal innovations with sparsity index 64 ($> M$) and 21 ($< M$), respectively, and (c) derivative of the two solutions. The two signal innovations in (b) are solutions of the same Lasso problem, but only $\mathbf{a}_{\mathrm{SLP}}^*$ is an extreme point of the solution set. The original signal is a second-order process ($\mathrm{L} = \mathrm{D}^2$) and the measurements are $M = 30$ nonuniform noisy samples (SNR = 40 dB). The parameters are $\lambda = 0.182$, $N = 400$, and grid step $\Delta = \frac{1}{80}$.

problem. Remarkably, for some specific cases, solving each of these steps is feasible which results in an exact finite-dimensional formulation (see for example [28, Section 2.4.2 and 2.4.3]).

VI. Illustrations

We discuss the results obtained for the cases when the measurements are random samples either of the signal itself or of its continuous-domain Fourier transform. The operators of interest are $\mathrm{L} = \mathrm{D}$ and $\mathrm{L} = \mathrm{D}^2$. The ground truth (GT) signal f_{GT} is solution of the stochastic differential equation $\mathrm{L}f_{\mathrm{GT}} = w_{\mathrm{GT}}$ [44] for the two cases when w_{GT} is

- **Impulsive Noise**. Here, the innovation w_{GT} is a compound-Poisson noise with Gaussian jumps, which corresponds to a sum of Dirac impulses whose amplitudes follow a Gaussian distribution. The corresponding process f_{GT} has then the particularity of being piecewise smooth [45]. This case is matched to the regularization operator $\|\mathrm{L}f\|_{\mathcal{M}}$ and is covered by Theorem 4 which states that the minima f_1^* for this regularization case is such that
$$w_1^* = \mathrm{L}f_1^* = \sum_{k=1}^{K} a_k \delta(\cdot - x_k), \tag{53}$$
which is a form compatible with a realization of an impulsive white noise.
- **Gaussian White Noise**. This case is matched to the regularization operator $\|\mathrm{L}f\|_{L_2}$. Unlike the impulsive noise, $w_2^* = \mathrm{L}f_2^*$ is not localized to finite points and therefore is a better model for the realization of a Gaussian white noise.

In all experiments, we also constrain the test signals to be compactly supported. This can be achieved by putting linear constraints on the innovations of the signal. In Sections VI-A and VI-C, we confirm experimentally that matched regularization recovers the test signals better than non-matched regularization. While reconstructing the Tikhonov and gTV solutions when the measurements are noisy, the parameter λ in (34) and (45) is tuned using a grid search to give the best recovered SNR.

A. Random Sampling

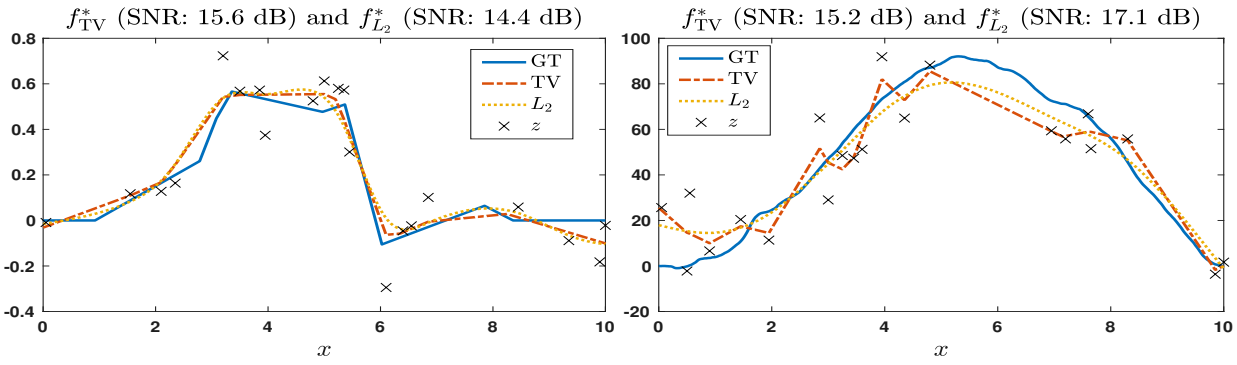
In this experiment, the measurement functionals are Dirac impulses with the random locations $\{x_m\}_{m=1}^M$. The regularization operator is $\mathrm{L} = \mathrm{D}^2$. It corresponds to $\rho_{\mathrm{D}^2}(x) = -\frac{1}{2}|x|$ and $\varphi_{\mathrm{D}^2}(x) = (\rho_{\mathrm{L}^*\mathrm{L}} * h_m)(x) = |x - x_m|^3/12$. The null space is $\mathcal{N}_{\mathrm{D}^2} = \mathrm{span}\{1, x\}$ for this operator. This means that the gTV-regularized solution is piecewise linear and that the L_2-regularized solution is piecewise cubic. We compare in Figures 3.a and 3.b the recovery from noiseless samples of a second-order process, referred to as ground truth (GT). It is composed of sparse (impulsive Poisson) and non-sparse (Gaussian) innovations, respectively [46]. The sparsity index—the number of impulses or non-zero elements—for the original sparse signal is 9. The solution for the gTV case is recovered with $\Delta = 0.05$ and $N = 200$. The sparsity index of the gTV solution for the sparse and Gaussian cases are 9 and 16, respectively. As expected, the recovery of the gTV-regularized reconstruction is better than that of the L_2-regularized solution when the signal is sparse. For the Gaussian case, the situation is reversed.

B. Multiple Solutions

We discuss the case when the gTV solution is non-unique. We show in Figure 2.a examples of solutions of the gTV-

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing



(a) Sparse Signal (b) Gaussian Signal

Fig. 3: Recovery of sparse (a) and Gaussian (b) second-order processes (GT) using $\mathrm{L} = \mathrm{D}^2$ from their nonuniform samples corrupted with 40 dB measurement noise.

regularized random-sampling problem obtained using FISTA alone (f^*_{F}) and FISTA + simplex (linear programming, f^*_{SLP}). In this case, $M = 30$, $\mathrm{L} = \mathrm{D}^2$, and $\lambda = 0.182$. The continuous-domain functions f^*_{F} and f^*_{SLP} have basis functions whose coefficients are the (non-unique) solutions of a given LASSO problem, as shown in Figure 2.b. The ℓ_1 norms of the corresponding coefficients are the same. Also, it holds that

$$\|\mathrm{D}^2 f^*_{\mathrm{F}}\|_{\mathcal{M}} = \|\mathrm{D}^2 f^*_{\mathrm{SLP}}\|_{\mathcal{M}} = \|\mathrm{D} f^*_{\mathrm{F}}\|_{\mathrm{TV}} = \|\mathrm{D} f^*_{\mathrm{SLP}}\|_{\mathrm{TV}}, \quad (54)$$

which implies that the TV norm of the slope of f^*_{F} and f^*_{SLP} are the same. This is evident from Figure 2.c. The arc-length of the two curves are the same. The signal f^*_{SLP} is piecewise linear ($21 < M$), carries a piecewise-constant slope, and is by definition, a non-uniform spline of degree 1. By contrast, f^*_{F} has many more knots and even sections whose slope appears to be piecewise-linear.

Theorem 4 asserts that the extreme points of the solution set of the gTV regularization need to have fewer than M knots. Remember that f^*_{SLP} is obtained by combining FISTA and simplex; this ensures that the basis coefficients of f^*_{SLP} are the extreme points of the solution set of the corresponding LASSO problem (Theorem 6) and guarantees that the number of knots is smaller than M.

This example shows an intuitive relationship between the continuous-domain and the discrete-domain formulations of inverse problems with gTV and ℓ_1 regularization, respectively. The nature of the continuous-domain solution set and its extreme points resonates with its corresponding discretized version. In both cases, the solution set is convex and the extreme points are sparse.

C. Random Fourier Sampling

Let now the measurement functions be $h_m(x) = \mathrm{rect}\left(\frac{x}{T}\right) \mathrm{e}^{-\mathrm{j}\omega_m x}$, where T is the window size. The samples are thus random samples of the continuous-domain Fourier transform of a signal restricted to a window. For the regularization operator $\mathrm{L} = \mathrm{D}$, the Green's function is $\rho_{\mathrm{D}}(x) = \frac{1}{2}\mathrm{sign}(x)$ and the basis is $\varphi_{\mathrm{D},m}(x) = \left(\frac{1}{2}|\cdot| * h_m\right)(x)$. Figure 4.a and 4.b correspond to a first-order process with sparse and Gaussian innovations, respectively. The grid step $\Delta = 0.05$, $M = 41$, and $N = 200$. The sparsity index of the gTV solution for the sparse and Gaussian cases is 36 and 39, respectively. For the original sparse signal (GT), it is 7. The oscillations of the solution in the L_2-regularized case are induced by the sinusoidal form of the the measurement functionals. This also makes the L_2 solution intrinsically smoother than its gTV counterpart. Also, the quality of the recovery depends on the frequency band used to sample.

In Figures 4.c and 4.d, we show the zoomed version of the recovered second-order process with sparse and Gaussian innovations, respectively. The grid step is $\Delta = 0.05$, $M = 41$ and $N = 200$. The operator $\mathrm{L} = \mathrm{D}^2$ is used for the regularization. This corresponds to $\rho_{\mathrm{D}^2}(x) = \frac{|x|}{2}$ and $\varphi_{\mathrm{D}^2,m}(x) = \left(\frac{1}{12}|\cdot|^3 * h_m\right)(x)$. The sparsity index of the gTV solution in the sparse and Gaussian cases is 10 and 36, respectively. For the original sparse signal (GT), it is 10. Once again, the recovery by gTV is better than by L_2 when the signal is sparse. In the Gaussian case, the L_2 solution is better.

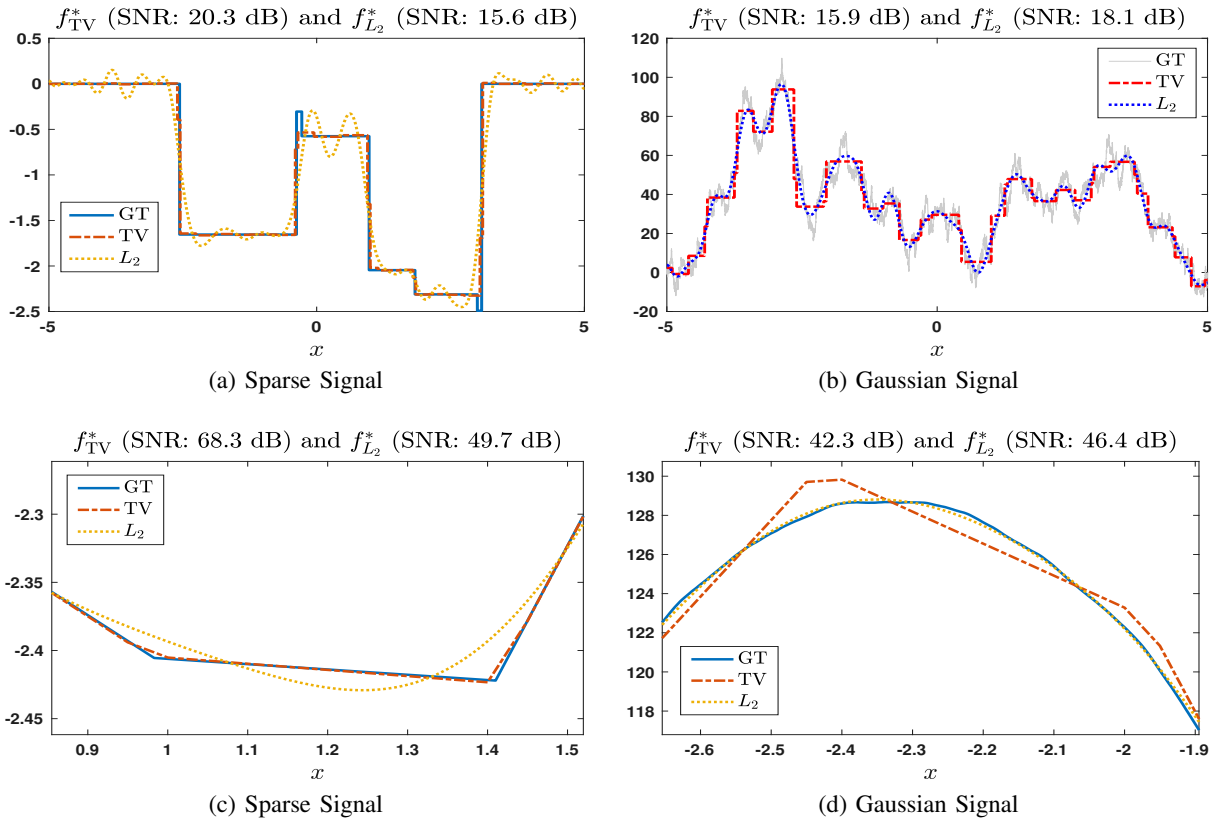
The effect of sparsity on the recovery of signals from their noiseless and noisy (40 dB SNR) Fourier samples are shown in Table 1. The sample frequencies are kept the same for all the cases. Here, $M = 41$, $N = 200$, $T = 10$, and the grid step $\Delta = 0.05$. We observe that reconstruction performances for random processes based on impulsive noise are comparable to that of Gaussian processes when the number of impulses increases. This is reminiscent of the fact that generalized-Poisson processes with Gaussian jumps are converging in law to corresponding Gaussian processes [47].

VII. Conclusion

We have shown that the formulation of continuous-domain linear inverse problems with Tikhonov- and total-variation-based regularizations leads to spline solutions. The nature of these splines is dependent on the Green's function of the regularization operator ($\mathrm{L}^*\mathrm{L}$) and L for Tikhonov and total variation, respectively. The former is better to reconstruct smooth signals; the latter is an attractive choice to reconstruct signals with sparse innovations. Representer theorems for the two cases come handy in the numerical reconstruction of the solution. They allow us to reformulate the infinite-dimensional

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing



(a) Sparse Signal (b) Gaussian Signal

(c) Sparse Signal (d) Gaussian Signal

Fig. 4: Recovery of first-order (first row) and second-order (second row) processes from their random noiseless Fourier samples. In all the cases, $M = 41$ and $N = 200$. In the interest of clarity, (c) and (d) contain the zoomed versions of the actual signals.

optimization as a finite-dimensional parameter search. The formulations and the results of this paper are summarized in Figure 5. The main aim of this paper was to compare the solutions of the two regularizations. We expect that similar results exist in higher dimensions since the theory can be generalized. However, the computations can also be expected to be challenging for signals defined over $\mathbb{R}^d$ with $d > 1$, for example, when considering images rather than signals.

Appendix A
Proof of Theorem 3

A. Abstract Representer Theorem

The result presented in this section is preparatory to Theorem 3. It is classical for Hilbert spaces [18, Theorem 16.1]. We give its proof for the sake of completeness.

Theorem 7. *Let $\mathcal{X}$ be a Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and let $h_1, \ldots, h_M \in \mathcal{X}'$ be linear and continuous functionals. Let $\mathcal{C} \in \mathbb{R}^M$ be a feasible convex compact set, meaning that there exists at least a function $f \in \mathcal{X}$ such that $\mathrm{H}\{f\} \in \mathcal{C}$. Then, the minimizer*

$$f^* = \arg\min_{f \in \mathcal{X}} \|f\|^2_{\mathcal{X}} \text{ s.t. } \mathrm{H}\{f\} \in \mathcal{C} \tag{55}$$

exists, is unique, and can be written as

$$f^* = \sum_{m=1}^{M} a_m h_m^{\#} \tag{56}$$

for some $\{a_m\}_{m=1}^M \in \mathbb{R}$, where $h_m^{\#} = \Pi h_m$ and $\Pi : \mathcal{X}' \to \mathcal{X}$ is the Riesz map of $\mathcal{X}$.

Proof. The feasibility of the set $\mathcal{C}$ implies that the set $\mathcal{C}_{\mathcal{X}} = \mathrm{H}^{-1}(\mathcal{C}) = \{f \in \mathcal{X} : \mathrm{H}\{f\} \in \mathcal{C}\} \in \mathcal{X}$, is nonempty. Since H is linear and bounded and since $\mathcal{C}$ is convex and compact, its preimage $\mathcal{C}_{\mathcal{X}}$ is also convex and closed. By Hilbert's projection theorem [48], the solution f^* exists and is unique as it is the projection of the null function onto $\mathcal{C}_{\mathcal{X}}$. Let the measurement of this unique point f^* be $\mathbf{z}_0 = \mathrm{H}\{f^*\}$.
The Riesz representation theorem states that $\langle h_m, f \rangle = \langle h_m^{\#}, f \rangle_{\mathcal{X}}$ for every $f \in \mathcal{X}$, where $h_m^{\#} \in \mathcal{X}$ is the unique Riesz conjugate of the functional h_m. We then uniquely decompose f^* as $f^* = f^{*\perp} + \sum_{m=1}^M a_m h_m^{\#}$, where $f^{*\perp}$ is orthogonal to the span of the $h_m^{\#}$ with respect to the inner product on $\mathcal{X}$ *i.e.*, $\mathrm{H}\{f^{*\perp}\} = \mathbf{0}$. The orthogonality also implies that

$$\|f^*\|^2_{\mathcal{X}} = \left\|f^{*\perp}\right\|^2_{\mathcal{X}} + \left\|\sum_{m=1}^{M} a_m h_m^{\#}\right\|^2_{\mathcal{X}}. \tag{57}$$

This means that the minimum norm is reached when $f^{*\perp} = 0$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

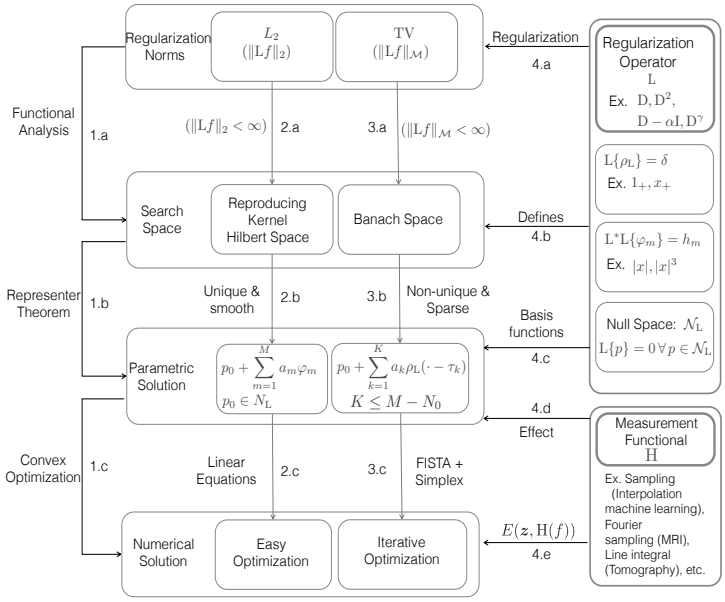


Fig. 5: Summary of the whole scheme. The regularization operator with a given norm {4.a} defines the search space for the solution{1.a, 4.b}. Representer theorems then give the parametric representation of the solution {1.b}. The numerical solution is then recovered by optimizing over the parameters to minimize $\mathrm{J}_{\mathrm{R}}(f|\mathbf{z})$ {1.c}.

No. of impulses	Sparsity	D TV	D L_2	D^2 TV	D^2 L_2
10	Strong	**19.60**	15.7	**52.08**	41.54
100	Medium	**16.58**	16.10	**41.91**	41.26
2000	Low	14.45	**16.14**	39.68	**41.40**
-	Gaussian	14.30	**16.32**	40.05	**41.23**

No. of impulses	Sparsity	D TV	D L_2	D^2 TV	D^2 L_2
10	Strong	**17.06**	11.52	**25.55**	24.60
100	Medium	**13.24**	10.94	**24.44**	24.24
2000	Low	10.61	**11.13**	25.80	**26.19**
-	Gaussian	10.40	**11.10**	24.95	**25.48**

TABLE I: Comparison of TV and L_2 recovery from their (left table) noiseless and (right table) noisy (with 40 dB SNR) random Fourier samples. The results have been averaged over 40 realizations.

while keeping $\mathrm{H}\{f^*\} = \mathbf{z}_0$, implying the form (56) of the solution.

□

B. Proof of Theorem 3

The proof of Theorem 3 has two steps. We first show that if Assumption 2 holds, then there is at least one solution and, moreover, if Assumption 2' holds, then the solution is unique. After this, we use Theorem 7 to deduce the form of the solution.

Existence of the Solution. We use the classical result on Hilbert spaces which states that a proper, coercive, lsc, and convex objective functional over a Hilbert space has a nonempty and convex set of minimizers [49].

Properness: By Assumption 2, $E(\mathbf{z}, \cdot)$ is proper. The regularization $\|\mathrm{L}f\|_{L_2}^2$ is proper by the definition of $\mathcal{X}_2$. This means that $\mathrm{J}_2(\cdot|\mathbf{z})$ is proper in $\mathcal{X}_2$.

Lower semi-continuity: $E(\mathbf{z}, \cdot)$ is lsc in $\mathbb{R}^M$, and $\mathrm{H} : \mathcal{X}_2 \to \mathbb{R}^M$ is continuous. Therefore, $E(\mathbf{z}, \mathrm{H}\{\cdot\})$ is lsc over $\mathcal{X}_2$. Similarly, by composition $f \mapsto \|\mathrm{L}f\|_{L_2}$ is continuous, hence lsc over $\mathcal{X}_2$. Since $\mathrm{J}_2(\cdot|\mathbf{z})$ is the sum of two lsc functionals, it is lsc as well.

Convexity: $E(\mathbf{z}, \cdot)$ and $\|\cdot\|_{L_2}^2$ are convex, and H and L are linear. Therefore, $\mathrm{J}_2(\cdot|\mathbf{z}) = E(\mathbf{z}, \mathrm{H}\{\cdot\}) + \|\mathrm{L}\cdot\|_{L_2}^2$ is convex too.

Coercivity: The measurement operator H is continuous and linear from $\mathcal{X}_2$ to $\mathbb{R}^M$; hence, there exists a constant A such that

$$\|\mathrm{H}\{f\}\|_2 \leq A\|f\|_{\mathcal{X}_2} \tag{58}$$

for every $f \in \mathcal{X}_2$. Likewise, the condition $\mathrm{H}\{p\} = \mathrm{H}\{q\} \Rightarrow p = q$ for $p, q \in \mathcal{N}_{\mathrm{L}}$ implies the existence of $B > 0$ such that [11, Proposition 8]

$$\|\mathrm{H}\{p\}\|_2 \geq B\|p\|_{\mathcal{N}_{\mathrm{L}}} \tag{59}$$

for every $p \in \mathcal{N}_{\mathrm{L}}$. As presented in the supplementary material (see [11] for more details), the search space $\mathcal{X}_2$ is a Hilbert space for the Hilbertian norm

$$\|f\|_{\mathcal{X}_2} = \|\mathrm{L}f\|_{L_2} + \|\mathrm{P}f\|_{\mathcal{N}_{\mathrm{L}}} \tag{60}$$

with P being the projector on $\mathcal{N}_{\mathrm{L}}$ introduced in (74). We set $p = \mathrm{P}f$ and $g = f - p$. Then, $g \in \mathcal{X}_2$ satisfies $\mathrm{L}g = \mathrm{L}f$ and

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

$\mathrm{P}g = 0$, and hence

$$\|g\|_{\mathcal{X}_2} = \|\mathrm{L}g\|_{L_2} + \|\mathrm{P}g\|_{\mathcal{N}_\mathrm{L}} = \|\mathrm{L}f\|_{L_2}. \tag{61}$$

Now consider a sequence of (generalized) functions $f_m \in \mathcal{X}_2, m \in \mathbb{N}$ such that $\|f_m\|_{\mathcal{X}_2} \to \infty$. We set $p_m = \mathrm{P}f_m$ and $g_m = f_m - p_m$. Assume by contradiction that $\mathrm{J}_2(f_m|\mathbf{z})$ is bounded. Then, $\|\mathrm{L}f_m\|_{L_2}$ and $\|\mathrm{H}\{f_m\}\|_2$ are bounded (for the latter, we use that $E(\mathbf{z},\cdot)$ is coercive). However, we have

$$\begin{aligned}\|\mathrm{H}\{f_m\}\|_2 &\geq \|\mathrm{H}\{p_m\}\|_2 - \|\mathrm{H}\{g_m\}\|_2 &(62)\\ &\geq B\|p_m\|_{\mathcal{N}_\mathrm{L}} - A\|g_m\|_{\mathcal{X}_2} &(63)\\ &= B\|f_m\|_{\mathcal{X}_2} - (A+B)\|\mathrm{L}f_m\|_{L_2} &(64)\end{aligned}$$

where we used respectively the triangular inequality in (62), the inequalities (58) and (59) in (63), and the relations $\|p_m\|_{\mathcal{N}_\mathrm{L}} = \|f_m\|_{\mathcal{X}_2} - \|\mathrm{L}f_m\|_{L_2}$ and $\|g_m\|_{\mathcal{X}_2} = \|\mathrm{L}f_m\|_{L_2}$ in (64). Since $\|\mathrm{L}f_m\|_{L_2}$ is bounded and $\|f_m\|_{\mathcal{X}_2} \to \infty$, we deduce that $\|\mathrm{H}\{f_m\}\|_2 \to \infty$, which is known to be false. Finally, we obtain a contradiction, proving the coercivity.

Since, $\mathrm{J}_2(\cdot|\mathbf{z})$ is proper, lsc, convex, and coercive on $\mathcal{X}_2$, therefore, it has at least one minimizer.

Uniqueness of the Solution. We now prove that if $E(\mathbf{z},\cdot)$ satisfies Assumption 2' then the solution is unique. We first show this for the case when Assumption 2'.i) is satisfied. We already know that the solution set is nonempty. It is then clear that the uniqueness is achieved if $\mathrm{J}_2(\cdot|\mathbf{z})$ is strictly convex. We now prove the convex functional $\mathrm{J}_2(\cdot|\mathbf{z})$ is actually strictly convex.

For $\beta \in (0,1)$, $f_A, f_B \in \mathcal{X}_2$, we denote $f_{AB} = \beta f_A + (1-\beta)f_B$. Then, the equality case $\mathrm{J}_2(f_{AB}|\mathbf{z}) = \beta\mathrm{J}_2(f_A|\mathbf{z}) + (1-\beta)\mathrm{J}_2(f_B|\mathbf{z})$ implies that $E(\mathbf{z}, f_{AB}) = \beta E(\mathbf{z}, f_A) + (1-\beta)E(\mathbf{z}, f_B)$ and $\|\mathrm{L}f_{AB}\|_{L_2} = \beta\|\mathrm{L}f_B\|_{L_2} + (1-\beta)\|\mathrm{L}f_B\|_{L_2}$, since the two parts of the functional are themselves convex. The strict convexity of $E(\boldsymbol{z},\cdot)$ and the norm $\|\cdot\|_2$ then implies that

$$\mathrm{L}f_A = \mathrm{L}f_B \text{ and } \mathrm{H}\{f_A\} = \mathrm{H}\{f_B\} \tag{65}$$

and, therefore, $(f_A - f_B) \in \mathcal{N}_\mathrm{L} \cap \mathcal{N}_\mathrm{H}$. Since $\mathcal{N}_\mathrm{L} \cap \mathcal{N}_\mathrm{H} = \mathbf{0}$ by Assumption 1, therefore, $f_A = f_B$. This demonstrates that $\mathrm{J}_2(\cdot|\mathbf{z})$ is strictly convex.

For Assumption 2'.ii), that is when $E(\mathbf{z},\cdot) = I(\mathbf{z},\cdot)$, the solution set can be written as

$$\mathcal{V}_2 = \arg\min_{f \in \mathrm{H}^{-1}\{\mathbf{z}\}} \|\mathrm{L}f\|_{L_2}^2. \tag{66}$$

where the set $\mathrm{H}^{-1}\{\mathbf{z}\} = \{f \in \mathcal{X}_2 : \mathrm{H}\{f\} = \mathbf{z}\}$ is nonempty since we assumed $I(\mathbf{z},\cdot)$ to be proper in the range of H.

According to [17, Theorem 1.1 and 1.2] given that the range of $\mathrm{L} : \mathcal{X}_2 \to L_2$ is closed in L_2, $\mathcal{V}_2$ in (66) is singleton . As discussed in the supplementary material, given any $w \in L_2$, we can always find an $f \in \mathcal{X}_2$ such that $\mathrm{L}f = w$. This means that the range of L is the whole L_2 which is clearly closed in L_2.

Form of the Minimizer. We first take the case when E satisfies Assumption 2'. Let f_2^* be the unique solution and $\mathbf{z}_0 = \mathrm{H}\{f_2^*\}$. One decomposes again $\mathcal{X}_2$ as the direct sum $\mathcal{X}_2 = \mathcal{Q} \oplus \mathcal{N}_\mathrm{L}$, where

$$\mathcal{Q} = \{f \in \mathcal{X}_2 : \ \langle f, p\rangle_{\mathcal{X}_2} = 0, \, \forall p \in \mathcal{N}_\mathrm{L}\}$$

is the Hilbert space with norm $\|\mathrm{L}\cdot\|_{L_2}$. In particular, we have that $f_2^* = q^* + p^*$ with $q^* \in \mathcal{Q}$ and $p^* \in \mathcal{N}_\mathrm{L}$.

Consider the optimization problem

$$\min_{g \in \mathcal{Q}} \|\mathrm{L}g\|_{L_2}^2 \text{ s.t. } \mathrm{H}\{g\} = (\mathbf{z}_0 - \mathrm{H}\{p^*\}). \tag{67}$$

According to Theorem 7, this problem admits a unique minimizer g^* such that $\Pi^{-1}g^* \in \mathcal{Q}' \cap \mathrm{Span}\{h_m\}_{m=1}^M$ where $\Pi^{-1} : \mathcal{X} \to \mathcal{X}'$ is the inverse of the Riesz map $\Pi : \mathcal{X}' \to \mathcal{X}$ and $Q' = \Pi^{-1}\mathcal{Q}$. The set $\mathcal{Q}' \cap \mathrm{Span}\{h_m\}_{m=1}^M$ is represented by $\sum_{m=1}^M a_m h_m$, where $\sum_m a_m \langle h_m, p\rangle = 0$ for every $p \in \mathcal{N}_\mathrm{L}$.

However, by definition, the function q^* also satisfies $\mathrm{H}\{q^*\} = (\mathbf{z}_0 - \mathrm{H}\{p^*\})$. Moreover, $\|\mathrm{L}q^*\|_{L_2}^2 \leq \|\mathrm{L}g^*\|_{L_2}^2$; otherwise, the function $\tilde{f} = g^* + p^* \in \mathcal{X}_2$ would satisfy $J_2(\tilde{f}|\mathbf{z}) < J_2(f_2^*|\mathbf{z})$, which is impossible. However, since (67) has a unique solution, we have $q^* = g^*$.

This proves that $f_2^* = \Pi\left\{\sum_{m=1}^M a_m h_m\right\} + p^*$. For any $q' \in \mathcal{Q}'$ the Riesz map $\Pi q' = q' * \rho_{\mathrm{L}*\mathrm{L}} + p_{q'}$ for some $p_{q'} \in \mathcal{N}_\mathrm{L}$ [17], [18]. Here $\rho_{\mathrm{L}^*\mathrm{L}}$ is the Green's function of the operator $(\mathrm{L}^*\mathrm{L})$ (see Definition 1). Therefore,

$$f_2^* = p_0 + \rho_{\mathrm{L}^*\mathrm{L}} * \left\{\sum_{m=1}^M a_m h_m\right\} \tag{68}$$

where $p_0 = (p_{q'} + p^*) \in \mathcal{N}_\mathrm{L}$ and where $\sum_m a_m \langle h_m, p\rangle = 0$ for every $p \in \mathcal{N}_\mathrm{L}$.

The component $\rho_{\mathrm{L}^*\mathrm{L}} * \left\{\sum_{m=1}^M a_m h_m\right\}$ in (68) can be written as, $\sum_{m=1}^M a_m \varphi_m$ provided that $\varphi_m = \rho_{\mathrm{L}^*\mathrm{L}} * h_m = \mathrm{F}^{-1}\left\{\frac{\widehat{h}_m}{|\widehat{\mathrm{L}}|^2}\right\}$ is well-defined. To show that this is the case, we decompose $h_m = \mathrm{Proj}_{\mathcal{Q}'}\{h_m\} + \mathrm{Proj}_{\mathcal{N}_\mathrm{L}'}\{h_m\}$ where $\mathrm{Proj}_{\mathcal{Q}'}$ and $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}'}$ are the projection operators on $\mathcal{Q}'$ and $\mathcal{N}_\mathrm{L}'$, respectively. Since, $\mathrm{Proj}_{\mathcal{Q}'}\{h_m\} \in \mathcal{Q}'$, as discussed earlier, $\rho_{\mathrm{L}*\mathrm{L}} * \ \mathrm{Proj}_{\mathcal{Q}'}\{h_m\}$ is well-defined.

Now, one can always select a basis $\{p_n\}_{n=1}^{N_0}$ such that $\mathcal{N}_\mathrm{L}' = \mathrm{Span}\{\phi_n\}_{n=1}^{N_0}$ with $\phi_n = \delta(\cdot - x_n)$ and $\langle \phi_m, p_n\rangle = \delta[m-n]$. The other component $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}'}\{h_m\} = \sum_{n=1}^{N_0} c_n \phi_n$ where $c_n = \langle h_m, p_n\rangle$. Therefore, $\rho_{\mathrm{L}*\mathrm{L}} * \mathrm{Proj}_{\mathcal{N}_\mathrm{L}'}\{h_m\}$ is a linear combination of shifted Green's functions, which proves that $\varphi_m = \mathrm{F}^{-1}\left\{\frac{\widehat{h}_m}{|\widehat{\mathrm{L}}|^2}\right\}$ is well defined.

For general case, when Assumption 2 is satisfied, we see that any solution $f_2^* \in \mathcal{V}_2$ also minimizes the following

$$\min_{f \in \mathrm{H}^{-1}\{\mathrm{H}\{f_2^*\}\}} \|\mathrm{L}f\|_{L_2}. \tag{69}$$

As discussed earlier, the minimizer of (69) is unique so that it is clearly f_2^*. We now use the same reasoning as for the cases of Assumption 2' to show that f_2^* takes the form (16). This concludes the proof.

Note that, even in the absence of convexity of $E(\mathbf{z},\cdot)$, results on the form of the solution can still be obtained.

Appendix B
Proof of Theorem 4

Similarly to the L_2 case, the proof has two steps. We first show that the set of minimizers is nonempty. We then connect the optimization problem to the one studied in [11, Theorem 2] to deduce the form of the extreme points. The functional

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

to minimize is $\mathrm{J}_1(f|\mathbf{z}) = E(\mathbf{z}, \mathrm{H}\{f\}) + \lambda\|\mathrm{L}f\|_{\mathcal{M}}$, defined over the Banach space $\mathcal{X}_1$.

Existence of Solutions. We first show that $\mathcal{V} = \arg\min_{f\in\mathcal{X}_1} \mathrm{J}_1(f|\mathbf{z})$ is nonempty, convex, and weak*-compact.

We rely on the generalized Weierstrass theorem presented in [49]: Any proper, lower semi-continuous (lsc) functional over a compact topological vector space reaches its minimum, from which we deduce the following result. We recall that the dual space $\mathcal{B}'$ of a Banach space $\mathcal{B}$ can be endowed with the weak*-topology, and that one can define a norm $\|f\|_{\mathcal{B}'} = \sup_{\|x\|_{\mathcal{B}}}\langle f, x\rangle$ for which $\mathcal{B}'$ is a Banach space.

Proposition 8. *Let $\mathcal{B}$ be a Banach space. Then, a functional $\mathrm{J} : \mathcal{B}' \to \mathbb{R}^+ \cup \{\infty\}$ which is proper, convex, coercive, and weak*-lsc is lower bounded and reaches its infimum. Moreover, the set $\mathcal{V} = \arg\min \mathrm{J}$ is convex and weak*-compact.*

Proof. Let $\alpha > \inf \mathrm{J}$. The coercivity implies that there exists $r > 0$ such that $\mathrm{J}(f) \geq \alpha$ as soon as $\|f\|_{\mathcal{B}'} > r$. The infimum of J can only be reached on $B_r = \{f \in \mathcal{B}', \|f\|_{\mathcal{B}'} \leq r\}$, hence we restrict our analysis to it. The Banach-Alaoglu theorem implies that B_r is weak*-compact. As a consequence, the functional J is proper and lsc on the compact space B_r endowed with the weak*-topology. According to the generalized Weierstrass theorem [49, Theorem 7.3.1], J reaches its infimum on B_r, hence on $\mathcal{X}'$.

Let $\mathcal{V} = \arg\min \mathrm{J}$ and $\alpha_0 = \min \mathrm{J}$. The convexity of J directly implies the one of $\mathcal{V}$. The set $\mathcal{V}$ is included in the ball B_{α_0} which is weak*-compact. Therefore, it suffices to show that $\mathcal{V}$ is weak*-closed to deduce that it is weak*-compact. Moreover, the weak*-lower semi-continuity is equivalent to the weak*-closedness of the level sets $\{f \in \mathcal{B}' : \mathrm{J}(f) \leq \alpha\}$ are weak*-closed. Applying this to $\alpha = \alpha_0$, we deduce that $\mathcal{V} = \{f \in \mathcal{B}' : \mathrm{J}(f) \leq \alpha_0\}$ is weak*-closed, as expected. □

We apply Proposition 8 to $\mathcal{B}' = \mathcal{X}_1$, which is the dual of the Banach space $\mathcal{B} = C_{\mathrm{L}}(\mathbb{R})$ introduced in [11] and recapped in the supplementary material. One has to show that the functional $\mathrm{J} = \mathrm{J}_1(\cdot|\mathbf{z})$ is coercive and weak*-lsc over $\mathcal{X}_1$. The coercivity is deduced exactly in the same way as for Theorem 3. The weak*-lower semi-continuity is deduced as follows. First, H is weak*-continuous by assumption and $E(\mathbf{z}, \cdot)$ is lsc; hence, the composition $f \mapsto E(\mathbf{z}, \mathrm{H}\{f\})$ is weak*-lsc. Similarly, the norm $\|\cdot\|_{\mathcal{M}}$ is weak*-lsc on $\mathcal{M}$ and $\mathrm{L} : \mathcal{X}_1 \to \mathcal{M}$ is continuous, hence $f \mapsto \|\mathrm{L}f\|_{\mathcal{M}}$ is weak*-continuous, and therefore weak*-lsc over $\mathcal{X}_1$. Finally, $\mathrm{J}_1(\cdot|\mathbf{z})$ is weak*-lsc over $\mathcal{X}_1$ as it is a sum of two weak*-lsc functionals.

Form of the Extreme Points. Let f_e be an extreme point of the set $\mathcal{V}_1$ and $\mathbf{z}_e = \mathrm{H}f_e$. Then f_e is also a member of the solution set

$$\mathcal{V}_e = \arg \min_{f=\mathrm{H}^{-1}\{\mathbf{z}_e\}} \|\mathrm{L}f\|_{\mathcal{M}}. \tag{70}$$

Since $\mathbf{z}_e$ is convex and compact, and the set $\mathrm{H}^{-1}\{\mathbf{z}_e\}$ is nonempty, we can apply Theorem 2 of [11] to deduce that $\mathcal{V}_e$ is convex and weak*-compact, together with the general form (19) of the extreme-points of $\mathcal{V}_e$.

Since $\mathcal{V}_e \subseteq \mathcal{V}_1$, and $f_e \in \mathcal{V}_e$ it can be easily shown that f_e is also an extreme point of $\mathcal{V}_e$. This proves that the extreme points of $\mathcal{V}_1$ admit the form (19).

Measurement of the solution set. We now show that in the case of Assumption 2' the measurement of the solution set is unique. We first prove this for the case of Assumption 2'.i). Let J_1^* be the minimum value attained by the solutions. Let f_A^* and f_B^* be two solutions. Let e_A, e_B be their corresponding E functional value and let r_A, r_B be their corresponding regularization functional value. Since the cost function is convex, any convex combination $f_{AB} = \beta f_A^* + (1-\beta) f_B^*$ is also a solution for $\beta \in [0, 1]$ with functional value J_1^*. Let us assume that $\mathrm{H}\{f_A^*\} \neq \mathrm{H}\{f_B^*\}$. Since $E(\mathbf{z}, \cdot)$ is strictly convex and $\mathrm{R}_1(f) = \|\mathrm{L}f\|_{\mathcal{M}}$ is convex, we get that

$$\begin{aligned} J_1^* &= E(\mathbf{z}, \mathrm{H}\{\beta f_A^* + (1-\beta) f_B^*\}) + \lambda \mathrm{R}_1(\beta f_A^* + (1-\beta) f_B^*) \\ &< \underbrace{\beta e_A + (1-\beta) e_B + \lambda\beta r_A + \lambda(1-\beta) r_B}_{J_1^*}. \end{aligned}$$

This is a contradiction. Therefore, $\mathrm{H}\{f_A^*\} = \mathrm{H}\{f_B^*\} = \mathrm{H}\{f_{AB}\}$.

In the case of Assumption 2'.ii), $E(\mathbf{z}, \cdot)$ is an indicator function. It is therefore obvious that all the solutions have the same measurement $\mathbf{z}$.

Appendix C
Proof of Theorem 6

We first state two propositions that are needed for the proof. Their proofs are given in the supplementary material.

Proposition 9 (Adapted from [10, Theorem 5])**.** *Let $\mathbf{z} \in \mathbb{R}^M$ and $\mathbf{H} \in \mathbb{R}^{M\times N}$, where $M < N$. Then, the solution set α_λ of*

$$\mathbf{a}^* = \arg \min_{\mathbf{a}\in\mathbb{R}^N} \left(\|\mathbf{z} - \mathbf{H}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1\right) \tag{71}$$

is a compact convex set and $\|\mathbf{a}\|_0 \leq M$, $\forall \mathbf{a} \in \alpha_{E,\lambda}$, where $\alpha_{E,\lambda}$ is the set of the extreme points of α_λ.

Proposition 10. *Let the convex compact set α_λ be the solution set of Problem (48) and let $\alpha_{E,\lambda}$ be the set of its extreme points. Let the operator $\mathrm{T} : \alpha_\lambda \to \mathbb{R}^N$ be such that $\mathrm{T}\mathbf{a} = \mathbf{u}$ with $u_m = |a_m|, m \in [1, \ldots, N]$. Then, the operator is linear and invertible over the domain α_λ and the range $\mathrm{T}\alpha_\lambda$ is convex compact such that the image of any extreme point $\mathbf{a}_E \in \alpha_{E,\lambda}$ is also an extreme point of the set $\mathrm{T}\alpha_\lambda$.*

The linear program corresponding to (50) is

$$\begin{aligned} (\mathbf{a}^*, \mathbf{u}^*) = \min_{\mathbf{a},\mathbf{u}} \sum_{n=1}^{N} u_n, \text{ subject to } & \mathbf{u} + \mathbf{a} \geq \mathbf{0}, \\ & \mathbf{u} - \mathbf{a} \geq \mathbf{0}, \\ & \mathbf{P}\mathbf{a} = \mathbf{z}. \end{aligned} \tag{72}$$

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

By putting $\mathbf{u}+\mathbf{a}=\mathbf{s}_1$ and $(\mathbf{u}-\mathbf{a})=\mathbf{s}_2$, the standard form of this linear program is

$$(\mathbf{s}_1^*,\mathbf{s}_2^*)=\min_{\mathbf{s}_1,\mathbf{s}_2}\left(\sum_{n=1}^{N}s_{1n}+s_{2n}\right),\ \text{s.t. } \mathbf{s}_1\geq\mathbf{0},\ \mathbf{s}_2\geq\mathbf{0},\ \mathbf{P}\mathbf{s}_1-\mathbf{P}\mathbf{s}_2\leq\mathbf{z}\ -\mathbf{P}\mathbf{s}_1+\mathbf{P}\mathbf{s}_2\leq-\mathbf{z}. \quad (73)$$

Any solution $\mathbf{a}^*$ of (72) is equal to $(\mathbf{s}_1^*-\mathbf{s}_2^*)$ for some solution pair (73). We denote the concatenation of any two independent points $\mathbf{s}_1^r,\mathbf{s}_2^r\in\mathbb{R}^N$ by the variable $\mathbf{s}^r=(\mathbf{s}_1^r,\mathbf{s}_2^r)\in\mathbb{R}^{2N}$. Then, the concatenation of the feasible pairs $\mathbf{s}^f=\left(\mathbf{s}_1^f,\mathbf{s}_2^f\right)$ that satisfies the constraints of the linear program (73) forms a polytope in $\mathbb{R}^{2N}$. Given that (73) is solvable, it is known that at least one of the extreme points of this polytope is also a solution. The simplex algorithm is devised such that its solution $\mathbf{s}_{\text{SLP}}^*=\left(\mathbf{s}_{1,\text{SLP}}^*,\mathbf{s}_{2,\text{SLP}}^*\right)$ is an extreme point of this polytope [34]. Our remaining task is to prove that $\mathbf{a}_{\text{SLP}}^*=\left(\mathbf{s}_{1,\text{SLP}}^*-\mathbf{s}_{2,\text{SLP}}^*\right)$ is an extreme point of the set α_λ, the solution set of the problem (48).

Proposition 9 claims that the solution set α_λ of the LASSO problem is a convex set with extreme points $\alpha_{E,\lambda}\in\mathbb{R}^N$. As α_λ is convex and compact, the concatenated set $\zeta=\{\boldsymbol{w}\in\mathbb{R}^{2N}:\boldsymbol{w}=(\mathbf{a}^*,\mathbf{u}^*),\mathbf{a}^*\in\alpha_\lambda\}$ is convex and compact by Proposition 10. The transformation $(\mathbf{a}^*,\mathbf{u}^*)=(\mathbf{s}_1^*-\mathbf{s}_2^*,\mathbf{s}_1^*+\mathbf{s}_2^*)$ is linear and invertible. This means that the solution set of (73) is convex and compact, too. The simplex solution corresponds to one of the extreme points of this convex compact set.

Since the map $(\mathbf{a}^*,\mathbf{u}^*)=(\mathbf{s}_1^*-\mathbf{s}_2^*,\mathbf{s}_1^*+\mathbf{s}_2^*)$ is linear and invertible, it also implies that an extreme point of the solution set of (73) corresponds to an extreme point of ζ. Proposition 10 then claims that this extreme point of ζ corresponds to an extreme point $\mathbf{a}_{\text{SLP}}\in\alpha_{\lambda,E}$.

REFERENCES

[1] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics*, vol. 4, pp. 1035–1038, 1963.

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. CRC press, 1998.

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

[4] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

[5] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[6] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[7] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.

[8] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 265–288, 1996.

[9] B. Efron, T. Hastie, and R. Tibshirani, "Discussion: The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2358–2364, Dec. 2007.

[10] M. Unser, J. Fageot, and H. Gupta, "Representer theorems for sparsity-promoting ℓ_1-regularization," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5167–5180, Sep. 2016.

[11] 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, Dec. 2017.

[12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[13] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[14] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," *Lecture Notes in Computer Science*, vol. 2111, pp. 416–426, 2001.

[15] G. Wahba, *Spline Models for Observational Data*. SIAM, 1990, vol. 59.

[16] ——, "Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV," *Advances in Kernel Methods-Support Vector Learning*, vol. 6, pp. 69–87, 1999.

[17] A. Y. Bezhaev and V. A. Vasilenko, *Variational theory of splines*. Springer, 2001.

[18] H. Wendland, *Scattered Data Approximation*. Cambridge University press, 2004, vol. 17.

[19] J. Kybic, T. Blu, and M. Unser, "Generalized sampling: A variational approach—Part I: Theory," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1965–1976, Aug. 2002.

[20] ——, "Generalized sampling: A variational approach—Part II: Applications," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1977–1985, Aug. 2002.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physics D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[22] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *International Journal of Computer Vision*, vol. 70, no. 3, pp. 241–255, Dec. 2006.

[23] S. Fisher and J. Jerome, "Spline solutions to L_1 extremal problems in one and several variables," *Journal of Approximation Theory*, vol. 13, no. 1, pp. 73–83, Jan. 1975.

[24] K. Bredies and H. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, Jan. 2013.

[25] E. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, Dec. 2013.

[26] Q. Denoyelle, V. Duval, and G. Peyré, "Support recovery for sparse super-resolution of positive measures," *Journal of Fourier Analysis and Applications*, vol. 23, no. 5, pp. 1153–1194, Oct. 2017.

[27] A. Chambolle, V. Duval, G. Peyré, and C. Poon, "Geometric properties of solutions to the total variation denoising problem," *Inverse Problems*, vol. 33, no. 1, p. 015002, Dec. 2016.

[28] A. Flinth and P. Weiss, "Exact solutions of infinite dimensional total-variation regularized problems," arXiv:1708.02157 [math.OC], 2017.

[29] M. Unser, "A representer theorem for deep neural networks," arXiv:1802.09210 [stat.ML], 2018.

[30] A. Braides, *Gamma-convergence for Beginners*. Clarendon Press, 2002, vol. 22.

[31] V. Duval and G. Peyré, "Sparse regularization on thin grids I: the Lasso," *Inverse Problems*, vol. 33, no. 5, p. 055008, 2017.

[32] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries-just discretize," in *Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 1043–1047.

[33] G. B. Dantzig, A. Orden, and P. Wolfe, "The generalized simplex method for minimizing a linear form under linear inequality restraints," *Pacific Journal of Mathematics*, vol. 5, no. 2, pp. 183–195, Oct. 1955.

[34] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Addison-Wesley Reading, MA, 1973, vol. 28.

[35] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[36] R. J. Tibshirani, "The LASSO problem and uniqueness," *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.

[37] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, Apr. 2008.

[38] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.

[39] A. Chambolle and C. Dossal, "On the convergence of the iterates of FISTA," *Journal of Optimization Theory and Applications*, vol. 166, no. 3, p. 25, 2015.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2860549, IEEE Transactions on Signal Processing

[40] E. Mammen and S. van de Geer, "Locally adaptive regression splines," *Annals of Statistics*, vol. 25, no. 1, pp. 387–413, 1997.
[41] 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.
[42] 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.
[43] Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre, "Exact solutions to super resolution on semi-algebraic domains in higher dimensions," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 621–630, 2017.
[44] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, Jun. 2005.
[45] M. Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 989–1006, Mar. 2011.
[46] ——, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2014.
[47] J. Fageot, V. Uhlmann, and M. Unser, "Gaussian and sparse processes are limits of generalized Poisson processes," arXiv:1702.05003 [math.PR], 2017.
[48] W. Rudin, *Real and Complex Analysis*. Tata McGraw-Hill Education, 1987.
[49] A. J. Kurdila and M. Zabarankin, *Convex functional analysis*. Springer Science & Business Media, 2006.
[50] R. Michael and B. Simon, *Methods of modern mathematical physics I: Functional analysis*. Academic Press, 1980.

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.