# Characterizing the Solution Set of Discrete-Time Inverse Problems as Sparse Superpositions of Decaying Shifted Heaviside Functions

1st Richard Oliveira
*Electrical and Computer Engineering Department*
*University of California, San Diego*
La Jolla, USA
raolivei@ucsd.edu

***Abstract*—In this paper we investigate the solution set of inverse problems that deal with the reconstruction of an infinite length, causal, finite energy discrete-time signal, as sparse superpositions of shifted and decaying Heaviside functions scaled by appropriate weights. Moreover, we showcase how the search for a sparse signal minimizer in time-domain eventually leads to regularizing the solution with a quadratic form as a function of the minimizer's weights, whose associated matrix is characterized by the signal's positive integer shifts, and scaling and decaying factors. We numerically validate our theoretical findings by showing that our approach yields a better reconstruction in terms of structure preservation and peak signal-to-noise ratio compared to the Least Absolute Shrinkage and Selection Operator (LASSO) method, and suggest a new direction of research in improving edge preservation during image reconstruction.**

***Index Terms*—inverse problem, Hardy space, Banach space, Gram matrix, quadratic form, regularization**

## I. Introduction

We start by considering the continuous-domain inverse problem, in which we are interested in recovering an infinite dimensional object $f : \mathbb{R}^d \to \mathbb{R}$, given a finite dimensional measurement

$$\mathbf{y} = \mathbf{H}\{f\} + \boldsymbol{\epsilon} \tag{1}$$

of dimension $\mathbb{R}^M$, where $\mathbf{H}$ is a measurement operator that acts on $f$, and $\boldsymbol{\epsilon}$ is a zero-mean Gaussian random vector.
This problem is clearly ill-posed, since we wish to reconstruct a continuous-domain object given a vector of measurements. Applications concerning signal recovery, are interested in a $f$ which is sparse in some domain. Practitioners in the field of computational imaging, and compressed sensing for example, are interested in sparse signal representations in frequency-domain [5], [14]-[17]. Because of this, it is common practice to assume a priori, that $f$ can be written as

$$f = \sum_{n\in\mathbb{Z}} \gamma_n \phi_n \tag{2}$$

where $\phi_j : \mathbb{R}^d \to \mathbb{R}$ is an atom belonging to a orthonormal dictionary, and $\{\gamma_n\}_{n\in\mathbb{Z}}$ is the set of expansion coefficients. The literature solves the problem of recovering $f$ given a finite number of measurements, by solving a regularized optimization problem, in which the norm associated to the regularization term is imposing sparsity on $f$ in frequency-domain [15]-[17], [20]. Conversely, in this paper, we consider signal reconstruction with sparse-promoting norms in time-domain, which differs from existing work, where the assumption of a sparse frequency representation of $f$ is not necessary, if the underlying signal to be reconstructed is indeed sparse in time-domain. Naturally, by assuming that $f$ can be written as (2), leads to regularized least-squares

$$\hat{\boldsymbol{\gamma}} \in \arg\min_{\boldsymbol{\gamma}\in\ell^p(\mathbb{Z})} \left\| \mathbf{y} - \mathbf{H}\left\{ \sum_{k\in\mathbb{Z}} \gamma_k \phi_k \right\} \right\|_2^2 + \tau \left\|\boldsymbol{\gamma}\right\|_p^p, \tag{3}$$

where $\tau > 0$ and $p \in [0, \infty)$ are adjustable hyperparameters. Since we are interested in looking for a sparse solution, it is common to set $p = 1$. The least-squares formulation in (3) is also referred to as the *synthesis* formulation. Alternatively, instead of assuming a prior structure on $f$, the search for a signal $f$ in a space with sparse-promoting norm, is referred to as the *analysis* formulation, or sometimes called the *variational* formulation, and takes the following form

$$\hat{f} \in \arg\min_{f\in\mathcal{X}^*} \left\|\mathbf{y} - \mathbf{H}\{f\}\right\|_2^2 + \tau \left\|f\right\|_{\mathcal{X}^*}^p, \tag{4}$$

where the search space $\mathcal{X}^*$ is the *native* space, to be viewed as a dual space. Recently, there have been works that studied the problem in (4), and explored Total Variation (TV) semi-norm regularization approaches, in which, the frequency representation of the candidate signal $f$ lives in a spectral Barron space of order $s$, endowed with a norm that measures the sparsity of the first $s$ fractional derivatives of the Fourier transform of $f$ [3]. Although continuous-domain methods in [3] are general, the TV-norm of the minimizer $\hat{f}$ almost unsurprisingly yields the $\ell^1$-norm of the weights associated to it's sparse superposition representation, thus in some ways agreeing with the well-established least-squares regularization term in (3).
From now on, we will focus on the *variational* formulation in

(4), and the signal to be reconstructed is discrete, infinite in length, causal and has finite energy. More specifically, we are looking for signals with membership in

$$\ell^2(\mathbb{N}) := \left\{ \mathbf{x} = \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ . \\ . \end{pmatrix} : \sum_{n=0}^{\infty} |x(n)|^2 < \infty \right\},$$

where $n \in \mathbb{N}$ (for our context, it is the set of positive integers that include 0). By setting $\mathcal{X}^* = \ell^2(\mathbb{N})$, the problem in (4) becomes

$$\hat{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \ell^2(\mathbb{N})} \|\mathbf{y} - \mathbf{H}\{\mathbf{x}\}\|_2^2 + \tau \|\mathbf{x}\|_{\ell^2(\mathbb{N})}^p . \tag{5}$$

Let the one-sided $\mathcal{Z}$-transform of a causal signal $\mathbf{x} \in \ell^2(\mathbb{N})$ be defined as

$$X(z) = \mathcal{Z}\{\mathbf{x}\} = \sum_{n=0}^{\infty} x(n) z^n, \tag{6}$$

where $z = re^{j\theta} \in \mathbb{C}$. Note that we omit the negative exponent in $z$ to maintain consistency with the the definitions that follow. Furthermore, the sum in (6) converges absolutely provided that $|z| < 1$. Additionally, we define the following set [12]:

$$\mathcal{K} = \left\{ X = \mathcal{Z}\{\mathbf{x}\} : \mathbf{x} \in \ell^2(\mathbb{N}) \right\}.$$

**Definition I.1.** *With $\mathcal{K}$, define for $X \in \mathcal{K}$ the Hardy norm:*

$$||X||_{\mathcal{H}^2} = \sqrt{\sup_{r \in (0,1)} \frac{1}{2\pi} \int_0^{2\pi} |X(re^{j\theta})|^2 \, d\theta}$$

*with corresponding Hardy space*

$$\mathcal{H}^2 = \{ X : X \in \mathcal{K}, \ ||X||_{\mathcal{H}^2} < \infty \}.$$

Historically, Hardy spaces have played, and play an important role in optimal control [21]. However, recent works have shown the subtle effectiveness of Hardy spaces in the context of recurrent neural network (RNN) learning [12], which differs from control paradigms. In this paper, we offer a fresh perspective, and take advantage of $\mathcal{H}^2$ in the context of signal estimation, and build upon what has been developed in the realm of regularization over Banach spaces [3], [4], and consider a non-trivial set of extreme points as solutions to (5) that retain causality and decaying structure. More specifically, we show that the convex hull of the extreme points of the unit ball (disk) in a Hardy Space $\mathcal{H}^2$, under the $\mathcal{Z}$-transform inversion operator, are solutions to (5) and have the following form

$$\hat{x}(n) = \sum_{k=1}^{K} \omega_k \alpha_k^{n-m_k} \sqrt{1-\alpha_k^2} \ u(n-m_k), \quad m_k \in \mathbb{N}, \tag{7}$$

where $\omega_k \in \mathbb{R}$, $\alpha_k \in [0,1]$, $m_k$ is a positive integer shift and $u(\cdot - m_k)$ is the Heaviside function for $k = 1, ..., K$, and $K \leq M$. Furthermore, we show that the $\ell^2(\mathbb{N})$-norm of $\hat{\mathbf{x}}$ is equivalent to measuring the energy of the weights $\omega_k$ in (7), i.e., $\hat{x}(n)$ is sparse in the sense that $\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})} = \|\boldsymbol{\omega}\|_{\boldsymbol{\Gamma}} = \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}$, where $\boldsymbol{\Gamma} = \mathbf{UAU}$ is a *Positive Semi-Definite* matrix, $\mathbf{A}$ is a *Gram* matrix, $\mathbf{U}$ is a diagonal matrix, and the matrix entries of $\mathbf{A}$ and $\mathbf{U}$ have a neat and closed-form expression, completely defined by $\alpha_k$ and $m_k$ for $k = 1, ..., K$. Moreover, we will numerically demonstrate that by employing the regularization scheme in (5) our results yield better image reconstruction in terms of structure preservation and peak signal-to-noise ratio (PSNR) with sparse digital images when compared to LASSO [20].

## II. Main Results

Extreme points in Banach spaces, have been studied extensively [6], [18], and the mapping of extreme points to extreme points between such spaces, under a well-suited operator, for signal reconstruction purposes have been recently tackled in a continuous-domain setting [3]. However, the current literature has not yet explored the role that Hardy spaces can have in signal reconstruction problems, and therefore, in this work, we consider the relationship between the extreme points of the unit ball in $\mathcal{H}^2$ and the extreme points in $\ell^2(\mathbb{N})$, under the $\mathcal{Z}$-transform mapping. We begin by noting that both $\mathcal{H}^2$ and $\ell^2(\mathbb{N})$ are Hilbert spaces and also well-established Banach spaces in the literature [1], [2]. Furthermore, $\ell^2(\mathbb{N})$ is a *reflexive* Hilbert space and is naturally its own pre-dual, i.e., $(\ell^2(\mathbb{N}))^* \simeq \ell^2(\mathbb{N})$ [6], [7].
Consider the mapping and inverse mapping between $\ell^2(\mathbb{N})$ and $\mathcal{H}^2$ [8], by letting the $\mathcal{Z}$-transform in (6) have the following forward and inverse operator definitions respectively [19]:

$$\mathcal{Z} : \ell^2(\mathbb{N}) \ni \mathbf{x} \mapsto \sum_{n=0}^{\infty} x(n)(\cdot)^n \in \mathcal{H}^2 \tag{8}$$

$$\mathcal{Z}^{-1} : \mathcal{H}^2 \ni X(z) \mapsto \frac{1}{2\pi j} \oint_C X(z) \, z^{-(\cdot)-1} \, dz \in \ell^2(\mathbb{N}), \tag{9}$$

where (8) maps a signal $\mathbf{x} \in \ell^2(\mathbb{N})$ to a convergent power series in $\mathcal{H}^2$, and (9) performs the inverse operation, where $C$ is a closed contour in the complex plane. Under the forward and inverse mappings defined in both (8) and (9), we proceed by establishing the *isometry* correspondence between $\ell^2(\mathbb{N})$ and $\mathcal{H}^2$ [11]:

$$\|\mathbf{x}\|_{\ell^2(\mathbb{N})} = \|\mathcal{Z}\{\mathbf{x}\}\|_{\mathcal{H}^2} . \tag{10}$$

Now, consider the extreme points on the unit ball in $\mathcal{H}^2$:

$$\mathcal{H}^2_{\text{ext}} = \{ X : \|X\|_{\mathcal{H}^2} \leq 1 \} \subset \mathcal{H}^2.$$

It is well-known that the set of points that live on the boundary of $\mathcal{H}^2_{\text{ext}}$ have the following form [11], [18]:

$$X^{\text{ext}}(z) = \sqrt{1-\alpha^2} \frac{z^m}{1-\alpha z}, \tag{11}$$

where $\alpha < 1$ and $m \in \mathbb{N}$ is some positive integer. Moreover, it is easy to check that $X^{\text{ext}}(z)$ has unit norm for $0 < \alpha < 1$, and therefore, this calculation is omitted. Consider the extreme points on the unit ball in (11) and by noting that the $\mathcal{Z}$-transform operator in (8) is an *isometric isomorphism* between $\ell^2(\mathbb{N})$ and $\mathcal{H}^2$, we have that unit norm convergent power series are mapped to unit norm signals in $\ell^2(\mathbb{N})$ under the inversion operator $\mathcal{Z}^{-1}$ in (9). Thanks to this, we arrive at defining the extreme points in $\ell^2(\mathbb{N})$ as

$$\mathcal{Z}^{-1}\left\{X^{\text{ext}}(z)\right\} = \alpha^{n-m}\sqrt{1-\alpha^2}\; u(n-m), \tag{12}$$

which are causal and have a decaying property. We are now ready to state the main results of this paper.

**Theorem II.1.** *Let $\mathcal{X}$ denote the pre-dual native space, with $\mathcal{X}^*$ being its dual space, such that $\mathcal{X} = \mathcal{X}^* = \ell^2(\mathbb{N})$. Consider a measurement process $\mathbf{H}\{\mathbf{x}\}$ defined as $(\langle \mathbf{h}_1, \mathbf{x}\rangle, ..., \langle \mathbf{h}_M, \mathbf{x}\rangle)^T \in \mathbb{R}^M$, where each measurement operator $\mathbf{h}_i \in \mathcal{X}$ for $i = 1, ..., M$ is linear and weak* continuous. Then, for any fixed measurement vector $\mathbf{y} \in \mathbb{R}^M$, the set of solutions*

$$\mathcal{U} := \arg\min_{\mathbf{x}\in\ell^2(\mathbb{N})} \|\mathbf{y} - \mathbf{H}\{\mathbf{x}\}\|_2^2 + \tau\,\|\mathbf{x}\|_{\ell^2(\mathbb{N})} \subset \ell^2(\mathbb{N}) \tag{13}$$

*where $\tau > 0$ is a tunable hyperparameter, is nonempty, convex, and weak* compact. Moreover, the extreme points of $\mathcal{U}$ can be expressed as signals of the form:*

$$\hat{x}(n) = \sum_{k=1}^{K} \omega_k \alpha_k^{n-m_k}\sqrt{1-\alpha_k^2}\; u(n-m_k), \quad m_k \in \mathbb{N} \tag{14}$$

*where $\omega_k \in \mathbb{R}$, $\alpha_k \in [0,1]$, $m_k$ is a positive integer shift and $u(\cdot - m_k)$ is the Heaviside function for $k = 1, ..., K$, and $K \leq M$. The convex hull of these extreme points constitutes the entire solution set for $n \in \mathbb{N}$.*

*Proof.* The proof follows by combining Theorem 2 with Theorem 3 from [4]. The combination of these two theorems characterize the solution set of variational problems as superpositions of extremal points of the unit ball associated to the regularization term [3].

□

Next, we show that by measuring the norm of the minimizer $\hat{x}(n)$ we arrive at measuring the energy associated to the weights $\omega_k$ for $k = 1, ..., K$ in (14).

**Proposition II.2.** *Let $\hat{x}(n)$ be the minimizer of* (13) *for $n \in \mathbb{N}$, and let $\boldsymbol{\omega} = (\omega_1, ..., \omega_K)^T$. Then*

$$\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})} = \|\boldsymbol{\omega}\|_{\boldsymbol{\Gamma}} = \boldsymbol{\omega}^T\mathbf{U}\mathbf{A}\mathbf{U}\boldsymbol{\omega}, \tag{15}$$

*where $\mathbf{A} \in \mathbb{R}^{K\times K}$ is a Gram matrix with entries*

$$\mathbf{A}_{ij} = \begin{cases} \frac{(\alpha_i)^{2m_i}}{1-\alpha_i^2} & \text{if } \; i = j \\ \frac{(\alpha_i\alpha_j)^{\max(m_i,m_j)}}{1-\alpha_i\alpha_j} & \text{if } \; i \neq j\ , \end{cases}$$

*and $\mathbf{U} \in \mathbb{R}^{K\times K}$ is a diagonal matrix with entries*

$$\mathbf{U}_{ij} = \begin{cases} \alpha_i^{-m_i}\sqrt{1-\alpha_i^2} & \text{if } \; i = j \\ 0 & \text{if } \; i \neq j\ , \end{cases}$$

*for $i, j = 1, ..., K$. Moreover, the matrix $\mathbf{U}\mathbf{A}\mathbf{U}$ is Positive Semi-Definite.*

*Proof.* Let $\beta_k = \omega_k\sqrt{1-\alpha_k^2}$ for $\omega_k$ and $\alpha_k$ in (14) and denote $\overline{(\cdot)}$ as the complex conjugation operation.

$$\begin{aligned}
\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})} &= \sum_{n=0}^{\infty}\left|\sum_{k=1}^{K}\beta_k\alpha_k^{n-m_k}\; u(n-m_k)\right|^2 \\
&= \sum_{n=0}^{\infty}\overline{\sum_{k=1}^{K}\beta_k\alpha_k^{n-m_k}\; u(n-m_k)}\sum_{l=1}^{K}\beta_l\alpha_l^{n-m_l}\; u(n-m_l) \\
&= \sum_{n=0}^{\infty}\sum_{k=1}^{K}\sum_{l=1}^{K}\overline{\beta_k}\,\overline{\alpha_k}^{n-m_k}\beta_l\alpha_l^{n-m_l}u(n-m_k)u(n-m_l).
\end{aligned}$$

Note that

$$u(n-m_k)u(n-m_l) = \begin{cases} 1 & \text{if } n \geq \max(m_k, m_l) \\ 0 & \text{otherwise}\ , \end{cases}$$

and therefore

$$\begin{aligned}
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\overline{\beta_k}\,\overline{\alpha_k}^{-m_k}\beta_l\alpha_l^{-m_l}\sum_{n=\max(m_k,m_l)}^{\infty}(\overline{\alpha_k}\alpha_l)^n \\
&\overset{(*)}{=} \sum_{k=1}^{K}\sum_{l=1}^{K}\beta_k\alpha_k^{-m_k}\beta_l\alpha_l^{-m_l}\frac{(\alpha_k\alpha_l)^{\max(m_k,m_l)}}{1-\alpha_k\alpha_l}.
\end{aligned}$$

In $(*)$, without loss of generality, we omit complex conjugation, since it is clear that $\alpha_k$ and $\omega_k$ in (14) are real, therefore $\overline{\beta}_k = \beta_k$, $\overline{\alpha}_k = \alpha_k$ and $\overline{\alpha_k}^{-m_k} = \alpha_k^{-m_k}$. Moreover, the geometric series starting at $n = \max(m_k, m_l)$ converges since $0 < \alpha_k\alpha_l < 1$. Now, we define the following vectors

$$\begin{aligned}
\boldsymbol{\alpha}_k &= \left(\frac{(\alpha_k\alpha_1)^{\max(m_k,m_1)}}{1-\alpha_k\alpha_1}, ..., \frac{(\alpha_k\alpha_K)^{\max(m_k,m_K)}}{1-\alpha_k\alpha_K}\right)^T, \\
\boldsymbol{\beta} &= \left(\beta_1\alpha_1^{-m_1}, ..., \beta_K\alpha_K^{-m_K}\right)^T.
\end{aligned} \tag{16}$$

Continuing the computation of $\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})}$ we have that

$$\begin{aligned}
\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})} &= \sum_{k=1}^{K}\beta_k\alpha_k^{-m_k}\boldsymbol{\beta}^T\boldsymbol{\alpha}_k \\
&= \boldsymbol{\beta}^T\underbrace{\begin{pmatrix} | & & | \\ \boldsymbol{\alpha}_1 & ... & \boldsymbol{\alpha}_K \\ | & & | \end{pmatrix}}_{\mathbf{A}}\boldsymbol{\beta}.
\end{aligned} \tag{17}$$

Given the vector definitions in (16), the matrix $\mathbf{A}$ in (17) has entries:

$$\mathbf{A}_{ij} = \begin{cases} \frac{(\alpha_i)^{2m_i}}{1-\alpha_i^2} & \text{if } \; i = j \\ \frac{(\alpha_i\alpha_j)^{\max(m_i,m_j)}}{1-\alpha_i\alpha_j} & \text{if } \; i \neq j\ , \end{cases}$$

for $i, j = 1, ..., K$. Clearly, since the matrix entries of $\mathbf{A}$ are bona-fide positive-definite kernels, we conclude that $\mathbf{A}$ is a *Gram* matrix and therefore *Positive Semi-Definite* [24]. Furthermore, the vector $\boldsymbol{\beta}$ can be rewritten as

$$\boldsymbol{\beta} = \mathbf{U}\boldsymbol{\omega}, \tag{18}$$

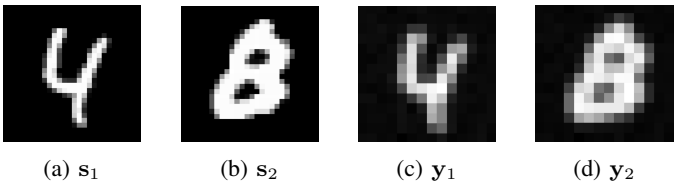
where the matrix $\mathbf{U}$ is diagonal with entries $\alpha_i^{-m_i}\sqrt{1-\alpha_i^2}$ for $i = 1, ..., K$ and $\boldsymbol{\omega} = (\omega_1, ..., \omega_K)^T$. By substituting (18) into (17) we obtain

$$\|\hat{\mathbf{x}}\|_{\ell^2(\mathbb{N})} = (\mathbf{U}\boldsymbol{\omega})^T\mathbf{A}(\mathbf{U}\boldsymbol{\omega}) = \boldsymbol{\omega}^T\mathbf{U}\mathbf{A}\mathbf{U}\boldsymbol{\omega},$$

and since $\mathbf{U}$ is diagonal, the matrix $\mathbf{UAU}$ is *Positive Semi-Definite*, which completes the proof. □

## III. Numerical Experiments

In this section we showcase our theoretical findings by performing some numerical experiments on both MNIST [9] and on a data set of real world CT images [10]. We will make use of these data sets since the $N \times N$ digital images are sparse in the spatial domain. Given a sample image $\mathbf{s} \in \mathbb{R}^{N^2}$ we perform the measurement process in (1), where $\mathbf{s}$ undergoes convolution followed by down-sampling under $\mathbf{H}$ and corrupted with $\boldsymbol{\epsilon}$. Once we obtain the measurement $\mathbf{y} \in \mathbb{R}^M$ generated by the sample $\mathbf{s}$, we numerically solve the optimization problem defined in (5), with $p = 1$. We consider three different runs. Let $\mathbf{s}_1$, $\mathbf{s}_2$ be the sample images associated to MNIST, as it is shown in Fig. 1a and Fig. 1b, and let $\mathbf{s}_3$ be the sample CT image as it is shown in Fig. 2a. Furthermore, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be the associated MNIST measurements and let $\mathbf{y}_3$ be the CT image measurement as it is shown in Fig. 1c, Fig. 1d and Fig. 2b, respectively. Other than qualitatively assessing the reconstructions, we use three performance metrics in order to compare our approach (which we will call $\Gamma$-regularization) with LASSO: Structural Similarity Index (SSIM) for structure preservation [23], PSNR and the Edge Preservation Index (EPI) for feature retention [13].



(a) $\mathbf{s}_1$ (b) $\mathbf{s}_2$ (c) $\mathbf{y}_1$ (d) $\mathbf{y}_2$

(e) $\hat{\mathbf{s}}_1^{\Gamma}$ (f) $\hat{\mathbf{s}}_2^{\Gamma}$ (g) $\hat{\mathbf{s}}_1^{\mathrm{L}}$ (h) $\hat{\mathbf{s}}_2^{\mathrm{L}}$

Fig. 1: From top left, we show the MNIST image samples $\mathbf{s}_1$ and $\mathbf{s}_2$, followed by their corrupted versions $\mathbf{y}_1$ and $\mathbf{y}_2$. Starting from bottom left we show the reconstructions via $\Gamma$-regularization, $\hat{\mathbf{s}}_1^{\Gamma}$ and $\hat{\mathbf{s}}_2^{\Gamma}$, followed by the LASSO reconstructions $\hat{\mathbf{s}}_1^{\mathrm{L}}$ and $\hat{\mathbf{s}}_2^{\mathrm{L}}$.

### A. $\Gamma$-regularization

Consider the reconstruction problem defined in (5) and let $\mathbf{s}_i$ be the sparse $N \times N$ image we wish to reconstruct, corresponding to the $i^{\text{th}}$ run, given a measurement $\mathbf{y}_i \in \mathbb{R}^M$ produced by the process in (1), where $M < N^2$. We wish to recover $\mathbf{s}_i$ by solving the following optimization problem

$$\hat{\mathbf{s}}_i^{\Gamma} \in \arg\min_{\mathbf{s}_i} \|\mathbf{y}_i - \mathbf{H}\{\mathbf{s}_i\}\|_2^2 + \tau \|\boldsymbol{\omega}_i\|_{\boldsymbol{\Gamma}_i}, \ i = 1, 2, 3 \tag{19}$$

where $\boldsymbol{\Gamma}_i = \mathbf{U}_i\mathbf{A}_i\mathbf{U}_i$. Fix a $K$ such that $K \leq M$ and construct the matrices $\mathbf{A}_i$ and $\mathbf{U}_i$ by choosing an appropriate set of $\alpha_k^i \in [0, 1]$ and positive integer shifts $m_k^i$, at run $i$, for $k = 1, ..., K$, such that they have the structure defined in proposition II.2. By definition, any choice of $\boldsymbol{\omega}_i$ will satisfy the *Positive Semi-Definite* condition of $\boldsymbol{\Gamma}_i$, and therefore, for our experiments we define the entries $\omega_k^i \in \mathbb{R}$ uniformly at random in $[-1, 1]$. By solving (19), with CVXPY [22], we obtain the estimated MNIST images $\hat{\mathbf{s}}_1^{\Gamma}$ and $\hat{\mathbf{s}}_2^{\Gamma}$ in Fig. 1e and Fig. 1f, respectively, followed by the reconstructed CT scan image in Fig. 2c.

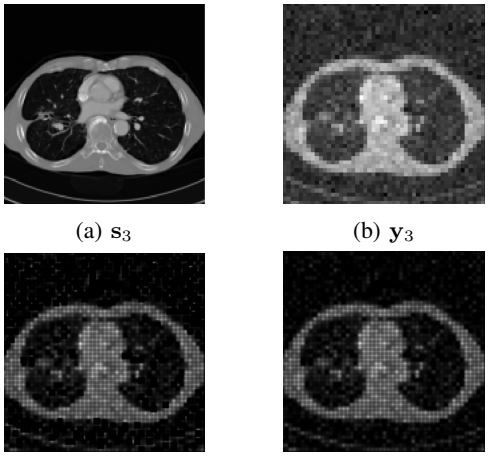
### B. LASSO

In this section we consider the usual reconstruction problem in (5) but instead of solving (19), we aim to solve LASSO:

$$\hat{\mathbf{s}}_i^{\mathrm{L}} \in \arg\min_{\mathbf{s}_i} \|\mathbf{y}_i - \mathbf{H}\{\mathbf{s}_i\}\|_2^2 + \tau \|\mathbf{s}_i\|_1, \ i = 1, 2, 3. \tag{20}$$

The only difference is that the regularization term is the standard $\ell^1$-norm of $\mathbf{s}_i$. Again, by solving (20) we obtain the LASSO reconstructed MNIST images in Fig. 1g and Fig. 1h, and the reconstructed CT image in Fig. 2d.

Qualitatively, by looking at the results in Fig. 1e and Fig. 1f, we can observe that our formulated reconstruction method yields better results when compared to LASSO, with respect to MNIST. Moreover, in Fig. 1g and Fig. 1h, this difference in the results is visible, since the LASSO method yields a quasi blurry reconstruction. The reconstructed CT images in Fig. 2c and Fig. 2d, on the other hand, are difficult to distinguish qualitatively. Therefore, a more quantitative comparison can be observed by looking at TABLE 1. We can see that by employing our $\Gamma$-regularization method, the reconstruction quality in terms of PSNR, is higher with respect to LASSO. Interestingly, if we look at how well our method preserved the overall structure of the sampled image, during reconstruction, we notice values that are closer to 1, in the SSIM column. In contrast, LASSO's estimates clearly have lower SSIM values. This means that our $\Gamma$-regularization approach performs better compared to LASSO approaches in terms of preserving image structure and overall reconstruction quality. On the other hand, both methods perform poorly in the EPI section, although $\Gamma$-regularization still yields higher EPI. This lack of success in preserving edges in the reconstruction can open doors in exploring regularization methods that enhance this feature. In fact, a possible way of tackling this can be found in considering regularization norms that measure sharp transitions in time-domain.



(a) $\mathbf{s}_3$ (b) $\mathbf{y}_3$

(c) $\hat{\mathbf{s}}_3^{\Gamma}$ (d) $\hat{\mathbf{s}}_3^{\mathrm{L}}$

Fig. 2: We show the CT scan image sample $\mathbf{s}_3$, followed by the measurement $\mathbf{y}_3$, $\Gamma$-regularization reconstruction $\hat{\mathbf{s}}_3^{\Gamma}$ and LASSO reconstruction $\hat{\mathbf{s}}_3^{\mathrm{L}}$.

TABLE I: Comparison between $\Gamma$-regularization and LASSO in terms of PSNR, SSIM, and EPI over multiple runs. Best results in bold.

| **Run** | **$\Gamma$-regularization** | | | **LASSO** | | |
|---|---|---|---|---|---|---|
| | ***SSIM*** | ***PSNR (dB)*** | ***EPI*** | ***SSIM*** | ***PSNR (dB)*** | ***EPI*** |
| $\mathbf{s}_1$ | **0.8040** | **17.61** | **0.0134** | 0.2731 | 10.79 | 0.0103 |
| $\mathbf{s}_2$ | **0.8189** | **16.31** | **0.0366** | 0.1035 | 6.69 | 0.0284 |
| $\mathbf{s}_3$ | **0.3996** | **20.05** | **0.0095** | 0.1275 | 11.44 | 0.0088 |

## IV. Conclusion

In conclusion, we have characterized the solution set of inverse problems concerning the recovery of a discrete-time signal, which is infinite in length, causal and has finite energy, as sparse superpositions of decaying shifted Heaviside functions scaled by appropriate weights, by using a Hardy space approach. Moreover, we have showcased that the overall philosophy of looking for a sparse signal minimizer in time-domain, eventually leads to regularizing the solution with a quadratic form, whose weight matrices are neatly characterized by the minimizer's positive integer shifts, and scaling and decaying factors. We also provided numerical experiments, using both MNIST and real world CT scan images, highlighting that our quadratic regularization approach yields better results in terms of SSIM and PSNR, compared to the widely used LASSO. Upon briefly suggesting that both methods need improvement in edge preservation during reconstruction, future work will be directed in improving this capability, while also exploring new forms of regularization in time-domain settings.

## Acknowledgment

The author would like to acknowledge the suggestions from the anonymous reviewers, that helped in improving the quality of this paper.

## References

[1] W. Rudin, *Functional Analysis*, 2nd ed. New York, NY, USA: McGraw-Hill, 1991.

[2] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces with Applications*, 3rd ed. New York, NY, USA: Academic Press, 2005.

[3] R. Parhi and R. D. Nowak, "On Continuous-Domain Inverse Problems With Sparse Superpositions of Decaying Sinusoids as Solutions," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5603–5607, 2022.

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

[5] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller, "DiffuserCam: lensless single-exposure 3D imaging," *Optica* 5, 1–9, 2018.

[6] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler, *Banach Space Theory: The Basis for Linear and Nonlinear Analysis*. Berlin, Germany: Springer, 2011.

[7] J. B. Conway, *A Course in Functional Analysis*, 2nd ed. New York, NY, USA: Springer, 1990.

[8] E. Kreyszig, *Introductory Functional Analysis with Applications*, 1st ed. New York, NY, USA: Wiley, 1978.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[10] K. Clark, et al., "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[11] P. Duren, *Theory of $H^p$ Spaces*. New York, NY, USA: Academic Press, 1970.

[12] C. Hutter, T. Allard, and H. Bölcskei, "Metric Entropy of Causal, Discrete-Time LTI Systems," unpublished.

[13] F. Sattar, L. Floreby, G. Salomonsson, and B. Lovstrom, "Image Enhancement Based on a Nonlinear Multiscale Method," *IEEE Transactions on Image Processing*, vol. 6, no. 6, pp. 888–895, June 1997.

[14] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[15] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, Jan. 2011.

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

[17] E. J. Candes 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.

[18] M. G. Krein and D. P. Milman, "On extreme points of regular convex sets," *Studia Math.*, vol. 9, pp. 133–138, 1940.

[19] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2010, ch. 3.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[21] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, "State-space solutions to standard $H^2$ and $H^\infty$ control problems," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 831–847, Aug. 1989.

[22] M. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2012.