Representer Theorems for Sparsity-Promoting ℓ_1 Regularization

Michael Unser, *Fellow, IEEE*, Julien Fageot, and Harshit Gupta

***Abstract*— We present a theoretical analysis and comparison of the effect of ℓ_1 versus ℓ_2 regularization for the resolution of ill-posed linear inverse and/or compressed sensing problems. Our formulation covers the most general setting where the solution is specified as the minimizer of a convex cost functional. We derive a series of representer theorems that give the generic form of the solution depending on the type of regularization. We start with the analysis of the problem in finite dimensions and then extend our results to the infinite-dimensional spaces $\ell_2(\mathbb{Z})$ and $\ell_1(\mathbb{Z})$. We also consider the use of linear transformations in the form of dictionaries or regularization operators. In particular, we show that the ℓ_2 solution is forced to live in a predefined subspace that is intrinsically smooth and tied to the measurement operator. The ℓ_1 solution, on the other hand, is formed by adaptively selecting a subset of atoms in a dictionary that is specified by the regularization operator. Beside the proof that ℓ_1 solutions are intrinsically sparse, the main outcome of our investigation is that the use of ℓ_1 regularization is much more favorable for injecting prior knowledge: it results in a functional form that is independent of the system matrix, while this is not so in the ℓ_2 scenario.**

***Index Terms*— Sparsity, compressed sensing, linear inverse problems, regularization, ℓ_1-norm minimization, total variation.**

I. Introduction

THE main advantage of using ℓ_2 (or Tikhonov) regularization for the resolution of ill-posed inverse problems is that it yields linear reconstruction algorithms; it is also backed by an elegant and solid mathematical theory [1]–[3]. However, it is not necessarily the method of choice anymore, except for routine reconstruction tasks. During the past decade, the research community has focused its efforts on more sophisticated iterative recovery schemes that exploit a remarkable property of signals called sparsity [4], [5]. The concept is central to the theory of compressed sensing [6], [7] and is driving the development of modern reconstruction algorithms [8]–[10].

There are essentially two strategies for achieving a sparse signal recovery. The first is the synthesis formulation where one attempts to reconstruct a signal from a small subset of atoms within a large dictionary of basis elements [5], [11], [12]. The sparsity constraint is usually enforced by minimizing the ℓ_1-norm of the expansion coefficients. The second strategy is the analysis formulation where the solution is constrained by minimizing a sparsity-promoting functional such as the total-variation semi-norm [8], [13]–[16] or some higher-order extension [17]–[20]. This latter strategy actually goes back much further since it falls within the general framework of regularization theory [21]–[23]. It also has the advantage of being compatible with statistical inference. For instance, one may specify a maximum a posteriori estimator by selecting a regularization functional (or Gibbs energy) that corresponds to the log-likelihood of a given probability model, including Markov random fields [24] or sparse stochastic processes [25]. It is well known that the synthesis and analysis formulations are equivalent for signal denoising when the sparsifying transform (or dictionary) is orthogonal and when the regularization functional is chosen to be the ℓ_1-norm of the expansion coefficients [26].

While the switch from an ℓ_2 to an ℓ_1 regularization necessitates the deployment of more sophisticated algorithms [8], [27], [28], there is increasing evidence that it results in higher-quality signal reconstructions, especially in the more challenging cases (compressed sensing) when there are less measurements than unknowns. The theory of compressed sensing also provides some guarantees of recovery for K-sparse signals under strict assumptions on the system matrix [6], [29], [30].

Our objective in this paper is to characterize and compare the effect of the two primary types of regularization on the solution of general convex optimization problems involving real-valued linear measurements. While the sparsity inducing property of the ℓ_1-norm is well documented and reasonably well understood by practitioners, we are only aware of a few mathematical results that make this explicit with the view of solving underdetermined systems of linear equations (e.g. [4], [31], [32]), typically under the assumption that the ℓ_1-minimizer is unique.

We have chosen to present our findings in the form of a series of representer theorems which go by pairs (ℓ_2 vs. ℓ_1 regularization) with all other aspects of the problem—*i.e.*, the choice of the (convex) data term and the regularization operator—being the same. A pleasing outcome is that our results reinforce the connection between the synthesis and analysis formulations of signal recovery since our ℓ_1 representer theorems can be interpreted as a "synthesis" solution to a class of optimiza-

Manuscript received February 2, 2016; revised June 20, 2016; accepted June 29, 2016. Date of publication August 4, 2016; date of current version August 16, 2016. This work was supported in part by the European Research Council through the European Unions Seventh Framework Programme (FP7/2007-2013) under ERC Grant 267439 (FUN-SP) and in part by the Romberg Professorship through the Interdisciplinary Center for Scientific Computing, University of Heidelberg.

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: michael.unser@epfl.ch; julien.fageot@epfl.ch; harshit.gupta@epfl.ch).

Communicated by H. Rauhut, Associate Editor for Signal Processing.
Digital Object Identifier 10.1109/TIT.2016.2590421

tion problems that is more typical of the regularization framework.

The paper is organized as follows. The scene is set in Section II with a brief discussion and comparison of the two primary schemes for signal recovery: the linear Tikhonov estimator (with $p = 2$) versus the non-linear basis pursuit estimator (with $p = 1$), where the only change is in the exponent of the regularization. We also document the property that one is able to control the sparsity of the latter estimator by varying the regularization parameter λ. In Section III, we focus on the finite-dimensional scenario ($\mathbf{x} \in \mathbb{R}^N$) and present our two main representer theorems that cover a broad family of convex optimization problems. As example of application, we prove that the extreme points of a total-variation optimization problem are necessarily piecewice-constant. In Section IV, we generalize our result to the infinite-dimensional setting ($x \in \ell_1(\mathbb{Z})$). The formulation becomes more technical as we need to invoke the weak$*$ topology to specify the full solution set of the generic ℓ_1-norm minimization problem. We also consider the scenario where the null space of the regularization operator L is non-trivial, which requires some more sophisticated developments (Theorem 19). The bottom line is that the generic form of the solution remains unchanged, while the sparsifying effect of ℓ_1-regularization is even more dramatic: the minimization process results into the collapsing of an infinity of degrees of freedom into a small finite number that is upper bounded by the number of measurements.

II. Motivation: ℓ_2 Versus ℓ_1 Regularization

In a linear inverse problem, the task is to recover some unknown signal $\mathbf{x} \in \mathbb{R}^N$ from a noisy set of linear measurements $\mathbf{y} = (y_1, \dots, y_M) \in \mathbb{R}^M$ such that

$$y_m = \langle \mathbf{h}_m, \mathbf{x} \rangle + n[m], \quad m = 1, \dots, M \tag{1}$$

where $n[m]$ is some unknown noise component that is typically assumed to be i.i.d. Gaussian. The measurement model is specified by the real-valued system matrix $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_M]^T$ of size $M \times N$. Our interest here is in the ill-posed scenario where M is (much) smaller than N (compressed sensing) or when the system matrix is poorly conditioned and not invertible. This ill-posedness is dealt with in practice by introducing some form of regularization. Since our objective here is to compare the regularizing effect of ℓ_2 vs. ℓ_1 norms, we shall start with the simplest scenario where the regularization is imposed directly upon $\mathbf{x}$. The more general case where the regularization is enforced in some transformed domain is addressed in the second part of Section III.

A. Simple Regularized Least-Squares Estimator

The most basic penalized least-squares (or Tikhonov) estimator of the signal $\mathbf{x}$ from the measurements $\mathbf{y}$ is specified by

$$\mathbf{x}_{\mathrm{LS}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2, \tag{2}$$

where $\lambda > 0$ is a hyper parameter that controls the strength of the regularization. The standard form of the solution is

$$\mathbf{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_N)^{-1}\mathbf{H}^T\mathbf{y},$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. This translates into a linear algorithm that can also be interpreted as a Wiener filter. We shall now invoke a lesser-known result that has some interesting conceptual implications. The proof is given in Appendix A for sake of completeness.

Proposition 1: For any matrix $\mathbf{H}$ of size $M \times N$ and $\lambda \in \mathbb{R}^+$, we have the identity

$$(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_N)^{-1}\mathbf{H}^T = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_M)^{-1}.$$

This allows us to rewrite the least-squares solution as

$$\mathbf{x}_{\mathrm{LS}} = \mathbf{H}^T\mathbf{a} = \sum_{m=1}^{M} a_m\mathbf{h}_m \tag{3}$$

where $\mathbf{a} = (\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_M)^{-1}\mathbf{y}$. We have thereby revealed the property that $\mathbf{x}_{\mathrm{LS}} \in \mathrm{span}\{\mathbf{h}_m\}_{m=1}^M$. Moreover, if we let $\lambda \to 0$, then the solution converges to $\mathbf{x}_0 = \mathbf{H}^+\mathbf{y}$ where $\mathbf{H}^+$ is the Moore-Penrose generalized inverse of $\mathbf{H}$ [2, Section 1.5.2]. By definition, $\mathbf{H}^+$ solves the classical least-squares approximation problem $\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ and extracts the solution $\mathbf{x}_0$ that has the minimum norm. If $\mathbf{H}^T\mathbf{H}$ is of full rank, then $\mathbf{H}^+ = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$, which is the classical pseudo-inverse of $\mathbf{H}$. Otherwise, which is the case of interest here, it returns the minimum-norm solution that is in the span of $\mathbf{H}^T$ as well.

While this simple linear reconstruction scheme works reasonably well when $M \geq N$, the situation is much less favorable for smaller M because the solution is forced to live in a space that is specified by the system matrix $\mathbf{H}$, and hence strongly problem-dependent.

B. Least-Squares Estimator With ℓ_1 Penalty

An alternative that has become increasingly popular in recent years is to substitute the squared ℓ_2-norm penalty by the ℓ_1-norm. This yields the so-called penalized basis pursuit (PBP) estimator

$$\mathbf{x}_{\mathrm{sparse}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{4}$$

where $\lambda > 0$ is a regularization parameter with the same role as before. To get some insight on the effect of the ℓ_1 regularization, we now look at the extreme scenario where there is a single measurement:

$$\min_{\mathbf{x}\in\mathbb{R}^N} |y_1 - \mathbf{h}_1^T\mathbf{x}|^2 + \lambda\|\mathbf{x}\|_1$$

For λ above some critical threshold, we get the trivial solution $\mathbf{x} = \mathbf{0}$. Otherwise, we obtain a "sparse" solution of the form

$$\mathbf{x}_{\mathrm{sparse}} = a_1\mathbf{e}_{n_1}$$

where $\{\mathbf{e}_n\}_{n=1}^N$ is the canonical basis of $\mathbb{R}^N$ and n_1 the index of the component of $\mathbf{h}_1$ that has the largest magnitude. This has to be compared with the corresponding ℓ_2 solution (3) which simplifies to

$$\mathbf{x}_{\mathrm{LS}} = a_1\mathbf{h}_1$$

with $a_1 = y/(\mathbf{h}_1^T\mathbf{h}_1 + \lambda)$. The contrast is striking: On the one hand, we have a solution that is completely sparse with

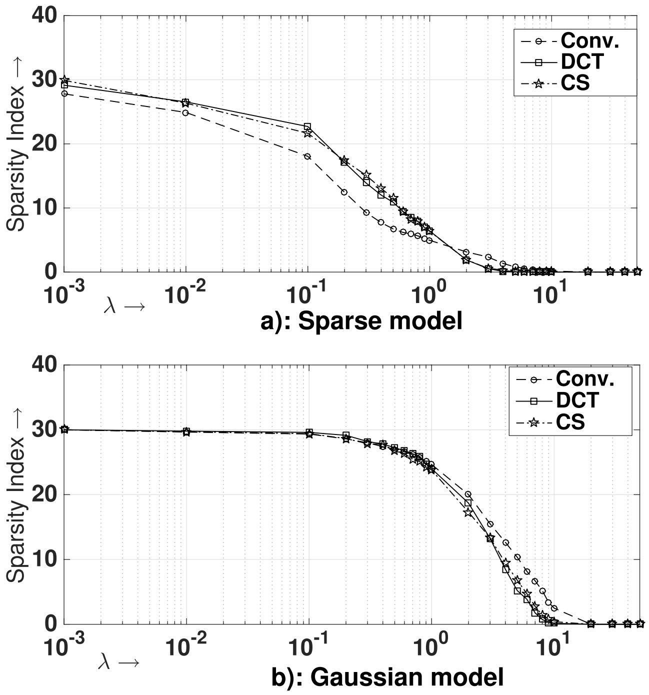

Fig. 1. Sparsity index $K = \|\mathbf{x}_{\text{sparse}}\|_0$ of the solution of (4) as a function of the regularization parameter λ for three different kinds of system matrices with $M = 30$ and $N = 120$. The simulated measurement model is described by (1) where $\mathbf{x}$ is i.i.d. (innovation model) with two distinct statistical models: (a) Bernoulli-Gauss with an average sparsity index of 8, (b) Gaussian and therefore, non-sparse. In all the simulations, the standard deviations of the signal is $\sigma_x = 3$, the input SNR is 35 dB and all the matrices are row normalized. The rows of the system matrix $\mathbf{H}$ are constructed as follows. (i) **Conv**: random shifts of a symmetric exponential $\mathrm{e}^{-|m|}$; (ii) **DCT**: random rows of a DCT matrix; (iii) **CS**: Gaussian i.i.d. with $\sigma = 1$. In all the cases, K reaches $M = 30$ as $\lambda \to 0$.

$\|\mathbf{x}_{\text{sparse}}\|_0 = 1$, while, on the other, we obtain a blurred rendition whose parametric form is dictated by the measurement vector $\mathbf{h}_1$. As it turns out, the contrasting behavior that has been identified for this very simple scenario is generic and transposable to a much broader class of optimization problems.

The other property that is well documented in the literature is the sparsifying effect of the regularization parameter λ in (4). When λ is very small and close to 0, the solution will typically have a sparsity index $\|\mathbf{x}_{\text{sparse}}\|_0 = M$ where M is the number of measurements. In order to promote sparser solutions, it then suffices to increase λ, as illustrated in Fig. 1. To show that this mechanism is universal and unrelated to the choice of the system matrix, we considered three representative scenario : (i) symmetric exponential convolution followed by a non-uniform sampling (Conv), (ii) random sampling of the discrete cosine transform (DCT) of the signal, and (iii) compressed sensing (CS) involving a system matrix whose components are i.i.d. Gaussian. The simulated measurements were generated according to (1) where $n[m]$ is AWG noise. The reconstruction was then performed using FISTA [27] for unconstrained ℓ_1 minimization. To verify that the control mechanism is independent of the suitability of the underlying signal model, we considered two extreme configurations. In the first set of simulations summarized in Fig. 1a, the ground-truth signal $\mathbf{x}$ is truly sparse with its majority of coefficients being zero—specifically, the components of $\mathbf{x}$ are i.i.d. with a Bernoulli-Gauss distribution. For the second set of experiments shown in Fig. 1b, we switched to a "non-sparse" model by taking $\mathbf{x}$ to be i.i.d. Gaussian. While there are differences in the shape of the graphs, the main point is that in all cases, $K = \|\mathbf{x}_{\text{sparse}}\|_0$ decreases monotonically with λ while its maximum value is bounded by M.

Besides the standard PBP form (4) favored by practitioners, there are two other possible formulations of the recovery problem. The first is the LASSO (Least Absolute Shrinkage and Selection Operator) defined as (see [26])

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \quad \text{s. t. } \|\mathbf{x}\|_1 \le \tau, \tag{5}$$

while the second is the (quadratically) constrained basis pursuit (CBP) estimator

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{s. t. } \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \le \sigma. \tag{6}$$

The key property for our purpose is that for any value of $\lambda \in \mathbb{R}^+$ in (4), it is possible to find some corresponding $\tau = \tau(\lambda) \ge 0$ and $\sigma = \sigma(\lambda) \ge 0$ (and vice versa) so that the PBP, LASSO, CBP problems are rigorously equivalent (see [32, Proposition 3.2, p 64]). The argument is that the minimizer of (6) $\big($resp., (5)$\big)$ saturates the inequality, which allows us to interpret (4) as the unconstrained form of the same minimization problem with Lagrange multiplier λ. The optimal tajectory $(\tau(\lambda), \sigma(\lambda))$ that is parametrized by λ is called the Pareto curve [33]. The same equivalence obviously also holds for $p = 2$. While the constrained version of the problem (6) with σ fixed is typically harder to solve numerically than (4), it is actually the form that lends itself best to a mathematical analysis, as we shall see next.

III. Finite Dimensional Analysis

In order to derive the general form of the solution of linear inverse problems with ℓ_2 versus ℓ_1 regularization constraints, we shall first enlarge the class of problems of interest by considering some arbitrary convex constraints on the so-called data term which involves the measurements $\mathbf{y}$. While this has the advantage of providing more general results, it has the even more remarkable effect of simplifying the mathematical derivations because it puts the problem in an abstract perspective that is more suitable for functional analysis.

A. Preliminaries

Let us start with a few definitions where $\mathcal{X}$ stands for an arbitrary (finite or infinite-dimensional) topological vector space. In this section, $\mathcal{X} = \mathbb{R}^N$.

Definition 2: A subset $\mathcal{C}$ of a vector space $\mathcal{X}$ is convex if $z = (tx + (1-t)y) \in \mathcal{C}$, for any $x, y \in \mathcal{C}$ and $t \in [0, 1]$; that is, if all the points that lie on the line connecting x to y are also included in $\mathcal{C}$.

Definition 3 (Projection on a Closed Convex Set): Let $\mathcal{X}$ be a vector space equipped with some norm $\|\cdot\|$. Then, the projection set of z on the closed convex set $\mathcal{U} \subset \mathcal{X}$ is

$$\arg\min_{x\in\mathcal{U}} \|x - z\| = \{x_0 \in \mathcal{U} : \|x_0 - z\| \le \|x - z\|, \ \forall x \in \mathcal{U}\}.$$

When the projection set reduces to a single point x_0, then x_0 is called the projection of z on $\mathcal{U}$ and we write

$$x_0 = \arg\min_{x\in\mathcal{U}} \|x - z\|. \tag{7}$$

To make the connection with the signal recovery formulations of Section II, we define the data-dependent closed convex set

$$\mathcal{U}(\mathbf{y};\sigma) = \left\{\mathbf{x}\in\mathbb{R}^N : \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \le \sigma\right\}$$

where σ is an adjustable control parameter. Hence, the classical projection problem (7) with $z = 0$ and $\mathcal{U} = \mathcal{U}(\mathbf{y};\sigma)$ yields the constrained form—that is, the CBP estimator (6)—of our initial signal recovery problem.

A potential difficulty when dealing with general convex optimization problems is that not all such problems have unique solutions. It is possible, however, to give a complete description of the solution set in terms of its extreme points.

Definition 4 (Extreme Point): Let E be a convex subset of some vector space $\mathcal{X}$. An extreme point of E is a point $x \in E$ that does not lie in any open line segment joining two distinct points of E.

The extreme points of a convex optimization problem are very special in that they lie on the frontier of the convex solution set which is then given by their convex hull. Obviously, the problem has a unique solution if and only if it has a single extreme point, as is generally the case with the ℓ_2-norm.

B. Finite-Dimensional Representer Theorems

Having set the context, we now proceed with the presentation of representer theorems for a broad family of convex optimization problems in relation to the type of regularization.

Theorem 5 (Convex Problem With ℓ_2 Minimization): Let $\mathrm{H} : \mathbb{R}^N \to \mathbb{R}^M : \mathbf{x} \mapsto \mathbf{H}\mathbf{x}$ with $M \le N$ be a linear measurement operator and $\mathcal{C}$ be a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\mathbb{R}^N$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C}) = \{\mathbf{x}\in\mathbb{R}^N : \mathbf{H}\mathbf{x}\in\mathcal{C}\}$, is nonempty (feasibility hypothesis). Then,

$$\mathcal{V} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_2^2 \quad \text{s.t. } \mathbf{H}\mathbf{x}\in\mathcal{C} \tag{8}$$

has a unique extreme point of the form

$$\mathbf{x}_{\mathrm{LS}} = \sum_{m=1}^{M} a_m \mathbf{h}_m = \mathbf{H}^T\mathbf{a}. \tag{9}$$

Proof: First, we observe that convexity (resp., closedness) is preserved through linear (resp., continuous) transformations so that the preimage $\mathcal{U}$ of $\mathcal{C}$ is guaranteed to be closed convex as well. In view of Definition 3, the solution is thereby given by the projection of the origin $\mathbf{z} = \mathbf{0}$ onto the closed convex set $\mathcal{U}$, which is known to be nonempty, because of the feasibility hypothesis. Our claim of unicity then follows from Hilbert's famous projection theorem for convex sets which states that the projection on a convex set in a Hilbert space always exists and reduces to a single point [34]. The Hilbert space here is $\mathbb{R}^N$ equipped with the inner product $\langle\mathbf{x}_1,\mathbf{x}_2\rangle = \mathbf{x}_1^T\mathbf{x}_2$.

Let $\mathbf{x}_0 = \arg\min_{\mathbf{x}\in\mathcal{U}}\|\mathbf{x}\|$ denote the unique solution of (8) and $\mathbf{y}_0 = \mathbf{H}\mathbf{x}_0$ be the image of that point through the measurement operator. Because the linear map $\mathbf{x}_0 \mapsto \mathbf{y}_0$ is consistent (*i.e.*, $\mathbf{y}_0 \in \mathcal{C}$) and the projection has minimum norm, the operation is reversible with $\mathbf{x}_0 = \mathbf{H}^+\mathbf{y}_0$ where $\mathbf{H}^+$ is the Moore-Penrose generalized inverse of $\mathbf{H}$ whose range is in the span of $\mathbf{H}^T$ (see brief discussion of the property of this inverse in Section II-A). In other words, there exists a unique $\mathbf{a}\in\mathbb{R}^N$ such that $\mathbf{x}_{\mathrm{LS}} = \mathbf{x}_0 = \mathbf{H}^T\mathbf{a} = \mathbf{H}^+(\mathbf{H}\mathbf{H}^T\mathbf{a})$. ∎

Let us note that the result in Theorem 5 is consistent with the elementary analysis of the Tikhonov estimator in Section II. Remarkably, the generic form of the solution remains valid for the complete class of convex optimization problems involving the same linear measurement model and the same quadratic regularization functional $\|\mathbf{x}\|_2^2$. The catch, of course, is that the general solution map is no longer linear. In other words, we should view Theorem 5 (as well as all subsequent representer theorems) as an existence/discretization result, meaning that it is still necessary to deploy some iterative algorithm (such a steepest-descent method) to actually find the optimal expansion vector $\mathbf{a}$.

We now present the ℓ_1 counterpart of Theorem 5. The statement of the problem is almost identically except for the fact that there can now be multiple extreme points.

Theorem 6 (Convex Problem With ℓ_1 Minimization): Let $\mathrm{H} : \mathbb{R}^N \to \mathbb{R}^M : \mathbf{x} \mapsto \mathbf{H}\mathbf{x}$ with $M < N$ be a linear measurement operator and $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\mathbb{R}^N$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C}) = \{\mathbf{x}\in\mathbb{R}^N : \mathbf{H}\mathbf{x}\in\mathcal{C}\}$, is nonempty (feasibility hypothesis). Then,

$$\mathcal{V} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{H}\mathbf{x}\in\mathcal{C}$$

is a nonempty, convex, compact subset of $\mathbb{R}^N$ with extreme points $\mathbf{x}_{\mathrm{sparse}}$ of the form

$$\mathbf{x}_{\mathrm{sparse}} = \sum_{k=1}^{K} a_k \mathbf{e}_{n_k} \tag{10}$$

with $K \le M$, $\{\mathbf{e}_n\}_{n=1}^N$ the canonical basis of $\mathbb{R}^N$, $n_k \in \{1,\dots,N\}$ for $k = 1,\dots,K$, and $\|\mathbf{x}_{\mathrm{sparse}}\|_1 = \sum_{k=1}^K |a_k|$.

Proof: Since $\mathcal{C}$ is convex (resp., closed) and H is linear (resp., continuous), the set $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$ is convex (resp., closed) as well. Therefore, $\mathcal{U}$ is a nonempty, convex, and closed subset of $\mathbb{R}^N$.

The function $\|\mathbf{x}\|_1$ is continuous from $\mathbb{R}^N \to \mathbb{R}^+$, and therefore admits a minimum (not necessarily unique) over any closed set, including $\mathcal{U}$, which ensures that $\mathcal{V}$ is nonempty. Therefore, let $\alpha = \min_{\mathbf{x}\in\mathcal{U}}\|\mathbf{x}\|_1$ and $\mathcal{B}$ be the closed ball of radius α for the ℓ_1-norm; that it, $\mathcal{B} = \{\mathbf{x}\in\mathbb{R}^N,\ \|\mathbf{x}\|_1 \le \alpha\}$. Then, the set $\mathcal{V} = \mathcal{U}\cap\mathcal{B}$ is convex and compact, as the intersection of a convex closed set with a convex compact set. Ultimately, this translates into $\mathcal{V}$ being nonempty, convex, and compact.

This allows us to invoke the Krein-Milman theorem (see [35, p. 75]), which tells us that a convex compact set, such as $\mathcal{V}$, is the closed convex hull of its extreme points.

Let us now consider an extreme point $\mathbf{x} = (x_1,\dots,x_N)$ of $\mathcal{V}$ whose number of non-zero entries is denoted by K, with a priori $K \in \{0,\dots,N\}$. We want to prove that $\mathbf{x}$ is of the form (10), which is equivalent to $K \le M$.

We shall proceed by contradiction and assume that $K = \|\mathbf{x}\|_0 \geq M+1$, meaning that there exists (at least) $(M+1)$ indices $n_1, \dots, n_{M+1} \in \{1, \dots N\}$ such that $x_{n_m} \neq 0$ for every m. We set $\bar{\mathbf{x}} = \mathbf{x} - \sum_{m=1}^{M+1} x_{n_m} \mathbf{e}_{n_m}$ and, for $m = 1, \dots, M+1$,

$$\mathbf{y}_m = x_{n_m} \mathbf{H} \mathbf{e}_{n_m}. \tag{11}$$

Since any collection of $(M+1)$ vectors in $\mathbb{R}^M$ is linearly dependent, there exists some constants c_m such that $\mathbf{c} = (c_1, \dots, c_{M+1}) \neq \mathbf{0}$ and $\sum_{m=1}^{M+1} c_m \mathbf{y}_m = \mathbf{0}$. Correspondingly, we define $\mathbf{x}_0 = \sum_{m=1}^{M+1} c_m x_{n_m} \mathbf{e}_{n_m}$. We also pick an $\epsilon \in \mathbb{R}$ with $|\epsilon| < 1/\max_m |c_m|$ such that $(1 + \epsilon c_m) > 0$ and $(1 - \epsilon c_m) > 0$ for all m. Since $\mathbf{H}\mathbf{x}_0 = \sum_{m=1}^{M+1} c_m \mathbf{y}_m = \mathbf{0}$, we have that $\mathbf{H}\mathbf{x} = \mathbf{H}(\mathbf{x} - \epsilon \mathbf{x}_0) = \mathbf{H}(\mathbf{x} + \epsilon \mathbf{x}_0)$ so that $(\mathbf{x} + \epsilon \mathbf{x}_0)$ and $(\mathbf{x} - \epsilon \mathbf{x}_0)$ are in $\mathcal{U}$. Moreover, because $\mathbf{x}_0$ and $\bar{\mathbf{x}}$ have disjoint support, we have that

$$\begin{aligned}
\|\mathbf{x} \pm \epsilon \mathbf{x}_0\|_1 &= \|\bar{\mathbf{x}} + \sum_{m=1}^{M+1} (1 \pm \epsilon c_{n_m}) x_{n_m} \mathbf{e}_{n_m}\|_1 \\
&= \|\bar{\mathbf{x}}\|_1 + \sum_{m=1}^{M+1} (1 \pm \epsilon c_{n_m}) |x_{n_m}| \\
&= \|\mathbf{x}\|_1 \pm \epsilon \sum_{m=1}^{M+1} c_{n_m} |x_{n_m}| \\
&= \alpha \pm \epsilon \sum_{m=1}^{M+1} c_{n_m} |x_{n_m}|.
\end{aligned}$$

If $\sum_{m=1}^{M+1} c_{n_m} |x_{n_m}| \neq 0$, then $(\mathbf{x} + \epsilon \mathbf{x}_0)$ or $(\mathbf{x} - \epsilon \mathbf{x}_0)$ has a ℓ_1-norm strictly smaller than α, which is impossible since the minimum over $\mathcal{U}$ is α. Hence, $\sum_{m=1}^{M+1} c_{n_m} |x_{n_m}| = 0$, and

$$\|\mathbf{x} + \epsilon \mathbf{x}_0\|_1 = \|\mathbf{x} - \epsilon \mathbf{x}_0\|_1 = \alpha. \tag{12}$$

In other words, there exists $\epsilon > 0$ such that $(\mathbf{x} + \epsilon \mathbf{x}_0)$, $(\mathbf{x} - \epsilon \mathbf{x}_0) \in \mathcal{V}$, which implies that

$$\mathbf{x} = \frac{1}{2}(\mathbf{x} + \epsilon \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \epsilon \mathbf{x}_0) \tag{13}$$

is not an extreme point of $\mathcal{V}$. This proves that $K \leq M$. ■

We like to mention a related result [32, Th. 3.1, p. 62] on the maximal cardinality of the support of the solution of the problem

$$\min \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{H}\mathbf{x} = \mathbf{y}$$

under the assumption that the problem admits a unique minimizer. It is also indicated there that the result does not carry over to the complex setting. Theorem 6 constitutes a substantial extension as it applies to a much broader class of problems—it also provides the structure of the full solution set for the more typical cases where the minimizer is not unique.

C. Incorporation of a Regularization Operator

To cover a broader spectrum of applications, we are also interested in problems involving a regularization operator or a dictionary. We shall now see that this extension is straightforward when the regularization functional is *coercive*; that is, when there exists a constant $A > 0$ such that $A\|\mathbf{x}\|_p \leq \|\mathbf{L}\mathbf{x}\|_p$ for all $\mathbf{x} \in \mathbb{R}^N$. In finite dimensions, this translates into $\mathbf{L}$ being an invertible matrix of size N. The analysis of the more challenging non-coercive scenario is deferred to Section IV-C.

Corollary 7 (Convex problem with ℓ_2 regularization): Let $\mathrm{H} : \mathbb{R}^N \to \mathbb{R}^M : \mathbf{x} \mapsto \mathbf{H}\mathbf{x}$ with $M < N$ be a linear measurement operator and $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\mathbb{R}^N$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$, is nonempty (feasibility hypothesis). $\mathbf{L}$ is an invertible regularization matrix of size N that can be chosen arbitrarily. Then,

$$\mathcal{V} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{L}\mathbf{x}\|_2^2 \quad \text{s.t. } \mathbf{H}\mathbf{x} \in \mathcal{C} \tag{14}$$

has a unique solution of the form

$$\mathbf{x}_{\mathrm{LS}} = \sum_{m=1}^{M} a_m \widetilde{\mathbf{h}}_m = \widetilde{\mathbf{H}}^T \mathbf{a}. \tag{15}$$

where $\widetilde{\mathbf{H}}^T = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{H}^T$.

Proof: Since $\mathbf{L}$ is invertible, we define the auxiliary variable $\mathbf{u} = \mathbf{L}\mathbf{x}$, which allows us to rewrite $\mathbf{y} = \mathbf{H}\mathbf{x} = \mathbf{G}\mathbf{u}$ with $\mathbf{G} = \mathbf{H}\mathbf{L}^{-1}$. Likewise, the convex set $\mathcal{C}$ in the space of the measurements $\mathbf{y}$ is linearly mapped into a nonempty convex set $\widetilde{\mathcal{U}}$ in the space of the auxiliary variable $\mathbf{u}$. We then apply Theorem 5, which yields the generic solution

$$\mathbf{u}_{\mathrm{LS}} = \mathbf{G}^T \mathbf{a} \Leftrightarrow \mathbf{x}_{\mathrm{LS}} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{H}^T \mathbf{a}.$$

■

Corollary 8 (Convex Problem With ℓ_1 Regularization): Let $\mathrm{H} : \mathbb{R}^N \to \mathbb{R}^M : \mathbf{x} \mapsto \mathbf{H}\mathbf{x}$ with $M < N$ be a linear measurement operator and $\mathcal{C}$ be a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\mathbb{R}^N$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$, is nonempty (feasibility hypothesis). $\mathbf{L}$ is an invertible regularization matrix of size N that can be chosen arbitrarily. Then,

$$\mathcal{V} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{L}\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{H}\mathbf{x} \in \mathcal{C} \tag{16}$$

is a nonempty, convex, compact subset of $\mathbb{R}^N$ with extreme points $\mathbf{x}_{\mathrm{sparse}}$ of the form

$$\mathbf{x}_{\mathrm{sparse}} = \sum_{k=1}^{K} a_k \mathbf{g}_{n_k} \tag{17}$$

with $K \leq M$ and $\|\mathbf{L}\mathbf{x}_{\mathrm{sparse}}\|_1 = \sum_{k=1}^{K} |a_k|$. The basis vectors $\mathbf{g}_{n_k}$ with indices $n_k \in \{1, \dots, N\}$ for $k = 1, \dots, K$ are taken within the N-dimensional dictionary

$$\mathbf{G}^T = [\mathbf{g}_1 \cdots \mathbf{g}_N] = (\mathbf{L}^{-1})^T.$$

Proof: The proof here too is based on the direct application of Theorem 6 with the auxiliary variable $\mathbf{u} = \mathbf{L}\mathbf{x}$. ■

The remarkable outcome is that the reconstruction space is now entirely determined by the regularization operator $\mathbf{L}$, and independent of the measurement setup, in sharp contrast with the ℓ_2 scenario in Corollary 7.

Corollary 8 tells us that the extreme points of the optimization problem (16) are constructed by picking the "best" $K \ll N$ elements within a dictionary that is specified by the row vectors of $\mathbf{L}^{-1}$. While this proves that the solution set is intrinsically sparse, it is primarily an existence result because Theorem 6 does not tell us which elements to pick (i.e., the value of the index n_k) nor the values of the weights a_k. Again, the powerfull aspect here is the generality of the result since it applies to a complete class of convex optimization problems.

An alternative formulation could be to specify the augmented vector $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$ with the implicit understanding that one is restricting our choice of candidates to those that are K-sparse with $K \leq M$. The optimal configuration would then be achieved when $\|\mathbf{a}\|_1 = \min_{\mathbf{x}\in\mathcal{U}} \|\mathbf{Lx}\|_1$.

D. The Special Case of Total Variation

To make the connection with the popular "total variation" scenario, we take $\mathbf{L} = \mathbf{D}$ as the finite difference operator

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad (18)$$

Its inverse is given by

$$\mathbf{D}^{-1} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 1 & \cdots & 1 & 1 \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

which is an upper triangular matrix of ones.

The interpretation of Corollary 8 is that the corresponding solution will then be formed by selecting a few rows of $\mathbf{D}^{-1}$ (or columns of $\mathbf{D}^{-1T}$), which results in a solution that is piecewise-constant with K jumps of amplitude a_k. The total variation of the solution is then measured by the ℓ_1 norm of the coefficient vector $\mathbf{a} = (a_1, \dots, a_K)$; *i.e.*,

$$\|\mathbf{x}_{\text{sparse}}\|_{\text{TV}} = \|\mathbf{D}\mathbf{x}_{\text{sparse}}\|_{\ell_1} = \|\mathbf{a}\|_{\ell_1}.$$

This is consistent with one of the earliest schemes used to solve compressed sensing problems [29]. The interest of our theorem is that it explains why the optimization of total variation *always* admits a piecewise-constant solution. While this behavior is well known and amply documented in the literature, we are not aware of any prior mathematical analysis that shows that the generic form of the solution (piecewise-constant) is actually independent of $\mathbf{H}$.

By contrast, there is no such decoupling in the ℓ_2 scenario where the influence of the regularization and the characteristic footprint of the system matrix are intertwined. Specifically, Corollary 7 tells us that the basis functions are now given by $\widetilde{\mathbf{H}}^T = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{H}^T$ which amounts to some smoothed version (doubly integrated) of $\mathbf{H}^T$. In particular, if $\mathbf{H}^T$ is taken to be the identity or a non-uniform sampling matrix, then the ℓ_2 solution becomes piecewise-linear with breakpoints (or knots) at the sampling locations, which is a rather different type of signal.

We conclude this section with an important remark concerning our use of the above finite-difference matrix. Indeed, another choice could have been the circulant matrix

$$\mathbf{D}_{\text{LSI}} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ -1 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad (19)$$

which is almost the same as (18), except for the additional -1 in the lower right. Now the major difference between $\mathbf{D}$ and $\mathbf{D}_{\text{LSI}}$ is that the latter, which maps into a circular convolution, annihilates constants. While this property is very desirable for regularization purposes, its downside for the present demonstration is that it spoils the invertibility requirement for the application of Corollaries 7 and 8. To handle such cases, we need the non-coercive counterparts of these results, which are presented in Section IV-C. At any rate, the bottom line for total variation is that the piecewise-constant form of the solution is preserved in either cases, the main difference being that $\mathbf{D}_{\text{LSI}}$ does not penalize constant signals (spanned by the first row vector of $\mathbf{D}^{-1}$).

IV. Infinite Dimensional Analysis

We will now extend our analysis to the infinite dimensional setting. While the basic ideas underlying the proofs remain the same, the formulation becomes more technical because we have to properly deal with topological issues; in particular, the complication that the unit ball in $\ell_1(\mathbb{Z})$ is no longer compact. Another substantial generalization is that we are also treating the very relevant case of regularization operators whose null space is non-trivial. To help the readers who are not so much at ease with functional analysis, we have done our best to clarify the presentation by including tutorial explanations.

A. Notation

Following the standard convention in signal processing, discrete signals or sequences are indexed using square brackets with the index running over $\mathbb{Z}$; i.e., $x[n]$ denotes the sample of the signal $x = x[\cdot]$ at location $n \in \mathbb{Z}$. Likewise, the infinite-dimensional counterpart of the canonical basis $\{\mathbf{e}_n\}_{n=1}^N$ is $\{\delta[\cdot - n]\}_{n\in\mathbb{Z}}$ where $\delta[\cdot - n_0]$ denotes the unit impulse at some fixed location n_0 (the dot "$\cdot$" is a placeholder for the domain variable of the input that is used to avoid notational confusion).

Instead of matrices, we shall now consider linear operators acting on suitable Banach spaces. These are denoted by capital letters. For instance, the operator $\mathrm{G} : \mathcal{X} \to \mathcal{Y}$ maps the space $\mathcal{X}$ (the domain of the operator) into $\mathcal{Y}$; its action is denoted by $x \mapsto y = \mathrm{G}\{x\}$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

B. Infinite-Dimensional Representer Theorems

We first formulate the optimization problem in the real-valued Hilbert space $\ell_2(\mathbb{Z})$ equipped with the ℓ_2-inner product $\langle x, y\rangle = \sum_{n\in\mathbb{Z}} x[n]y[n]$.

Theorem 9 (Convex Problem With $\ell_2(\mathbb{Z})$ Minimization): Let us consider the following:

- $\mathrm{H}: \ell_2(\mathbb{Z}) \to \mathbb{R}^M$, $x \mapsto (\langle h_1, x\rangle, \dots, \langle h_M, x\rangle)$ is a linear measurement operator such that $\|\mathrm{H}\{x\}\| \le B\|x\|_{\ell_2}$ for some constant $B > 0$ and every $x \in \ell_2(\mathbb{Z})$;
- $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\ell_2(\mathbb{Z})$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C}) = \{x \in \ell_2(\mathbb{Z}) : \mathrm{H}\{x\} \in \mathcal{C}\}$, is nonempty (feasibility hypothesis).

Then, the problem

$$\mathcal{V} = \arg\min_{x \in \ell_2(\mathbb{Z})} \|x\|_{\ell_2} \quad \text{s.t. } \mathrm{H}\{x\} \in \mathcal{C} \tag{20}$$

has a unique solution of the form

$$x_{\mathrm{LS}} = \sum_{m=1}^{M} a_m h_m = \mathrm{H}^*\{\mathbf{a}\}. \tag{21}$$

Proof: The first part of the argument is the same as in the proof of Theorem 5. The linear operator H is bounded, and therefore continuous. Hence, $\mathcal{U}$ (the preimage of $\mathcal{C}$ through a linear and continuous transformation) is closed and convex, while the minimizer of (20) $x_{\mathrm{LS}} = x_0$ is unique (by Hilbert's projection theorem).

The second part is now handled in a softer manner by using a geometric argument. Let $\mathcal{M} = \mathrm{span}\{h_k\}_{k=1}^M$ and $\mathcal{M}^\perp = \{x \in \ell_2(\mathbb{Z}) : \mathrm{H}\{x\} = \mathbf{0}\}$ be the orthogonal complement of $\mathcal{M}$, which also coincides with the null-space of H. Since $\ell_2(\mathbb{Z}) = \mathcal{M} \oplus \mathcal{M}^\perp$, every $x \in \ell_2(\mathbb{Z})$ has a unique decomposition as $x = u + u^\perp$ with $u \in \mathcal{M}$ and $u^\perp \in \mathcal{M}^\perp$. Then, the solution x_0 can be written as $x_0 = u_0 + u_0^\perp$. Since $\mathrm{H}\{x_0\} = \mathrm{H}\{u_0\}$, x_0 and u_0 both lie in $\mathcal{U}$. As x_0 is the solution of (20), we have

$$\|x_0\|^2 \le \|u_0\|^2 \Rightarrow \|u_0 + u_0^\perp\|^2 = \|u_0\|^2 + \|u_0^\perp\|^2 \le \|u_0\|^2$$
$$\Rightarrow \|u_0^\perp\| = 0 \Leftrightarrow u_0^\perp = 0$$

Thus, $x_0 = u_0$ implying that $x_0 \in \mathrm{span}\{h_m\}_{m=1}^M$, which can be written in the form of (21). ■

As expected, (21) is the infinite-dimensional counterpart of (9) where the measurement vectors play the central role in the solution.

Let us now focus our attention on $\ell_1(\mathbb{Z})$, which is the Banach space associated with the norm $\|x\|_{\ell_1} = \sum_{n\in\mathbb{Z}} |x[n]|$. The complication there is to properly handle the potential issue of non-uniqueness. Since $\ell_1(\mathbb{Z})$ has an infinite number of dimensions, the unit ball $\mathcal{B} = \{x \in \ell_1(\mathbb{Z}) : \|x\|_{\ell_1} \le 1\}$ is not compact anymore for the Banach topology. However, by considering a weaker notion of convergence on $\ell_1(\mathbb{Z})$, we recover compactness and are able to generalize Theorem 6 for infinite sequences. The space of sequences that vanish at $\pm\infty$ is denoted by $c_0(\mathbb{Z})$. It is a Banach space when endowed with the supremum norm. The space $\ell_1(\mathbb{Z})$ is the topological dual of $c_0(\mathbb{Z})$. We can therefore define the weak*-topology on $\ell_1(\mathbb{Z})$; that is, the topology associated with the following notion of convergence: a sequence $(x_m)_{m\in\mathbb{N}}$ of elements of $\ell_1(\mathbb{Z})$ converges to 0 for the weak*-topology if

$$\sum_{n\in\mathbb{Z}} \left|a[n]x_m[n]\right| \underset{m\to\infty}{\longrightarrow} 0 \tag{22}$$

for every $a \in c_0(\mathbb{Z})$. Note that the sum $\sum_{n\in\mathbb{Z}} \left|a[n]x[n]\right|$ is always finite for $a \in c_0(\mathbb{Z})$ and $x \in \ell_1(\mathbb{Z})$. As suggested by the name, the weak*-topology is weaker than the usual Banach topology. Indeed, the convergence to 0 for the ℓ_1-norm implies the convergence to 0 for the weak*-topology due to the relation

$$\sum_{n\in\mathbb{Z}} \left|a[n]x[n]\right| \le \|a\|_{\ell_\infty}\|x\|_{\ell_1}. \tag{23}$$

We say that a subset of $\ell_1(\mathbb{Z})$ is weak*-closed (weak*-compact, respectively) if it is closed (compact, respectively) for the weak*-topology. The crucial point for us is that the ball $\mathcal{B}$ is weak*-compact in $\ell_1(\mathbb{Z})$, as implied by the Banach-Alaoglu theorem [35, p. 68]. For more details on the weak*-topology, we refer the reader to [35, Section 3.11].

Theorem 10 (Convex Problem With $\ell_1(\mathbb{Z})$ Minimization): Let us consider the following:

- $\mathrm{H}: \ell_1(\mathbb{Z}) \to \mathbb{R}^M$ is a linear measurement operator such that $\|\mathrm{H}\{x\}\| \le A\|x\|_{\ell_1}$ for some constant $A > 0$ and every $x \in \ell_1(\mathbb{Z})$;
- $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\ell_1(\mathbb{Z})$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C}) = \{x \in \ell_1(\mathbb{Z}) : \mathrm{H}\{x\} \in \mathcal{C}\}$, is nonempty (feasibility hypothesis).

Then,

$$\mathcal{V} = \arg\min_{x \in \ell_1(\mathbb{Z})} \|x\|_{\ell_1} \quad \text{s.t. } \mathrm{H}\{x\} \in \mathcal{C}$$

is a nonempty, convex, weak*-compact subset of $\ell_1(\mathbb{Z})$ with extreme points x_{sparse} of the form

$$x_{\mathrm{sparse}} = \sum_{k=1}^{K} a_k \delta[\cdot - n_k] \tag{24}$$

with $K = \|x_{\mathrm{sparse}}\|_0 \le M$, $n_k \in \mathbb{Z}$ for $k = 1, \dots, K$ and $\|x_{\mathrm{sparse}}\|_{\ell_1} = \sum_{k=1}^K |a_k|$.

Proof: The fact that $\mathcal{V}$ is nonempty, convex, and weak*-compact follows from classical theorems in convex analysis, as detailed in Appendix B. The form of the extreme points is then established using the same argumentation as in the proof of Theorem 6. ■

C. Extensions for Non-Coercive Regularization Functionals

In Section III-C, we have seen that there is no major difficulty in extending the representer theorems for more general scenarios involving an invertible regularization operator L. The concept carries over to infinite dimensions as well under the same assumption that the mapping is injective; that is, when the null space of the operator is trivial ($\mathcal{N}_{\mathrm{L}} = \{0\}$).

We shall now show that we can do much more and handle the non-coercive cases where the null space of the regularization operator

$$\mathcal{N}_{\mathrm{L}} = \left\{q : \mathbb{Z} \to \mathbb{R} \;\middle|\; \mathrm{L}\{q\} = 0\right\} = \mathrm{span}\{p_n\}_{n=1}^{N_0} \tag{25}$$

is finite dimensional of size N_0 where we are assuming that the p_n (basis elements) are linearly independent. The null space of L has a privileged role in the problem formulation because it incurs no penalty. This has the effect of promoting solutions whose null-space component is the largest possible. For instance, in the case of the finite-difference operator, any constant signal results in a zero-cost solution.

While such an extended setting is very attractive from a practical perspective, it introduces a higher level of difficulty because the operator L is no longer invertible in the usual (two-sided) sense. Yet, we shall see that it is still possible to specify some proper right inverse via the introduction of suitable boundary conditions. But prior to that, we need to spell out the conditions that ensure that an operator is well defined over $\ell_p(\mathbb{Z})$, the cases of interest being $p = 1, 2$.

In our framework, the concrete description of a linear operator G is provided by its kernel (or generalized impulse response) $G[k,l] = \mathrm{G}\{\delta[\cdot - l]\}[k]$. To make things more concrete, simply think of $G[k,l]$ as an infinite-dimensional matrix that is applied to the signal $x = x[\cdot] = (x[l])_{l\in\mathbb{Z}}$.

Definition 11: Given some sequence (or discrete signal) $x = (x[k])_{k\in\mathbb{Z}}$, we say that $\mathrm{G}\{x\}$ is well defined if

$$\sum_{l\in\mathbb{Z}} \big|G[k,l]x[l]\big| < +\infty$$

for any fixed $k \in \mathbb{Z}$ where $G[\cdot,\cdot]$ is the kernel of the operator. The output signal $\mathrm{G}\{x\}$ is then specified by $\mathrm{G}\{x\}[k] = \sum_{l\in\mathbb{Z}} G[k,l]x[l]$ for $k \in \mathbb{Z}$.

Definition 12: A sequence $x = (x[k])_{k\in\mathbb{Z}}$ is said to be of slow growth if there exists an integer $n_0 \in \mathbb{Z}$ and a constant $A > 0$ such that

$$|x[k]| \le A(1 + |k|)^{n_0} \quad \text{for all } k \in \mathbb{Z}.$$

The space of such sequences is denoted by $\mathcal{S}'(\mathbb{Z})$. It is the discrete counterpart of $\mathcal{S}'(\mathbb{R})$ (Schwartz's space of tempered distributions). As the notation suggests, $\mathcal{S}'(\mathbb{Z})$ is actually the topological dual of $\mathcal{S}(\mathbb{Z})$: the space of rapidly-decreasing sequences [36].

Proposition 13: The generic linear operator $\mathrm{G} : x \mapsto y = \mathrm{G}\{x\}$ is well defined over $\ell_p(\mathbb{Z})$ if and only if its kernel satisfies

$$\|G[k,\cdot]\|_{\ell_{p'}} < \infty \tag{26}$$

for any $k \in \mathbb{Z}$ where $p' = p/(p-1)$ is the conjugate exponent of $p \in [1,\infty]$. Moreover, G is bounded from $\ell_1(\mathbb{R}) \to \ell_\infty(\mathbb{R})$ if and only if

$$\sup_{k,l\in\mathbb{Z}} \big|G[k,l]\big| < \infty. \tag{27}$$

Proof: The sufficiency of (26) is established by using Hölder's inequality to construct the estimate

$$\sum_{l\in\mathbb{Z}} |G[k,l]x[l]| \le \|G[k,\cdot]\|_{\ell_{p'}} \|x\|_{\ell_p}.$$

Conversely, if there is some $k_0 \in \mathbb{Z}$ such that $G[k_0,\cdot] \notin \ell_{p'}(\mathbb{Z})$, we can construct a worst-case signal $x \in \ell_p(\mathbb{Z})$ such that $\sum_{l\in\mathbb{Z}} |G[k_0,l]x[l]|$ diverges (since the Hölder inequality is sharp).

By taking the supremum of the above estimate for $p = 1$, we get

$$\|y\|_{\ell_\infty} = \|\mathrm{G}\{x\}\|_{\ell_\infty} \le \left(\sup_{k,l\in\mathbb{Z}} |G[k,l]|\right) \|x\|_{\ell_1}.$$

for all $x \in \ell_1(\mathbb{Z})$, which shows that (27) implies that G is bounded from $\ell_1(\mathbb{Z}) \to \ell_\infty(\mathbb{Z})$. The necessity is established by considering $w_l = \delta[\cdot - l] \in \ell_1(\mathbb{Z})$ with $\|w_l\|_{\ell_1} = 1$. Since $\mathrm{G}\{w_l\} = G[\cdot,l]$, we have that

$$\|\mathrm{G}\{w_l\}\|_{\ell_\infty} = \|G[\cdot,l]\|_{\ell_\infty} = \sup_{k\in\mathbb{Z}} |G[k,l]|,$$

for any $l \in \mathbb{Z}$. On the other hand, the boundedness of G implies that

$$\|\mathrm{G}\{w_l\}\|_{\ell_\infty} = \sup_{k\in\mathbb{Z}} |G[k,l]| \le \|\mathrm{G}\| \tag{28}$$

where

$$\|\mathrm{G}\| \le \sup_{k,l\in\mathbb{Z}} |G[k,l]|.$$

Since (28) must hold for all $l \in \mathbb{Z}$ including the value that achieves the supremum, we conclude that the bound is sharp. ■

We are now ready to specify the vector spaces over which the global optimization is going to take place as

$$\ell_{p,\mathrm{L}}(\mathbb{Z}) = \{x : \mathbb{Z} \to \mathbb{R} \text{ s.t. } \|\mathrm{L}\{x\}\|_{\ell_p} < \infty\}, \tag{29}$$

with $p = 1, 2$. By definition, the operator L maps $\ell_{p,\mathrm{L}}(\mathbb{Z})$ into $\ell_p(\mathbb{Z})$. Our first step is to establish that $\ell_{p,\mathrm{L}}(\mathbb{Z})$ is a bona fide Banach space. The difficulty is that $\|\mathrm{L}\{\cdot\}\|_{\ell_p}$ is only a semi-norm on $\ell_{p,\mathrm{L}}(\mathbb{Z})$; that is, it has all the properties of a norm except that $\|\mathrm{L}\{x\}\|_{\ell_p} = 0$ does not imply that $x = 0$. This is resolved by factoring out the null space of the operator.

Proposition 14: Let $\|\cdot\|_{\mathcal{N}_\mathrm{L}}$ be some admissible norm for the finite-dimensional null space $\mathcal{N}_\mathrm{L}$ and $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}$ a projection operator from $\ell_{p,\mathrm{L}}(\mathbb{Z})$ into $\mathcal{N}_\mathrm{L}$. Then, $\ell_{p,\mathrm{L}}(\mathbb{Z})$ defined by (29) is a Banach space for the composite norm

$$\|\mathrm{L}\{x\}\|_{\ell_p} + \|\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{x\}\|_{\mathcal{N}_\mathrm{L}}.$$

Proof: We recall that the elements x_Q of the quotient space $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L}$ are equivalence classes on $\ell_{p,\mathrm{L}}(\mathbb{Z})$ such that, for $x \in \ell_{p,\mathrm{L}}(\mathbb{Z})$, $x_Q = \{x + q : q \in \mathcal{N}_\mathrm{L}\}$. Since the quotient space does not distinguish between elements $x, y \in \ell_{p,\mathrm{L}}(\mathbb{Z})$ such that $x - y \in \mathcal{N}_\mathrm{L}$, we can endow it with the norm $\|\mathrm{L}\{x_Q\}\| := \|\mathrm{L}\{x\}\|_p$, where x is any member of the equivalence class x_Q. This shows that $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L}$ is a Banach space, while the same property obviously holds for $\mathcal{N}_\mathrm{L}$. It follows that the direct sum of those two spaces, $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L} + \mathcal{N}_\mathrm{L}$, is a Banach space for the sum-norm $\|(x_Q, q)\| = \|\mathrm{L}\{x_Q\}\|_{\ell_p} + \|q\|_{\mathcal{N}_\mathrm{L}}$ with $x_Q \in \ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L}$ and $q \in \mathcal{N}_\mathrm{L}$. The final step is to specify the isomorphism between $\ell_{p,\mathrm{L}}(\mathbb{Z})$ and $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L} + \mathcal{N}_\mathrm{L}$ via the relation $x \mapsto (x_Q, \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{x\})$ where x_Q is the equivalence class of x in $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_\mathrm{L}$ and $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{x\}$ the projection of x into $\mathcal{N}_\mathrm{L}$. To make the link completely explicit, we further identify x_Q with $x - \mathrm{Proj}_{\mathcal{N}_\mathrm{L}}\{x\}$ which is the unique element of x_Q whose projection onto $\mathcal{N}_\mathrm{L}$ is zero. The reverse map is then simply $(x_Q, q) \mapsto x_Q + q$, which spans the complete space $\ell_{p,\mathrm{L}}(\mathbb{Z})$. As a consequence, $\ell_{p,\mathrm{L}}(\mathbb{Z})$ inherits the Banach space structure of the direct sum. ■

Let us note that Proposition 14 is a high-level statement that holds for any admissible norm $\|\cdot\|_{\mathcal{N}_\mathrm{L}}$ and projection operator $\mathrm{Proj}_{\mathcal{N}_\mathrm{L}}$. It turns out that the exact choice of these elements has no influence on the Banach topology of $\ell_{p,\mathrm{L}}(\mathbb{Z})$. The explanation lies in the fact that the null space $\mathcal{N}_\mathrm{L}$ is finite-dimensional and that all finite-dimensional norms are topologically equivalent. The finite-dimensionality of $\mathcal{N}_{\mathrm{L}}$ also guarantees the existence of the projector $\mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}: \ell_{p,\mathrm{L}}(\mathbb{Z}) \to \mathcal{N}_{\mathrm{L}}$ (by the Hahn-Banach theorem); the main point is that the latter should be seen as an extension of the identity map $i: \mathcal{N}_{\mathrm{L}} \to \mathcal{N}_{\mathrm{L}}$ to the whole space $\ell_{p,\mathrm{L}}(\mathbb{Z})$. In the sequel, we will fix these elements in order to properly invert the operator L. This will be achieved by imposing N_0 linear boundary conditions, as will be made explicit in Theorem 16.

Definition 15 (Admissible Regularization Operator): A linear operator $\mathrm{L}: \ell_{p,\mathrm{L}}(\mathbb{Z}) \to \ell_p(\mathbb{Z})$ is called admissible if

1) it has a finite-dimensional null space $\mathcal{N}_{\mathrm{L}} = \{q \in \ell_{p,\mathrm{L}}(\mathbb{Z}) : \mathrm{L}\{q\} = 0\}$ spanned by some basis $\boldsymbol{p} = (p_1, \dots, p_{N_0})$;
2) it is right-invertible in the sense that there exists a kernel $\rho_{\mathrm{L}} \in \mathcal{S}'(\mathbb{Z} \times \mathbb{Z})$ (the space of bi-infinite matrices with slow-growing rows and columns) with the property that $\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} = \delta[\cdot - l]$.

It is important to note that the fundamental solution of $\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} = \delta[\cdot - l]$ (or discrete Green's function) is not unique (unless $N_0 = 0$) since any kernel of the form $q_l + \rho_{\mathrm{L}}[\cdot, l]$ with $q_l \in \mathcal{N}_{\mathrm{L}}$ is acceptable as well. We shall now show that there are some privileged forms that result in an inversion that is stable over $\ell_2(\mathbb{Z})$.

Theorem 16 (Stable Right-Inverse Operator): Let L be an admissible regularization operator in the sense of Definition 15. We also assume that we are given some corresponding set of biorthogonal analysis functionals $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{N_0})$ with $\phi_n \in \ell'_{2,\mathrm{L}}(\mathbb{Z})$ (the continuous dual of $\ell_{2,\mathrm{L}}(\mathbb{Z})$) such that $\langle \phi_m, p_n \rangle = \delta[m - n]$. Then,

$$\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \left\{x : \mathbb{Z} \to \mathbb{R} \;\middle|\; \mathrm{L}\{x\} \in \ell_2(\mathbb{Z}) \text{ and } \langle \boldsymbol{\phi}, x \rangle = \mathbf{0}\right\}$$

is a Hilbert space equipped with the inner product $\langle f, g \rangle_{\mathrm{L}} = \langle \mathrm{L}\{f\}, \mathrm{L}\{g\} \rangle$. Moreover, there exists an isometric map $\mathrm{L}_{\boldsymbol{\phi}}^{-1}: \ell_2(\mathbb{Z}) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ such that

$$\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} = \left\{x = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w : w \in \ell_2(\mathbb{Z})\right\}.$$

The operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is uniquely specified through the following properties

1) Right-inverse property: $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$ for all $w \in \ell_2(\mathbb{Z})$
2) Boundary conditions: $\langle \boldsymbol{\phi}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w \rangle = \mathbf{0}$ for all $w \in \ell_2(\mathbb{Z})$

and its kernel is given by

$$g_{\boldsymbol{\phi}}[k, l] = \rho_{\mathrm{L}}[k, l] - \sum_{n=1}^{N_0} p_n[k] q_n[l], \tag{30}$$

with $q_n[l] = \langle \rho_{\mathrm{L}}[\cdot, l], \phi_n \rangle$ and ρ_{L} such that $\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} = \delta[\cdot - l]$.

Proof: We start by proving that $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ equipped with the inner product

$$\langle x_1, x_2 \rangle_{\mathrm{L}} = \langle \mathrm{L}\{x_1\}, \mathrm{L}\{x_2\} \rangle = \langle \mathrm{L}^*\mathrm{L}\{x_1\}, x_2 \rangle \tag{31}$$

is a Hilbert space. The only delicate aspect there is to establish the unicity property of the inner product: $\langle x_0, x_0 \rangle_{\mathrm{L}} = 0 \Leftrightarrow x_0 = 0$. To that end, we observe that the condition $\langle \mathrm{L}\{x_0\}, \mathrm{L}\{x_0\} \rangle = 0$ is equivalent to $x_0 \in \mathcal{N}_{\mathrm{L}}$. Thanks to the biorthogonality of $\boldsymbol{p}$ and $\boldsymbol{\phi}$, we also know that

$$q = \sum_{n=1}^{N_0} \langle \phi_n, q \rangle p_n$$

for all $q \in \mathcal{N}_{\mathrm{L}}$. Finally, we use the boundary conditions $\langle \boldsymbol{\phi}, x_0 \rangle = \mathbf{0}$ to conclude that $x_0 = 0$.

The idea is then to first establish the properties 1) and 2) of the operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ on the space of rapidly-decreasing sequences $\mathcal{S}(\mathbb{Z})$ to avoid any technical problems related to the splitting and interchange of sums. Since the space $\mathcal{S}(\mathbb{Z})$ equipped with the standard weighted-ℓ_2 Fréchet topology is dense in $\ell_2(\mathbb{Z})$ [36], we are then able to extend the properties by continuity.

For notational purpose, we introduce the operator $\mathrm{G}: x \mapsto \sum_{l\in\mathbb{Z}} \rho_{\mathrm{L}}[\cdot, l] x[l]$, which is well defined over $\mathcal{S}(\mathbb{Z})$ as long as $\rho_{\mathrm{L}}[\cdot, \cdot] \in \mathcal{S}'(\mathbb{Z} \times \mathbb{Z})$. By assuming that $w \in \mathcal{S}(\mathbb{Z})$, we can therefore rewrite $x = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\}$ as

$$x = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} = \mathrm{G}\{w\} - \sum_{n=1}^{N_0} p_n \langle q_n, w \rangle.$$

Next, we apply the operator L, which yields

$$\begin{aligned}\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} &= \mathrm{L}\left\{\sum_{l\in\mathbb{Z}} w[l]\rho_{\mathrm{L}}[\cdot, l]\right\} - \sum_{n=1}^{N_0} \underbrace{\mathrm{L}\{p_n\}}_{=0} \langle q_n, w \rangle \\ &= \sum_{k\in\mathbb{Z}} w[l]\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} \\ &= \sum_{k\in\mathbb{Z}} w[l]\delta[\cdot - l] = w\end{aligned}$$

where we have used the defining properties $\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} = \delta[\cdot - l]$ and $\mathrm{L}\{p_n\} = 0$ for $n = 1, \cdots, N_0$. In particular, this implies that

$$\|\mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\}\|_{\mathrm{L}}^2 = \langle \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\}, \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} \rangle_{\mathrm{L}} = \|w\|_{\ell_2}^2 \tag{32}$$

for all $w \in \mathcal{S}(\mathbb{Z})$, which shows that $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is bounded in the ℓ_2 norm.

As for the boundary conditions, we first observe that

$$\begin{aligned}q_n[l] &= \langle \rho_{\mathrm{L}}[\cdot, l], \phi_n \rangle \\ &= \sum_{k\in\mathbb{Z}} \rho_{\mathrm{L}}[k, l]\phi_n[k] = \mathrm{G}^*\{\phi_n\}[l]\end{aligned}$$

where G^* is the adjoint of G. We then make use of the biorthogonality property $\langle \phi_m, p_n \rangle = \delta[m - n]$ to evaluate the inner product of $\mathrm{L}_{\boldsymbol{\phi}}^{-1} w$ with ϕ_m as

$$\begin{aligned}\langle \phi_m, \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w\} \rangle &= \langle \phi_m, \mathrm{G}\{w\} \rangle - \sum_{n=1}^{N_0} \langle \phi_m, p_n \rangle \langle q_n, w \rangle \\ &= \langle \phi_m, \mathrm{G}\{w\} \rangle - \langle q_m, w \rangle \\ &= \langle \mathrm{G}^*\{\phi_m\}, w \rangle - \langle \mathrm{G}^*\{\phi_m\}, w \rangle = 0,\end{aligned}$$

which shows that the boundary conditions are satisfied. In doing so, we have effectively shown that $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ continuously maps $\mathcal{S}(\mathbb{Z})$ into $\mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$. Again, since $\mathcal{S}(\mathbb{Z})$ is dense in $\ell_2(\mathbb{Z})$, the boundary conditions do also extend to $\ell_2(\mathbb{Z})$ by continuity.

As final step, we invoke the Hahn-Banach theorem in conjunction with the ℓ_2 bound (32) to extend the domain of the operator to all of $\ell_2(\mathbb{Z})$. This allows us to conclude that $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ continuously maps $\ell_2(\mathbb{Z}) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$. In fact, $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \ell_2(\mathbb{Z}) \to \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}}$ is an isometry that provides a stable inverse of the operator $\mathrm{L} : \mathcal{H}_{\mathrm{L},\boldsymbol{\phi}} \to \ell_2(\mathbb{Z})$. In other words, we have shown that the operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ whose kernel is specified by (30) is such that $\mathrm{L}\mathrm{L}_{\boldsymbol{\phi}}^{-1} w = w$, $\langle \boldsymbol{\phi}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w\rangle = \mathbf{0}$ and $\|\mathrm{L}_{\boldsymbol{\phi}}^{-1} w\|_{\mathrm{L}} = \|w\|_{\ell_2}$ for all $w \in \ell_2(\mathbb{Z})$. ■

Since $\ell_p(\mathbb{Z}) \subseteq \ell_2(\mathbb{Z})$ for $p \in [1, 2]$, one can obviously also restrict the domain of the inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ to $\ell_p(\mathbb{Z})$ with the insurance that Properties 1) and 2) are met for $w \in \ell_p(\mathbb{Z})$.

As demonstration of usage, let us consider the finite-difference operator D, which is specified as

$$x \mapsto \mathrm{D}\{x\} = x[\cdot] - x[\cdot - 1].$$

This operator is the infinite-dimensional counterpart of $\mathbf{D}$ in Section III-D. It is shift-invariant, and its Fourier symbol is $(1 - \mathrm{e}^{-\mathrm{j}\omega})$, which exhibits a single zero at $\omega = 0$. Consequently, D has a one-dimensional null space $\mathcal{N}_{\mathrm{D}} = \mathrm{span}\{p_1\} \subseteq \ell_\infty(\mathbb{Z})$ that is spanned by the "constant" signal $p_1[k] = 1$. The simplest choice of biorthogonal analysis vector is $\phi_1 = \delta[\cdot]$ with the property that

$$q = \langle \delta[\cdot], q\rangle p_1 = q[0] p_1$$

for all $q \in \mathcal{N}_{\mathrm{D}}$. A possible choice of fundamental solution is $\rho_{\mathrm{D}}[k, l] = \mathbb{1}_+[k - l]$ with the property that $\mathrm{D}\{\mathbb{1}_+[\cdot - l]\} = \delta[\cdot - l]$. The application of Theorem 16 then yields the kernel of the corresponding right-inverse operator $\mathrm{D}_{\phi_1}^{-1}$:

$$g_\delta[k, l] = \mathbb{1}_+[k - l] - \mathbb{1}_+[-l].$$

Its stability is revealed by observing that, for $k_0 \geq 1$, $g_\delta[k_0, l] = \mathbb{1}_{\{1,\dots,k_0\}}[l]$, which is compactly supported of size k_0, and hence included in $\ell_{p'}(\mathbb{Z})$ for all $p' \geq 1$. This guarantees that $x \mapsto \mathrm{D}_{\phi_1}^{-1}\{x\}$ is well defined for any $x \in \ell_p(\mathbb{Z})$ with $p \geq 1$ (see Proposition 13).

This is in contrast with the "canonical" shift-invariant inversion mechanism $x \mapsto y = \mathbb{1}_+ * x$ (moving sum filter), which is ill-defined on $\ell_p(\mathbb{Z})$ for $p > 1$.

The main point that we want to make here is that the inversion task is not trivial (because the standard system-theoretic solution is not directly applicable), but that it can nevertheless be achieved in a principled fashion by applying the constructive procedure described in Theorem 16. In essence, the second term in (30) is a mathematical correction that makes the (right)-inverse operator ℓ_p-stable for $1 \leq p \leq 2$.

D. Extended Regularization Theory

We have now all the tools in hand to make the Banach structure of $\ell_{p,\mathrm{L}}(\mathbb{Z})$ suggested by Proposition 14 explicit. This, in turn, will allow us to derive the generic form of the optimizer for $p = 1, 2$.

Theorem 17 (Direct Sum Decomposition): Let L be a regularization operator that admits a stable right-inverse $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ of the form specified by Theorem 16. Then, any $x \in \ell_{p,\mathrm{L}}(\mathbb{Z})$ with $p \in [1, 2]$ has a unique representation as

$$x = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + q,$$

where $w = \mathrm{L}\{x\} \in \ell_p(\mathbb{Z})$ and $q = \sum_{n=1}^{N_0} \langle \phi_n, x\rangle p_n \in \mathcal{N}_{\mathrm{L}}$. Moreover, $\ell_{p,\mathrm{L}}(\mathbb{Z})$ is a Banach space equipped with the norm

$$\|x\|_{p,\mathrm{L},\boldsymbol{\phi}} = \|\mathrm{L}\{x\}\|_{\ell_p} + \|\langle x, \boldsymbol{\phi}\rangle\|_2. \quad (33)$$

Proof: The right-inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ is obviously well-defined for $w \in \ell_p(\mathbb{R}) \subseteq \ell_2(\mathbb{Z})$. Let $x_1, x_2 \in \ell_{p,\mathrm{L}}(\mathbb{Z})$ be such that $\mathrm{L}\{x_1\} = \mathrm{L}\{x_2\} = w$. By definition of the null space, this is equivalent to $(x_1 - x_2) = q \in \mathcal{N}_{\mathrm{L}}$. Conversely, let $x = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w$ with $w \in \ell_p(\mathbb{Z})$. Then, the condition $\|\mathrm{L}\{x\}\|_p = \|w\|_p < \infty$ ensures that $x \in \ell_{p,\mathrm{L}}(\mathbb{Z})$. This allows us to deduce that $\ell_{p,\mathrm{L}}(\mathbb{Z})$ is the sum of $\mathcal{N}_{\mathrm{L}} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$ and

$$\begin{aligned}\mathcal{B}_{p,\mathrm{L},\boldsymbol{\phi}} &= \{x = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w : w \in \ell_p(\mathbb{Z})\}\\ &= \{x \in \ell_{p,\mathrm{L}}(\mathbb{Z}) : \langle \phi_n, x\rangle = 0, n = 1, \dots, N_0\},\end{aligned}$$

where $\mathcal{B}_{p,\mathrm{L},\boldsymbol{\phi}}$ is a Banach space equipped with the norm $\|\mathrm{L}\{\cdot\}\|_{\ell_p}$. Its completeness is inherited from the one of $\ell_p(\mathbb{Z})$ and the fact that the inverse operator $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ performs an isometric mapping $\ell_p(\mathbb{Z}) \to \mathcal{B}_{p,\mathrm{L},\boldsymbol{\phi}}$. Moreover, since $\langle \boldsymbol{\phi}, \mathrm{L}_{\boldsymbol{\phi}}^{-1} w\rangle = \mathbf{0}$ (boundary conditions) and the ϕ_m are biorthogonal to the p_n, we find that the null-space component q is given by

$$q = \sum_{n=1}^{N_0} \langle \phi_n, x\rangle p_n = \mathrm{Proj}_{\mathcal{N}_{\mathrm{L}}}\{x\}$$

It is therefore specified by its expansion coefficients $\langle \boldsymbol{\phi}, x\rangle = (\langle \phi_1, x\rangle, \dots, \langle \phi_{N_0}, x\rangle)$ whose ℓ_2-norm is $\|\langle \boldsymbol{\phi}, x\rangle\|_2$. This shows that the decomposition $x = \mathrm{L}_{\boldsymbol{\phi}}^{-1} w + q$, where $w = \mathrm{L}\{x\} \in \ell_p(\mathbb{Z})$ and $q \in \mathcal{N}_{\mathrm{L}}$, is unique, which also translates into $\ell_{p,\mathrm{L}}(\mathbb{Z}) = \mathcal{B}_{p,\mathrm{L},\boldsymbol{\phi}} \oplus \mathcal{N}_{\mathrm{L}}$ because $\ell_{p,\mathrm{L},\boldsymbol{\phi}}(\mathbb{Z}) \cap \mathcal{N}_{\mathrm{L}} = \{0\}$. The final part of the argument is the same as in Proposition 14 with $\mathcal{B}_{p,\mathrm{L},\boldsymbol{\phi}}$ being isomorphically equivalent to the quotient space $\ell_{p,\mathrm{L}}(\mathbb{Z})/\mathcal{N}_{\mathrm{L}}$. ■

Using Theorem 17, we now proceed to provide the results for convex optimization with ℓ_1 and ℓ_2 regularizers. The technical part concerning the weak*-compactness of the solution set is taken care in Appendix B.

Theorem 18 (Convex Problem With ℓ_2 Regularization): Let us consider the following:

- $\mathrm{L} : \ell_{2,\mathrm{L}}(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ is an admissible regularization operator in the sense of Definition 15;
- $\mathrm{H} : \ell_{2,\mathrm{L}}(\mathbb{Z}) \to \mathbb{R}^M : x \mapsto (\langle h_1, x\rangle, \dots, \langle h_M, x\rangle)$ is a linear measurement operator such that, for any $x \in \ell_{2,\mathrm{L}}(\mathbb{Z})$,

$$A\|\langle x, \boldsymbol{\phi}\rangle\|_2 \leq \|\mathrm{H}\{x\}\|_2 \leq B(\|\mathrm{L}\{x\}\|_{\ell_2} + \|\langle x, \boldsymbol{\phi}\rangle\|_2) \quad (34)$$

  for some constants $A, B > 0$ and $\boldsymbol{\phi}$ as in Theorem 16;
- $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^M$ such that its preimage in $\ell_2(\mathbb{Z})$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C}) = \{x \in \ell_{2,\mathrm{L}}(\mathbb{Z}) : \mathrm{H}\{x\} \in \mathcal{C}\}$, is nonempty (feasibility hypothesis);
- $\mathrm{L}_{\boldsymbol{\phi}}^{-1} : \ell_2(\mathbb{Z}) \to \ell_{2,\mathrm{L}}(\mathbb{Z})$ is a stable right-inverse of L as specified in Theorem 16.

Then,

$$\mathcal{V} = \arg \min_{x \in \ell_{2,\mathrm{L}}(\mathbb{Z})} \|\mathrm{L}\{x\}\|_{\ell_2} \quad \text{s.t.} \quad \mathrm{H}\{x\} \in \mathcal{C}$$

is a nonempty, convex, weak*-compact subset of $\ell_{2,\mathrm{L}}(\mathbb{Z})$ with solutions of the form

$$x_{\mathrm{LS}} = \sum_{m=1}^{M} a_m \tilde{h}_m + p_0 \quad \text{with} \quad \tilde{h}_m = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\{h_m\}, \tag{35}$$

where $\boldsymbol{a} = (a_1, \dots, a_M)$ is a fixed element of $\mathbb{R}^M$ and $p_0 \in \mathcal{N}_{\mathrm{L}}$ a null-space component that describes the full solution set; i.e., $p_0 \in \mathrm{H}^{-1}\left(\mathcal{C} - \mathrm{H}\left\{\sum_{m=1}^{M} a_m \tilde{h}_m\right\}\right) \cap \mathcal{N}_{\mathrm{L}}$. In particular, when $\mathcal{C}$ reduces to a single point, then the solution is unique.

Proof: The property that $\mathcal{V}$ is non-empty, convex and weak*-compact is covered by Lemma 20 in Appendix B. Consider the set $\mathcal{C}_{\mathrm{L}} = \{z + \mathrm{H}\{p\},\ z \in \mathcal{C},\ p \in \mathcal{N}_{\mathrm{L}}\}$. We define the new optimization problem

$$\mathcal{W} = \arg \min_{w \in \ell_2(\mathbb{Z})} \|w\|_2 \text{ s.t. } (\mathrm{HL}_{\boldsymbol{\phi}}^{-1})\{x\} \in \mathcal{C}_{\mathrm{L}}. \tag{36}$$

$\mathcal{C}_{\mathrm{L}}$ is closed and convex as the sum of two closed and convex sets, $\mathcal{C}$ and $\mathrm{H}(\mathcal{N}_{\mathrm{L}})$. Moreover, we easily show that

$$\mathcal{U}_0 := (\mathrm{HL}_{\boldsymbol{\phi}}^{-1})^{-1}(\mathcal{C}_{\mathrm{L}}) = \mathrm{L}(\mathrm{H}^{-1}(\mathcal{C})). \tag{37}$$

Since the set $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$ is nonempty by assumption, the same holds true for $\mathcal{U}_0 = \mathrm{L}(\mathcal{U})$. We are therefore fulfilling the conditions of Theorem 9, from which we deduce that there exists a unique minimizer $w_{\mathrm{LS}} = (\mathrm{HL}_{\boldsymbol{\phi}}^{-1})^*\{\boldsymbol{a}\}$ in $\mathcal{W}$, with $\boldsymbol{a} \in \mathbb{R}^M$.

Let $x_0 \in \mathcal{V}$, which is decomposed as $x_0 = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_0\} + p_0$ with $w_0 \in \ell_2(\mathbb{Z})$ and $p_0 \in \mathcal{N}_{\mathrm{L}}$. Then, $\mathrm{L}\{x_0\} = w_0$ and $x_0 \in \mathcal{U}$, hence $w_0 \in \mathcal{U}_0$. Likewise, for any $w \in \mathcal{U}_0$, there exists $x \in \mathcal{U}$ such that $\mathrm{L}\{x\} = w$. Since $x_0 \in \mathcal{V}$ and $x \in \mathcal{U}$, we have $\|w\|_{\ell_2} = \|\mathrm{L}\{x\}\|_{\ell_2} \geq \|\mathrm{L}\{x_0\}\|_{\ell_2} = \|w_0\|_{\ell_2}$. As this relation is true for every $w \in \mathcal{U}_0$, $w_0 \in \mathcal{W}$ and therefore $w_0 = w_{\mathrm{LS}}$. This shows that $x_0 = \mathrm{L}_{\boldsymbol{\phi}}^{-1}(\mathrm{HL}_{\boldsymbol{\phi}}^{-1})^*\{\boldsymbol{a}\} + p_0$. Next, we define $x_1 = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_0\}$ and simplify its expression as

$$\begin{aligned} x_1 &= \mathrm{L}_{\boldsymbol{\phi}}^{-1}(\mathrm{HL}_{\boldsymbol{\phi}}^{-1})^*\{\boldsymbol{a}\} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\mathrm{H}^*\{\boldsymbol{a}\} \\ &= \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\left\{\sum_{m=1}^{M} a_m h_m\right\} = \sum_{m=1}^{M} a_m \tilde{h}_m \end{aligned}$$

where $\tilde{h}_m = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\mathrm{L}_{\boldsymbol{\phi}}^{-1*}\{h_m\}$. Since $\mathrm{H}\{x_0\} = \mathrm{H}\{x_1\} + \mathrm{H}\{p_0\} \in \mathcal{C}$ by definition, we deduce that p_0 necessarily lies in $\mathrm{H}^{-1}(\mathcal{C} - \mathrm{H}\{x_1\})$. Conversely, any element of the form (35) is clearly in $\mathcal{U}$, and hence in $\mathcal{V}$ when $\boldsymbol{a}$ is chosen optimally. ■

Theorem 19 (Convex Problem With ℓ_1 Regularization): Let us consider the following:

- $\mathrm{L} : \ell_{1,\mathrm{L}}(\mathbb{Z}) \to \ell_1(\mathbb{Z})$ is an admissible regularization operator in the sense of Definition 15;
- $\rho_{\mathrm{L}}[\cdot,\cdot]$ is a kernel such that $\mathrm{L}\{\rho_{\mathrm{L}}[\cdot, l]\} = \delta[\cdot - l]$ for all $l \in \mathbb{Z}$.
- $\mathrm{H} : \ell_{1,\mathrm{L}}(\mathbb{Z}) \to \mathbb{R}^M$ is a linear measurement operator such that, for any $x \in \ell_{1,\mathrm{L}}(\mathbb{Z})$,

$$A\|\langle x, \boldsymbol{\phi}\rangle\|_2 \leq \|\mathrm{H}\{x\}\|_2 \leq B(\|\mathrm{L}\{x\}\|_{\ell_1} + \|\langle x, \boldsymbol{\phi}\rangle\|_2) \tag{38}$$

for some constants $A, B > 0$ and $\boldsymbol{\phi}$ as in Theorem 16.

- $\mathcal{C}$ is a convex compact subset of $\mathbb{R}^M$ such that its preimage in $\ell_{1,\mathrm{L}}(\mathbb{Z})$, $\mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$, is nonempty (feasibility hypothesis).

Then,

$$\mathcal{V} = \arg \min_{x \in \ell_{1,\mathrm{L}}(\mathbb{Z})} \|\mathrm{L}\{x\}\|_{\ell_1} \quad \text{s.t. } \mathrm{H}\{x\} \in \mathcal{C}$$

is a nonempty, convex, weak*-compact subset of $\ell_{1,\mathrm{L}}(\mathbb{Z})$ with extreme points of the form

$$x_{\mathrm{sparse}} = \sum_{k=1}^{K} a_k \rho_{\mathrm{L}}[\cdot, n_k] + \sum_{n=1}^{N_0} b_n p_n \tag{39}$$

with $K \leq M$, $n_k \in \mathbb{Z}$, $a_k, b_n \in \mathbb{R}$, and $\|\mathrm{L}\{x_{\mathrm{sparse}}\}\|_{\ell_1} = \sum_{k=1}^{K} |a_k|$.

Proof: Here too, we refer to Lemma 20 with $p = 1$ for the non-emptiness, convexity, and weak*-compactness of $\mathcal{V}$. The remainder of the proof is essentially the same as the one of Theorem 6. For a fixed extreme point x, we assume that $\mathrm{L}\{x\}$ is not K-sparse and that we can find at least $M + 1$ elements $n_1, \dots, n_{M+1}$ such that $\mathrm{L}\{x\}[n_k] \neq 0$ and we show that x is not an extreme point. The final observation is that $x_{\mathrm{sparse}} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_\infty\} + p_\infty$ can be rewritten as (39) by using the explicit form of the kernel of $\mathrm{L}_{\boldsymbol{\phi}}^{-1}$ given by (30). ■

Once again, it is instructive to compare the solutions of the ℓ_2 and ℓ_1 regularization problems covered by Theorems 18 and 19. The first fundamental difference is that the solution of the ℓ_2 problem is constrained to live in a *fixed* finite-dimensional subspace of ℓ_2, while the reconstruction space for the ℓ_1 problem is *adaptive* and determined by the problem and the data at hand. Interestingly, the first property remains valid for the ℓ_2 regularization even if the solution of the extended problem in Theorem 18 is no longer unique because of the additional degrees of freedom offered by the null space component. The second distinction is in the form of the basis functions: In the ℓ_2 case, there is a characteristic intertwining between the effect of the measurement and regularization operators, while in the ℓ_1 scenario the basis functions are chosen within a dictionary $\{\rho_{\mathrm{L}}[\cdot, n]\}_{k \in \mathbb{Z}}$ whose form is completely determined by the regularization operator L. This part of the story is completely in line with the findings of Section III so that all the comments that have been made there are still pertinent.

The novel aspect in our two last representer theorems is the appearance of the second parametric term $p_0 = \sum_{n=1}^{N_0} b_n p_n$, which encodes the component that is in the null space of the operator. As already mentioned, the role of p_0, whose regularization cost is zero, is fundamental because it tries to fullfil the constraints as much as possible in order to decrease the ℓ_1 or ℓ_2 penalty associated with the first component. While the possibility of applying a regularization operator whose null space is non-trivial is immensely useful in practice, it requires a more sophisticated mathematical treatment. The enabling ingredient is the construction and proof of existence of a stable right-inverse operator under very weak hypotheses (Theorem 16) which also constitutes one of the contribution of this work.

We believe that the stability bounds used in the statement of our infinite-dimensional representer theorems are the weakest possible hypotheses for this kind of optimization problem. The upper bound on $\|\mathrm{H}\{x\}\|_2$ is the explicit way of indicating that the measurement operator is well-defined in the sense that it continuously maps $\ell_{p,\mathrm{L}}(\mathbb{Z}) \rightarrow \mathbb{R}^M$; as far as we know, this latter hypothesis (which is often implicit) is necessary for the mathematical analysis of any inverse problem. Hence, the only constraining hypothesis is the lower bound in (34) and (38), which is required to counteract the lack of coercivity of the regularization functional $\|\mathrm{L}\{x\}\|_p$. It makes the problem well-posed over the (very small) subspace $\mathcal{N}_\mathrm{L}$; in other words, the measurements should be rich enough to allow us to unambiguously reconstruct the null-space component of the signal. For instance, in the case of TV (i.e., $\mathrm{L} = \mathrm{D}$), there should at least be one measurement functional h_m such that $\langle h_m, 1\rangle \neq 0$, which is a very mild constraint. Also note that the non-coercive scenario has the additional restriction that the convex set $\mathcal{C}$ should be bounded.

E. Connection With Splines

In Section III-D, we have seen that the extremal points of finite-dimensional linear inverse problems with a total-variation regularization are necessarily piecewise-constant, which suggests a connection with splines. We recall that splines are continuous-domain entities (i.e., functions) that are defined classically as the solution of a quadratic-energy minimization problem subject to (linear) interpolation constraints [37], [38]. The concept is transposable to the discrete domain as well, which leads to the related notion of *discrete splines* with the regularization operator $\mathrm{L} = \mathrm{D}^n$ being the nth power of the finite-difference operator D. Existence results are also available for discrete splines with ℓ_p regularization for $p \geq 1$ [39], but the explicit form of these splines has only been worked out explicitly for $p = 2$. This corresponds to the simplified setting $\mathrm{H}\{x\} = (x[k_1], \dots, x[k_M])$ (non-uniform sampling operator) and $\mathcal{C} = \mathbf{y} = (y_1, \dots, y_M) \in \mathbb{R}^M$ in Theorem 18, which imposes the interpolation constraints $x[k_1] = y_1, \dots, x[k_M] = y_M$. It is well known that this problem admits a unique solution, which is the discrete counterpart of a polynomial spline interpolant of degree $2n-1$ with knots at the k_m's [39], [40].

In order to specify the solution of the ℓ_1 variant of the interpolation problem, we observe that D^n admits a discrete shift-invariant Green's function $\rho_n[\cdot]$ that is the n-fold convolution of the discrete step $\mathbb{1}_+$ and hence a (discrete) one-sided polynomial of degree $n-1$. The corresponding form of the extreme points in Theorem 19 is $\sum_{k=1}^{K} a_k \rho_n[\cdot - n_k] + p_0$ where the null-space component p_0 is a (discrete) polynomial of degree $n-1$. In other words, they are discrete splines of degree $n-1$ with data-dependent knots $(n_k)_{k=1}^K$ and $K \leq M$. Besides the reduction of the polynomial degree of the spline, the key difference with the ℓ_2 scenario is that the position of the knots is adaptive and not known a priori. Yet, the truly remarkable finding here is that this functional form of the solution remains valid for *any* convex linear inverse problems with nth-order ℓ_1-regularization, far beyond the classical spline setting.

Finally, we have recently managed to (literally) connect the dots (that is, the samples of the signal) by developing a functional framework that is the continuous-domain counterpart of the present theory; this is the topic of a forthcoming paper whose name says it all [41].

V. Conclusion

In this paper, we have characterized the form of the solution of general linear inverse problems with convex constraints and ℓ_1 vs. ℓ_2 regularization. We have started from the simplest finite-dimensional scenario and worked our way up progressively to the more challenging family of (infinite-dimensional) inverse problems covered by the Representer Theorems 18 and 19. We have striven for the maximal generality and the weakest possible assumptions in order to cover the majority of convex signal recovery problems encountered in practice. We believe that these functional descriptions of the solution should be of interest to researchers working in the field.

The primary message that emerges from this investigation is the superiority of ℓ_1 over ℓ_2 regularization for injecting prior knowledge on the solution. For instance, the minimization of $\|\mathrm{D}^2\{x\}\|_{\ell_1}$ where D^2 is the 2nd-order difference operator produces solutions that are piecewise-linear irrespective of the system's matrix H and the number of measurements. There is no such independence between the characteristic form of the solution and the system matrix in the case of ℓ_2 regularization.

Appendix A
Proof of Proposition 1

Let $\mathbf{H}$ be an arbitrary matrix of size $M \times N$ and $\mathbf{I}_N$ the identity matrix of size N. We start by noting that

$$(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_N)\mathbf{H}^T = \mathbf{H}^T\mathbf{H}\mathbf{H}^T + \lambda\mathbf{H}^T = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_M)$$

The underlying hypothesis that $\lambda > 0$ ensures that both $(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_N)$ and $(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_M)$ are invertible. This allows us to deduce that

$$(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I}_N)^{-1}\mathbf{H}^T = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I}_M)^{-1},$$

which is the desired result.

Appendix B
Convexity and Weak*-Compactness of Solution Set

Here, we establish the convexity and weak*-compactness of the sets of minimizers for the infinite-dimensional optimization problems of Section IV. This result is preparatory for the proof of all representer theorems.

When the operator L is invertible (including the simplest case of the identity), the functional that we minimize is coercive, convex, and lower semi-continuous. In that case, Lemma 20 below can be deduced from standard results in convex optimization [42, Sec. II-1]. However, when the operator L has a non-trivial null space, the functional $\|\mathrm{L}\{x\}\|_p^p$ is not coercive anymore and the proof must be adapted. This is the main contribution of Lemma 20.

Lemma 20: For $1 \leq p \leq 2$ fixed, let us consider the following:

- $\mathrm{L} : \ell_{p,\mathrm{L}}(\mathbb{Z}) \to \ell_p(\mathbb{Z})$ is an admissible regularization operator in the sense of Definition 15;
- $\mathrm{H} : \ell_{p,\mathrm{L}}(\mathbb{Z}) \to \mathbb{R}^M$ is a linear measurement operator such that, for any $x \in \ell_{p,\mathrm{L}}(\mathbb{Z})$,

$$A\|\langle x, \boldsymbol{\phi}\rangle\|_2 \leq \|\mathrm{H}\{x\}\|_2 \leq B(\|\mathrm{L}\{x\}\|_{\ell_p} + \|\langle x, \boldsymbol{\phi}\rangle\|_2) \tag{40}$$

  for some constants $A, B > 0$ and $\boldsymbol{\phi}$ as in Theorem 16;
- $\mathcal{C}$ is a compact convex subset of $\mathbb{R}^M$ such that its preimage in $\ell_p(\mathbb{Z}), \mathcal{U} = \mathrm{H}^{-1}(\mathcal{C})$, is nonempty (feasibility hypothesis).

Then,

$$\mathcal{V} = \arg \min_{x \in \ell_{p,\mathrm{L}}(\mathbb{Z})} \|\mathrm{L}\{x\}\|_{\ell_p} \quad \text{s.t. } \mathrm{H}\{x\} \in \mathcal{C}$$

is a nonempty, convex, weak*-compact subset of $\ell_{p,\mathrm{L}}(\mathbb{Z})$.

When the operator L is a bijection the conclusion remains valid for any closed convex, but not necessarily bounded, set $\mathcal{C}$.

Proof: The measurement operator H is linear by assumption and bounded on $\ell_{p,\mathrm{L}}(\mathbb{Z})$ due to (38); therefore, it is continuous. The set $\mathcal{U}$ is closed convex as the preimage of a closed convex set by the linear and continuous map H.

Next, we show that $\mathcal{V}$ is nonempty. Let (x_n) be a sequence of elements of $\mathcal{U}$ such that $\|\mathrm{L}\{x_n\}\|_{\ell_p}$ decreases to $\beta = \inf_{x \in \mathcal{U}} \|\mathrm{L}\{x\}\|_{\ell_p}$. Based on Theorem 17, we decompose $x_n = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_n\} + p_n$ in a unique way with $w_n \in \ell_p(\mathbb{Z})$ and $p_n \in \mathcal{N}_{\mathrm{L}}$. Then, $\|w_n\|_{\ell_p} = \|\mathrm{L}\{x_n\}\|_{\ell_p}$ is bounded. Moreover, thanks to the lower bound in (40), we have

$$\begin{aligned}\|\langle p_n, \boldsymbol{\phi}\rangle\|_2 &\leq \frac{1}{A}\|\mathrm{H}\{p_n\}\|_2 = \frac{1}{A}\|\mathrm{H}\{x_n\} - \mathrm{HL}_{\boldsymbol{\phi}}^{-1}\{w_n\}\|_2 \\ &\leq \frac{1}{A}\left(\|\mathrm{H}\{x_n\}\|_2 + \|\mathrm{HL}_{\boldsymbol{\phi}}^{-1}\{w_n\}\|_2\right).\end{aligned} \tag{41}$$

The $\mathrm{H}\{x_n\}$ are inside the bounded set $\mathcal{C}$ so that $\|\mathrm{H}\{x_n\}\|_2$ is bounded as well. Moreover, the composed operator $\mathrm{HL}_{\boldsymbol{\phi}}^{-1}$ is continuous from $\ell_p(\mathbb{Z})$ to $\mathbb{R}^M$ and (w_n) is bounded in $\ell_p(\mathbb{Z})$, so that $\|\mathrm{HL}_{\boldsymbol{\phi}}^{-1}\{w_n\}\|_2$ is bounded too. This shows that $\|\langle p_n, \boldsymbol{\phi}\rangle\|_2$ is bounded. The space $\mathcal{N}_{\mathrm{L}}$ being finite-dimensional, we can therefore extract a subsequence of (p_n) that converges to $p_\infty \in \mathcal{N}_{\mathrm{L}}$. Since the sequence (w_n) is bounded in $\ell_p(\mathbb{Z})$, we also extract a subsequence that converges to $w_\infty \in \ell_p(\mathbb{Z})$ for the weak*-topology. Finally, a double extraction allows us to consider $x_{\varphi(n)} = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_{\varphi(n)}\} + p_{\varphi(n)}$ that converges to $x_\infty = \mathrm{L}_{\boldsymbol{\phi}}^{-1}\{w_\infty\} + p_\infty$ for the weak*-topology on $\ell_{p,\mathrm{L}}(\mathbb{Z})$. Then, the space $\mathcal{U}$ is closed and therefore weak*-closed; hence, $x_\infty \in \mathcal{U}$ as a weak*-limit of elements in $\mathcal{U}$. Moreover, $\|\mathrm{L}\{x_\infty\}\|_{\ell_p} \leq \|\mathrm{L}\{x_{\varphi(n)}\}\|_{\ell_p} \to \beta$. Since $x_\infty \in \mathcal{U}$, we also have $\|\mathrm{L}\{x_\infty\}\|_{\ell_p} \geq \beta$ and therefore $x_\infty \in \mathcal{V}$, which is therefore nonempty.

Moreover, we can write $\mathcal{V} = \mathcal{U} \cap \mathcal{B}$ with $\mathcal{B} = \{x \in \ell_{p,\mathrm{L}}(\mathbb{Z}),\ \|\mathrm{L}\{x\}\|_{\ell_p} \leq \beta\}$. The space $\mathcal{B}$ is convex and weak*-compact in $\ell_{p,\mathrm{L}}(\mathbb{Z})$ due to the Banach-Alaoglu theorem. Therefore, $\mathcal{V}$ is itself convex and weak*-compact as the intersection of two convex sets, one being weak*-compact and the other weak*-closed.

Finally, when the null space of L is trivial, the bound (41) is not required, so that we do not need the compactness of $\mathcal{C}$. ∎

References

[1] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math.*, Dokl. 5, pp. 1035–1038, 1963.

[2] A. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA, USA: SIAM, 1996.

[3] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Boca Raton, FL, USA: CRC Press, 1998.

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

[5] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

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

[8] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[9] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[10] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[12] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[14] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[15] T. Goldstein and S. Osher, "The split Bregman method for l_1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[16] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, no. 2, pp. 167–188, 1997.

[17] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *Int. J. Comput. Vis.*, vol. 70, no. 3, pp. 241–255, 2006.

[18] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[19] F. I. Karahanoglu, I. Bayram, and D. Van De Ville, "A signal processing approach to generalized 1-D total variation," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5265–5274, Nov. 2011.

[20] M. Benning, C. Brune, M. Burger, and J. Müller, "Higher-order TV methods—Enhancement via Bregman iteration," *J. Sci. Comput.*, vol. 54, nos. 2–3, pp. 269–310, 2013.

[21] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.

[22] N. B. Karayiannis and A. N. Venetsanopoulos, "Regularization theory in image restoration—The stabilizing functional approach," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 7, pp. 1155–1179, Jul. 1990.

[23] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[24] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[25] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[26] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[28] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[29] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[30] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comp. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008.

[31] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.

[32] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer, 2013.

[33] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[34] P. G. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications*. Philadelphia, PA, USA: SIAM, 2013.

[35] W. Rudin, *Functional Analysis*, (International Series in Pure and Applied Mathematics), 2nd ed. New York, NY, USA: McGraw-Hill, 1991.

[36] F. Trèves, *Topological Vector Spaces, Distributions and Kernels*. New York, NY, USA: Dover, 2006.

[37] P. M. Prenter, *Splines and Variational Methods*. New York, NY, USA: Wiley, 1975.

[38] L. L. Schumaker, *Spline Functions: Basic Theory*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[39] O. L. Mangasarian and L. L. Schumaker, "Discrete splines via mathematical programming," *SIAM J. Control*, vol. 9, no. 2, pp. 174–183, 1971.

[40] I. J. Schoenberg, "Spline functions and the problem of graduation," *Proc. Nat. Acad. Sci. USA*, vol. 52, no. 4, pp. 947–950, Oct. 1964.

[41] M. Unser, J. Fageot, and J. P. Ward. (2016). "Splines are universal solutions of linear inverse problems with generalized-TV regularization." [Online]. Available: https://arxiv.org/abs/1603.01427

[42] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*. Philadelphia, PA, USA: SIAM, 1999.

Michael Unser (M'89–SM'94–F'99) is professor and director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His primary area of investigation is biomedical image processing. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has published over 250 journal papers on those topics. He is the author with P. Tafti of the book, An Introduction to Sparse Stochastic Processes, Cambridge University Press 2014.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda USA, conducting research on bioimaging.

Dr. Unser has held the position of associate Editor-in-Chief (2003-2005) for the IEEE Transactions on Medical Imaging. He is currently member of the editorial boards of SIAM J. Imaging Sciences and Foundations and Trends in Signal Processing. He is the founding chair of the technical committee on Bio Imaging and Signal Processing (BISP) of the IEEE Signal Processing Society. Prof. Unser is an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is the recipient of several international prizes including three IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010).

Julien Fageot graduated from the École Normale Supérieure, Paris, France, in 2012. He received the M.Sc. degree in mathematics from the Université Paris-Sud, France, in 2009 and the M.Sc. degree in imaging science from the École Normale Supérieure, Cachan, France, in 2011. He is currently pursuing the Ph.D. degree with the Biomedical Imaging Group under the direction of M. Unser. He is mainly working on random processes and their applications to signal processing. His research interests include stochastic models for sparse signals and spline theory.

Harshit Gupta received the B. Tech. in Electronics and Communication Engineering in 2015 from the Indian Institute of Technology, Guwahati, India. He is currently pursuing a Ph.D. degree with the Biomedical Imaging Group under the direction of M. Unser. His research focuses on splines, regularization theory and the resolution of inverse problems in imaging.