Robust Sparse Analysis Regularization

Samuel Vaiter, Gabriel Peyré, Charles Dossal, and Jalal Fadili

***Abstract*—This paper investigates the theoretical guarantees of ℓ^1-analysis regularization when solving linear inverse problems. Most of previous works in the literature have mainly focused on the sparse synthesis prior where the sparsity is measured as the ℓ^1 norm of the coefficients that synthesize the signal from a given dictionary. In contrast, the more general analysis regularization minimizes the ℓ^1 norm of the correlations between the signal and the atoms in the dictionary, where these correlations define the analysis support. The corresponding variational problem encompasses several well-known regularizations such as the discrete total variation and the fused Lasso. Our main contributions consist in deriving sufficient conditions that guarantee exact or partial analysis support recovery of the true signal in presence of noise. More precisely, we give a sufficient condition to ensure that a signal is the unique solution of the ℓ^1-analysis regularization in the noiseless case. The same condition also guarantees exact analysis support recovery and ℓ^2-robustness of the ℓ^1-analysis minimizer vis-à-vis an enough small noise in the measurements. This condition turns to be sharp for the robustness of the sign pattern. To show partial support recovery and ℓ^2-robustness to an arbitrary bounded noise, we introduce a stronger sufficient condition. When specialized to the ℓ^1-synthesis regularization, our results recover some corresponding recovery and robustness guarantees previously known in the literature. From this perspective, our work is a generalization of these results. We finally illustrate these theoretical findings on several examples to study the robustness of the 1-D total variation, shift-invariant Haar dictionary, and fused Lasso regularizations.**

***Index Terms*—Analysis regularization, fused Lasso, inverse problems, ℓ^1-minimization, noise robustness, sparsity, synthesis regularization, total variation (TV), union of subspaces, wavelets.**

I. Introduction

A. Inverse Problems and Signal Priors

THIS paper considers the stability of regularized inverse problems using sparsity-promoting priors. The forward model in many data acquisition scenarios can be formulated as the action of a linear mapping on some unknown (sought-after) signal contaminated by an additive noise. This takes the form

$$y = \Phi x_0 + w \tag{1}$$

where $y \in \mathbb{R}^Q$ are the observations, $x_0 \in \mathbb{R}^N$ is the unknown signal to recover, w is the noise supposed to be of bounded ℓ^2-norm, and Φ is a bounded linear operator which maps the signal domain $\mathbb{R}^N$ into the observation domain $\mathbb{R}^Q$ where generally $Q \leqslant N$. Even when $Q = N$, the mapping Φ is in general ill-conditioned or even singular. This makes the problem of solving for an accurate approximation of x_0 from the system (1) ill-posed; see, for instance, [1] for an introduction to inverse problems.

However, the situation radically changes if one has some prior information about the underlying object x_0. Regularization is a popular way to impose such a prior, hence making the search for solutions feasible. The general variational problem we consider can be stated as

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda R(x) \tag{2}$$

where the first term is the data fidelity reflecting ℓ^2-boundedness of the noise, and R is an appropriate (prior) regularization term through which some regularity is enforced on the recovered signal. The regularization parameter $\lambda > 0$ should be adapted to balance between the allowed fraction of noise level and regularity as dictated by the prior on x_0.

For noiseless observations, i.e., $w = 0$, taking the limit $\lambda \to 0$, we end up solving the constrained problem

$$\min_{x \in \mathbb{R}^N} R(x) \quad \text{subject to} \quad \Phi x = y. \tag{3}$$

A popular class of priors are quadratic forms $R(x) = \langle x, Kx \rangle$ where K is a symmetric semidefinite positive kernel. Problems (2) and (3) then correspond to Tikhonov regularization, which typically induces some kind of uniform smoothness in the recovered signal. More advanced priors that have received considerable interest in the recent years rely on nonquadratic, generally nonsmooth, functionals such as those promoting sparsity of the signal in some transform domain (e.g., its wavelet transform or its derivatives). These sparsity priors are at the heart of this paper. They will be discussed in more detail after some necessary definitions and notations are first introduced in the following section.

B. Notations

Throughout this paper, we focus on real vector spaces. The variable x will denote a vector in $\mathbb{R}^N$, y will be a vector in $\mathbb{R}^Q$, and α will be a vector in $\mathbb{R}^P$.

The sign vector $\mathrm{sign}(\alpha)$ of $\alpha \in \mathbb{R}^P$ is

$$\forall i \in \{1, \ldots, P\}, \quad \mathrm{sign}(\alpha)_i = \begin{cases} +1 & \text{if} \quad \alpha_i > 0 \\ 0 & \text{if} \quad \alpha_i = 0 \\ -1 & \text{if} \quad \alpha_i < 0 \,. \end{cases}$$

Manuscript received March 25, 2012; revised October 01, 2012; accepted December 02, 2012. Date of publication December 12, 2012; date of current version March 13, 2013. This work was supported in part by the SIGMA-Vision European Research Council Grant.

S. Vaiter and G. Peyré are with the CEREMADE, CNRS, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France (e-mail: samuel.vaiter@ceremade.dauphine.fr; peyre@ceremade.dauphine.fr).

C. Dossal is with the Institut Mathématiques de Bordeaux, Université Bordeaux 1, 33405 Talence Cedex, France (e-mail: charles.dossal@math.u-bordeaux1.fr).

J. Fadili is with the GREYC, CNRS-ENSICAEN-Université de Caen, 14050 Caen Cedex, France (e-mail: jalal.fadili@ensicaen.fr).

Communicated by M. Elad, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2012.2233859

Its support is

$$\operatorname{supp}(\alpha) = \{i \in \{1,\ldots,P\} \setminus \alpha_i \neq 0\}.$$

For a subset $I \subset E$, $|I|$ will denote its cardinality, and $I^c = E \backslash I$ its complement.

The p,q-operator (induced) norm of a matrix M is

$$\|M\|_{p,q} = \max_{x \neq 0} \frac{\|Mx\|_q}{\|x\|_p}.$$

The matrix M_J for J a subset of $\{1,\ldots,P\}$ is the submatrix whose columns are indexed by J. Similarly, the vector s_J is the restriction of s to the entries of s indexed by J.

The matrix Id is the identity matrix, where the underlying space will be clear from the context. For any matrix M, M^+ is its Moore–Penrose pseudoinverse and M^* is its adjoint. $M^{+,*}$ is the adjoint of the Moore–Penrose pseudoinverse of M.

C. Synthesis and Analysis Sparsity Priors

1) Synthesis Sparsity Prior: Sparse regularization is a popular class of priors to model natural signals and images; see, for instance, [2]. We recall that a dictionary $D = (d_i)_{i=1}^P$ is a (possibly redundant, i.e., $P > N$) collection of P atoms $d_i \in \mathbb{R}^N$. It can also be viewed as a linear mapping from $\mathbb{R}^P$ to $\mathbb{R}^N$ which is used to synthesize a signal $x \in \operatorname{Im}(D) \subseteq \mathbb{R}^N$ as

$$x = D\alpha = \sum_{i=1}^{P} \alpha_i d_i$$

where α is the coefficient vector that synthesizes x from the dictionary D.

In its simplest form, the sparsity of coefficients $\alpha \in \mathbb{R}^P$ is measured using the ℓ^0 pseudonorm

$$R_0(\alpha) = \|\alpha\|_0 = |\operatorname{supp}(\alpha)|.$$

Minimizing (2) or (3) with $R = R_0$ is however known to be NP-hard; see, for instance, [3]. Several workarounds have been proposed to alleviate this difficulty. A first family of methods relies on greedy algorithms [4]. The most popular ones are matching pursuit [5] and orthogonal matching pursuit [6], [7]. A second family of methods, which is the focus of this paper, relies on convex relaxation which amounts to replacing the ℓ^0 pseudonorm by the ℓ^1 norm [8].

The sparsest set of coefficients, according to the ℓ^1 norm, defines a signal prior which is the image of $\|.\|_1$ under D

$$R_S(x) = \min_{\alpha \in \mathbb{R}^P} \|\alpha\|_1 \quad \text{subject to} \quad x = D\alpha.$$

Therefore, any solution x of (2) using $R = R_S$ can be written as $x = D\alpha$ where α is a solution of

$$\min_{\alpha \in \mathbb{R}^P} \frac{1}{2}\|y - \Psi\alpha\|_2^2 + \lambda\|\alpha\|_1 \tag{4}$$

where $\Psi = \Phi D$. ℓ^1-regularization was first considered in the statistical community in [9] where it was coined Lasso. Note that it was originally introduced as an ℓ^1-ball constrained optimization and in the overdetermined case. It is also known in the signal processing community as basis pursuit denoising [10]. Such a problem corresponds to the so-called sparse synthesis regularization as sparsity is assumed on the coefficients α that synthesize the signal $x = D\alpha$. In the noiseless case, the constrained problem (3) becomes

$$\min_{\alpha \in \mathbb{R}^P} \|\alpha\|_1 \quad \text{subject to} \quad y = \Psi\alpha \tag{5}$$

which goes by the name of basis pursuit after [10]. Taking $D = \mathrm{Id}$ amounts to assuming sparsity of the signal itself, and was used for instance for sparse spike train deconvolution in seismic imaging [11]. Sparsity in orthogonal as well as redundant wavelet dictionaries is popular to model natural signals and images that exhibit certain singularities [2].

2) Analysis Sparsity Prior: Analysis regularization corresponds to using $R = R_A$ in (2) where

$$R_A(x) = \|D^* x\|_1 = \sum_{i=1}^{P} |\langle d_i, x\rangle|$$

in which case (2) reads

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\|D^* x\|_1. \tag{$\mathcal{P}_\lambda(y)$}$$

Of course, D^* is not in general the adjoint operator of a full rank dictionary D. Note that the analysis problem $(\mathcal{P}_\lambda(y))$ is more general than the synthesis one (4) because the latter is recovered by taking $D = \mathrm{Id}$ and $\Psi = \Phi$ in the former.

As the objective in $(\mathcal{P}_\lambda(y))$ is proper (i.e., not infinite everywhere), continuous, and convex, it is a classical existence result that the set of (global) minimizers is nonempty and compact if and only if

$$\operatorname{Ker} \Phi \cap \operatorname{Ker} D^* = \{0\}. \tag{H_0}$$

From now on, we suppose that this condition holds.

In the noiseless case, the ℓ^1-analysis equality-constrained problem is

$$\min_{x \in \mathbb{R}^N} \|D^* x\|_1 \quad \text{subject to} \quad \Phi x = y. \tag{$\mathcal{P}_0(y)$}$$

One of the most popular analysis sparsity-inducing regularizations is the total variation (TV), which was first introduced for denoising (in a continuous setting) in [12]. It roughly corresponds to taking D^* as a derivative operator. Typically, for 1-D discrete signals, D can be taken as a dictionary of forward finite differences D_{DIF} where

$$D_{\mathrm{DIF}} = \begin{pmatrix} -1 & & & 0 \\ +1 & -1 & & \\ & +1 & \ddots & \\ & & \ddots & -1 \\ 0 & & & +1 \end{pmatrix}. \tag{6}$$

The corresponding prior R_A favors piecewise constant signals and images. A comprehensive review of TV regularization can be found in [13].

The theoretical properties of TV regularization have been previously studied. A distinctive feature of this regularization is its tendency to yield a staircasing effect, where discontinuities not present in the original data might be artificially created by the regularization. This effect has been studied by Nikolova in the discrete case in a series of papers; see, e.g., [14] and in [15] in the continuous setting. The stability of the discontinuity set of the solution of the 2-D continuous TV-based denoising problem is investigated in [16]. Section IV-C shows how our results also shed some light on this staircasing effect for 1-D discrete signals.

It is also possible to use a dictionary D of shift invariant wavelets so that the corresponding regularization term R_A can be viewed as a multiscale (higher order) TV [17]. Such a prior tends to favor piecewise regular signals and images. From a numerical standpoint, an extensive study is reported in [18] using these redundant dictionaries to highlight differences between synthesis and analysis sparsity priors for inverse problems.

As a last example of sparse analysis regularization, we would like to mention the fused Lasso [19], where D is the concatenation of a discrete derivative and a weighted identity. The corresponding prior R_A promotes both sparsity of the signal and its derivative, hence favoring the grouping of nonzero coefficients in blocks.

3) Synthesis Versus Analysis Priors: In a synthesis prior, the vector α that synthesizes the signal x from the dictionary D is sparse, whereas in an analysis prior, the correlation between the signal x and the atoms in the D is sparse. Some insights on the relation and distinction between analysis and synthesis-based sparsity regularizations were first given in [20]. When D is orthogonal, and more generally when D is square and invertible, $(\mathcal{P}_\lambda(y))$ and the Lasso entail equivalent regularizations in the sense that the set of minimizers of one problem can be retrieved from that of an equivalent form of the other through a bijective change of variable. However, when D is redundant, synthesis and analysis regularizations differ significantly.

D. Union of Subspaces Model

As analysis regularization involves the sparsity of the correlation vector D^*x, it is thus natural to keep track of the support of D^*x. To fix terminology, we define this support and its complement.

Definition 1: The *D-support* I of a vector $x \in \mathbb{R}^N$ is $I = \operatorname{supp}(D^*x) \subset \{1,\dots,P\}$. Its *D-cosupport* J is $J = I^c = \{1,\dots,P\} \setminus I$.

A signal x such that D^*x is sparse lives in a subspace $\mathcal{G}_J$ of small dimension whose formal definition is as follows.

Definition 2: Given a dictionary D, and J a subset of $\{1,\dots,P\}$, the *cospace* $\mathcal{G}_J$ is defined as

$$\mathcal{G}_J = \mathrm{Ker} D_J^*$$

where we recall that D_J is the subdictionary whose columns are indexed by J.

Following the cosparse model introduced in [21], the signal space can thus be decomposed as

$$\mathbb{R}^N = \bigcup_{k \in \{0,\dots,N\}} \Theta_k$$

where

$$\Theta_k = \{\mathcal{G}_J \setminus J \subseteq \{1,\dots,P\} \text{ and } \dim \mathcal{G}_J = k\} \tag{7}$$

which is dubbed *union of subspaces* of dimension k.

The union of subspaces associated with synthesis regularization, i.e., $D = \mathrm{Id}$, corresponds to Θ_k as the set of axis-aligned subspaces of dimension k. For the 1-D TV prior, where $D = D_{\mathrm{DIF}}$ as defined in (6), Θ_k is the set of piecewise constant signals with $k-1$ steps. Several examples of subspaces Θ_k, including those corresponding to translation invariant wavelets, are discussed in [21].

More general union of subspaces models (not necessarily corresponding to analysis regularizations) has been introduced in sampling theory to model various types of nonlinear signal ensembles; see, for instance, [22]. Union of subspaces models has been extensively studied for the recovery from pointwise sampling measurements [22] and compressed sensing measurements [23]–[26].

E. Organization of This Paper

The rest of this paper is organized as follows. Section II details our main contributions. Section III draws some connections with relevant previous work. Section IV illustrates our results on some examples. The proofs are deferred to Section V.

II. CONTRIBUTIONS

This paper proves the following three main results.

1) *Robustness to small noise*: we provide a sufficient condition on x_0 ensuring that the solution of $(\mathcal{P}_\lambda(y))$ is unique, lives in the same cospace, and is close to x_0 when w is small enough.
2) *Noiseless identifiability*: under the same sufficient condition, x_0 is guaranteed to be the unique solution of $(\mathcal{P}_0(y))$ when $w = 0$.
3) *Robustness to bounded noise*: we then give a sufficient condition that depends on the D-cosupport of x_0 under which the solution of $(\mathcal{P}_\lambda(y))$ is unique and close to x_0 for an arbitrary bounded noise w, with the proviso that λ is large enough.

Each contribution will be rigorously described in a corresponding section.

It is worth mentioning that our results will extend previously known ones in the synthesis case; see, for instance, [27]–[31]. Additionally, there are only a few recent works that we are aware of and which give provable guarantees using analysis regularization for exact recovery in the noiseless case [21] or accurate and robust recovery in the noisy case [32]–[37]. We will discuss this prior literature in detail in Section III. Nevertheless, to the best of our knowledge, it appears that our work is the first that addresses the aforementioned three questions in the analysis case.

For some cosupport J, the invertibility of Φ on $\mathcal{G}_J$ will play a pivotal role in our theory. This is achieved by imposing that

$$\mathrm{Ker}\Phi \cap \mathcal{G}_J = \{0\}. \tag{H_J}$$

To get the gist of the importance of (H_J), consider the noiseless case where we want to recover a D-sparse signal x_0 from

$y = \Phi x_0$. Let J be the D-cosupport of x_0 and assume that it is known. As $x_0 \in \mathcal{G}_J \cap \{x : y = \Phi x\}$, for x_0 to be uniquely recovered from y, (H_J) must be verified. Conversely, if x_0 is such that (H_J) does not hold, then any $x_0 + h$, with $h \in \mathrm{Ker}\Phi \cap \mathcal{G}_J$, is also a candidate solution, i.e., $x_0 + h \in \mathcal{G}_J \cap \{x : y = \Phi x\}$. Clearly, one cannot reconstruct such D-sparse objects.

With assumption (H_J) at hand, we are in a position to define the following matrix whose role will be clarified shortly.

Definition 3: Let J be a D-cosupport. Suppose that (H_J) holds. We define the operator $A^{[J]}$ as

$$A^{[J]} = U\left(U^*\Phi^*\Phi U\right)^{-1} U^* \tag{8}$$

where U is a matrix whose columns form a basis of $\mathcal{G}_J$.

It is worth noting that the action of $A^{[J]}$ on a vector can be computed without explicitly constructing a basis of $\mathcal{G}_J$ by solving the quadratic problem

$$A^{[J]}u = \underset{D_J^* x = 0}{\arg\min}\ \frac{1}{2}\|\Phi x\|^2 - \langle x,\, u\rangle.$$

A. Robustness to Small Noise

Our first contribution consists in showing that ℓ^1-analysis regularization is robust to a small enough noise under a sufficient condition that depends on the sign of $D^* x_0$ and its D-cosupport. This condition will be formulated via the following criterion.

Definition 4: Let $s \in \{-1, 0, +1\}^P$, I its support, and $J = I^c$. Suppose that (H_J) holds. The analysis *identifiability criterion* $\mathbf{IC}$ of s is defined as

$$\mathbf{IC}(s) = \min_{u \in \mathrm{Ker} D_J} \|\Omega^{[J]} s_I - u\|_\infty$$

where

$$\Omega^{[J]} = D_J^+(\Phi^*\Phi A^{[J]} - \mathrm{Id}) D_I.$$

We have the following theorem.

Theorem 1: Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-support I and D-cosupport $J = I^c$. Let $y = \Phi x_0 + w$. Assume that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Then, there exist constants $c_J > 0$ and $\tilde{c}_J > 0$ satisfying

$$\frac{\|w\|_2}{T} < \frac{\tilde{c}_J}{c_J} \quad \text{and} \quad T = \min_{i \in \{1, \cdots, |I|\}} |D_I^* x_0|_i$$

such that if λ is chosen according to

$$c_J \|w\|_2 < \lambda < T\tilde{c}_J$$

the vector

$$\hat{x}^\star = x_0 + A^{[J]}\Phi^* w - \lambda A^{[J]} D_I \mathrm{sign}(D^* x_0) \tag{9}$$

is the unique solution of $(\mathcal{P}_\lambda(y))$. Moreover

$$\hat{x}^\star \in \mathcal{G}_J \quad \text{and} \quad \mathrm{sign}(D^* x_0) = \mathrm{sign}(D_I^* \hat{x}^\star).$$

In plain words, Theorem 1 asserts that when $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$, the support and sign pattern of $D^* x_0$ are exactly recovered by solving $(\mathcal{P}_\lambda(y))$ with λ wisely chosen and provided that the nonzero entries of $D_I^* x_0$ are large enough compared to noise. In addition, if λ is chosen proportional to the noise level, (9) implies

$$\|\hat{x}^\star - x_0\|_2 = O(\|w\|_2).$$

Remark 1: One may question the benefit of minimizing over $\mathrm{Ker} D_J$ in the criterion $\mathbf{IC}$. First note that $\mathbf{IC}(s)$ is upper bounded by $\|\Omega^{[J]} s_I\|_\infty$. For D with maximally linear independent columns, $\mathrm{Ker} D_J = \{0\}$ and $\mathbf{IC}$ is large. On the other hand, when $\mathrm{Ker} D_J$ is large, minimizing the (translated) ℓ^∞-norm over $\mathrm{Ker} D_J$ is likely to produce lower values of $\mathbf{IC}$. In a nutshell, linear dependences among the columns of D, in some sense, are desirable to optimize the value of $\mathbf{IC}$. This is in agreement with the observations of [21].

At this stage, one may wonder whether the sufficient condition $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$ can be weakened while ensuring both sign consistency and cospace recovery by solving $(\mathcal{P}_\lambda(y))$ in presence of small noise. The following proposition provides a first answer by proving that the condition is in some sense necessary.

Proposition 1: Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J. Let $y = \Phi x_0 + w$. Suppose that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) > 1$. If

$$\frac{1}{\lambda}\|\Pi^{[J]} w\|_\infty < \mathbf{IC}(\mathrm{sign}(D^* x_0)) - 1 \tag{10}$$

where $\Pi^{[J]} = D_J^+ \Phi^*(\Phi A^{[J]}\Phi^* - \mathrm{Id})$, then for any solution $x^\star$ of $(\mathcal{P}_\lambda(y))$, we have

$$\mathrm{sign}(D^* x_0) \neq \mathrm{sign}(D^* x^\star).$$

In plain words, for signals x_0 with $\mathbf{IC}(\mathrm{sign}(D^* x_0)) > 1$, the associated sign vector and D-support cannot be simultaneously identified by solving $(\mathcal{P}_\lambda(y))$ even with a small noise for the range of λ obeying (10).

B. Noiseless Identifiability

In the noiseless case, $w = 0$, the criterion $\mathbf{IC}$ can be used to test identifiability. A vector x_0 is said to be *identifiable* if x_0 is the unique solution of $(\mathcal{P}_0(\Phi x_0))$. We will prove the following theorem.

Theorem 2: Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J. Suppose that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Then, x_0 is identifiable.

The conclusions of Proposition 1 remain valid even in the noiseless case.

Corollary 1: Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J. Suppose that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) > 1$. Then, for any $\lambda > 0$ and any solution $x^\star$ of $(\mathcal{P}_\lambda(\Phi x_0))$

$$\mathrm{sign}(D^* x_0) \neq \mathrm{sign}(D^* x^\star).$$

When $\mathbf{IC}(\mathrm{sign}(D^* x_0)) = 1$, Proposition 1 and Corollary 1 do not allow to conclude. In Section IV-C, a family of signals x_0 is built such that $\mathbf{IC}(\mathrm{sign}(D^* x_0)) = 1$, and where we show that depending on the noise structure, recovery can be possible or not.

C. Robustness to Bounded Noise

Let us now turn to robustness to an arbitrary bounded noise. To this end, we introduce the following criterion, which is a strengthened version of the $\mathbf{IC}$ criterion.

Definition 5: The analysis *recovery criterion (RC)* of $I \subset \{1, \ldots, P\}$ is defined as

$$\mathbf{RC}(I) = \max_{\substack{p_I \in \mathbb{R}^{|I|} \\ \|p_I\|_\infty \leqslant 1}} \min_{u \in \mathrm{Ker} D_J} \|\Omega^{[J]} p_I - u\|_\infty.$$

It is clear that if I is the D-support of x_0, $\mathbf{RC}(I) < 1$ implies $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Moreover, $\mathbf{RC}$ depends solely on the D-support while $\mathbf{IC}$ relies both on the D-support and the sign vector $\mathrm{sign}(D^* x_0)$.

In Theorem 1, the assumption on T plays a pivotal role: if T is too small, there is no way to distinguish the small components of $D^* x_0$ from the noise. If no assumption is made on T, it turns out that one can nevertheless expect robustness to an arbitrary bounded noise if the parameter λ is large enough. In this case, solving $(\mathcal{P}_\lambda(y))$ allows us to recover a unique vector which lives in the same cospace $\mathcal{G}_J$ as the unknown signal x_0, and whose ℓ^2 distance from x_0 is within a factor of the noise level.

Theorem 3: Let I be a fixed D-support, $J = I^c$ its associated D-cosupport. Let $y = \Phi x_0 + w$. Suppose that (H_J) holds. If $\mathbf{RC}(I) < 1$ and

$$\lambda = \rho \|w\|_2 \frac{c_J}{1 - \mathbf{RC}(I)} \quad \text{with} \quad \rho > 1$$

where

$$c_J = \|D_J^+ \Phi^* (\Phi A^{[J]} \Phi^* - \mathrm{Id})\|_{2,\infty}$$

then for every x_0 of D-support I, problem $(\mathcal{P}_\lambda(y))$ has a unique solution $x^\star$ whose D-support is included in I and $\|x_0 - x^\star\|_2 = O(\|w\|_2)$. More precisely

$$\|x_0 - x^\star\|_2 \leqslant \|A^{[J]}\|_{2,2} \|w\|_2 \left(\|\Phi\|_{2,2} + \frac{\rho c_J}{1 - \mathbf{RC}(I)} \|D_I\|_{2,\infty} \right).$$

III. Related Works

A. Previous Works on Synthesis Identifiability and Robustness

There is an extensive literature on guarantees for identifiability and robustness to noise of ℓ^1 sparse synthesis regularization, i.e., Lasso in (4). In [28], Fuchs introduced a synthesis identifiability criterion $\mathbf{IC}_S$, which is a specialization of our $\mathbf{IC}$ to the case where $D = \mathrm{Id}$.

Definition 6: Let $s \in \{-1, 0, +1\}^P$, I its support, and J its cosupport. We suppose Ψ_I is full rank. The criterion $\mathbf{IC}_S$ of a sign vector s associated with a support I is defined as

$$\mathbf{IC}_S(s) = \|\Omega^S s_I\|_\infty \quad \text{where} \quad \Omega^S = \Psi_J^* \Psi_I^{+,*}.$$

Let us point out that the full rank assumption on Ψ_I is a particularization of (H_J) to the synthesis prior case.

The following result is proved in [28]. We restate it here for completeness.

Theorem ([28]): Let $\alpha_0 \in \mathbb{R}^P$ be a fixed vector of support I. If Ψ_I has full rank and $\mathbf{IC}_S(\mathrm{sign}(\alpha_0)) < 1$, then α_0 is identifiable, i.e., it is the unique solution of (4) for $y = \Psi \alpha_0$.

Note that the aforementioned condition is also known as the irrepresentable condition in the statistical literature.

The work of Tropp [29], [30] in the synthesis case developed a sufficient noise robustness condition built upon the so-called exact recovery coefficient (ERC) of the support.

Definition 7: The ERC of $I \subset \{1, \ldots, P\}$ is defined as

$$\mathbf{ERC}(I) = \|\Omega^S\|_{\infty,\infty}.$$

Note again that while $\mathbf{IC}_S(s)$ depends both on the sign and the support, $\mathbf{ERC}$ depends only on the support and we have the inequality $\mathbf{IC}_S(s) \leqslant \mathbf{ERC}(I)$.

It is proved in [29] that $\mathbf{ERC}(I) < 1$ is a sufficient condition for partial support recovery and ℓ^2-consistency by solving the Lasso.

Theorem ([29]): Let I be a fixed support. Suppose that Ψ_I has full rank. If $\mathbf{ERC}(I) < 1$ and λ large enough, then for every α_0 of support I, problem (4) with $y = \Psi \alpha_0 + w$ has a unique solution $\alpha^\star$ whose support is included in I and $\|\alpha_0 - \alpha^\star\|_2 = O(\|w\|_2)$.

By noticing that when $D = \mathrm{Id}$, $\mathrm{Ker} D_J = \{0\}$, and by definition of the operator norm $\|\cdot\|_{\infty,\infty}$, we easily conclude that our criteria $\mathbf{IC}$ and $\mathbf{RC}$ are equivalent to $\mathbf{IC}_S$ and $\mathbf{ERC}$.

Proposition 2: If $D = \mathrm{Id}$, then $\mathbf{IC}(\mathrm{sign}(D^* x_0)) = \mathbf{IC}_S(\mathrm{sign}(D^* x_0))$ and $\mathbf{RC}(I) = \mathbf{ERC}(I)$.

There are of course many other sufficient conditions in the literature which provably guarantee uniqueness, identifiability, and noise robustness in the ℓ^1-synthesis regularization case; see [38] for a thorough review. Among the most popular, we have coherence-based conditions and those based on the RIP which plays a central role in the compressed sensing theory [31], [39].

In the inverse problems community, efforts have been undertaken to derive results of robustness to arbitrary bounded noise (so-called convergence rates), for ℓ^1-synthesis regularization to solve ill-posed linear inverse problems. In the regularization theory, the source or range condition as well as a restricted invertibility condition on Φ are generally imposed; see, e.g., [40]–[44] and references therein. For instance, the authors in [43] have shown that a strengthened version of the source condition generalizing $\mathbf{IC}_S(s) < 1$ is a necessary and sufficient condition for noise robustness with the rate $O(\|w\|_2)$. This source condition is detailed in (11) for the more general analysis setting. However, these results do not say anything about the sign and support recovery.

B. Previous Works on Analysis Identifiability and Robustness

It is only very recently that recovery and noise robustness theoretical guarantees of ℓ^1-analysis sparse regularization have been investigated. The previous works that we are aware of are [21] and [32]–[37].

Taking a compressed sensing perspective with a generalization of the RIP (called D-RIP) on Φ, and assuming that D is a tight frame, the authors [32] prove that ℓ^1-analysis regularization allows accurate and robust recovery from noisy measurements uniformly over *all* signals that are (even nearly) D-sparse. Needell and Ward [37] also give a provable guarantee of robust recovery for images from compressed measurements via TV regularization. As usual, the RIP-based guarantees are uniform and the (D-)RIP is satisfied for Gaussian matrices and other random ensembles. This setting is thus quite far from ours.

The work of [21] is much closer to ours. It studies noiseless identifiability using ℓ^0 and ℓ^1 sparse analysis regularization. Their result on ℓ^1-analysis noiseless identifiability is the following whose proof is inspired from an extension of the null space property [45] to the ℓ^1-analysis case.

Theorem ([21]): Let x_0 be a fixed vector whose D-support is I and D-cosupport J. Let M^* be any basis matrix of $\mathrm{Ker}\Phi$. Assume that $D_J^* M^*$ has full rank and $\mathbf{IC}_0(\mathrm{sign}(D^* x_0)) < 1$, where

$$\mathbf{IC}_0(s) = \|\daleth_I s_I\|_\infty \quad \text{where} \quad \daleth_I = (MD_J)^+ M D_I.$$

Then, x_0 is identifiable.

Note that $\mathbf{IC}_0(s) < 1$ does not imply $\mathbf{IC}(s) < 1$ neither the opposite. Moreover, unlike $\mathbf{IC}$, $\mathbf{IC}_0$ does not reduce to $\mathbf{IC}_S$ in the ℓ^1-synthesis case; see the discussion on their fundamental differences in [21, Sec. 5.3]. In the noisy case, $\mathbf{IC}_0(\mathrm{sign}(D^* x_0)) < 1$ is not sufficient to guarantee stability of the sign pattern to noise, even a small one. More precisely, if $\mathbf{IC}_0(\mathrm{sign}(D^* x_0)) < 1$ but $\mathbf{IC}(\mathrm{sign}(D^* x_0)) > 1$, then according to Proposition 1, any solution $x^\star$ of $(\mathcal{P}_\lambda(y))$, for λ satisfying (10), will violate the sign agreement property, i.e., $\mathrm{sign}(D^* x_0) \neq \mathrm{sign}(D^* x^\star)$. Robustness guarantees of ℓ^1-analysis regularization by an appropriate strengthening of the analysis equivalent of the null space property remain an open problem.

Turning to the inverse problems literature, some authors have established linear convergence rates. For instance, in [36], convergence (robustness) rates for convex regularizations R have been derived with respect to the Bregman divergence under a source condition. The Bregman divergence measures the distance between the regularization term R and its affine approximation at the true solution. Analysis-type regularizations where D^* is not necessarily injective, such as the TV, fall within the class of regularization functionals they considered. The author in [35] derived more general linear convergence rates for a large class of positively homogeneous convex sparsity promoting regularization functionals R, including analysis-type ones, under a source condition and a suitable restricted injectivity condition on Φ. The convergence was established with respect to the error in the solution measured in terms of the regularization functional. Specialized to the case of ℓ^1-analysis regularization, this result reads.

Theorem ([35]): Let $x_0 \in \mathbb{R}^N$ of D-support I and $y = \Phi x_0 + w$ such that $\|w\| = \epsilon$. Assume also that there exist α such that

$$\alpha \in \partial \| \cdot \|_1 (D^* x_0) \quad \text{and} \quad D\alpha \in \mathrm{Im}\Phi^* \tag{11}$$

(source condition). Let $J \subseteq I^c$ such that $\|\alpha_J\|_\infty < 1$. Suppose that (H_J) holds with such J. Then, for λ proportional to ϵ, there exists C independent of ϵ such that

$$\|D^*(x^\star - x_0)\|_2 \leqslant C\epsilon.$$

Interestingly, for $J = I^c$, if (H_J) is satisfied, $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$ implies that the source and restricted injectivity conditions stated in the previous theorem are in force. More precisely, the following holds.

Proposition 3: Let $x_0 \in \mathbb{R}^N$ of D-cosupport J such that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Then, the source and restricted injectivity conditions of Theorem ([35]) hold. The claimed convergence is therefore also valid.

However, in none of these works in the inverse problem literature, robustness with respect to the ℓ^2-norm, i.e., ℓ^2-distance of the solution from the true one, was established for general D. Of course, if D^* were injective, ℓ^2-robustness would follow immediately from [35]. In addition, their results do not allow to conclude anything about the sign and D-support recovery unless there is no noise.

IV. Examples

This section details algorithms to compute the criteria $\mathbf{IC}$ and $\mathbf{RC}$, together with a detailed study of three ℓ^1-analysis regularizations: TV, which when D is the shift-invariant Haar dictionary, and the fused Lasso. The source code used to produce the numerical results is available online at github.com/svaiter/robust_sparse_analysis_regularization.

A. Computing Sparse Analysis Regularization

It is not the main scope of this paper to give a comprehensive treatment of provably convergent minimization schemes that can be used to solve $(\mathcal{P}_\lambda(y))$. We describe one possible efficient algorithm to do so which originates from the realm of nonsmooth convex optimization theory, and more precisely, proximal splitting.

In the case where $\Phi = \mathrm{Id}$ (denoising), $(\mathcal{P}_\lambda(y))$ is strictly (actually strongly) convex, and one can compute its unique solution $x^\star$ by solving an equivalent Fenchel–Rockafellar dual problem [46]

$$x^\star = y + D\alpha^\star \quad \text{where} \quad \alpha^\star \in \underset{\|\alpha\|_\infty \leqslant \lambda}{\mathrm{arg\,min}} \|y + D\alpha\|_2^2.$$

The dual problem can be solved using, e.g., projected gradient descent or a multistep accelerated version of it.

In the general case, we advocate the use of a primal-dual algorithm such as the relaxed Arrow–Hurwicz scheme recently revitalized in [47]. This algorithm is designed to minimize the sum of two proper lower semicontinuous convex functions, one of which is composed of a linear bounded operator. To put problem $(\mathcal{P}_\lambda(y))$ in a form amenable to apply this scheme, we can rewrite it as follows:

$$\min_{x \in \mathbb{R}^N} F(K(x)) \text{ where } \begin{cases} F : (g, u) \mapsto \frac{1}{2}\|y - g\|_2^2 + \lambda \|u\|_1 \\ K : x \mapsto (\Phi x, D^* x). \end{cases}$$

The primal-dual algorithm requires the computation of the proximity operator of F which is a separable and simple function, i.e., its proximity operator is easy to compute. Recall that the proximity operator prox_f of a proper lower semicontinuous function and convex f is defined as

$$\forall x \in \mathbb{R}^N, \quad \operatorname{prox}_f(x) = \operatorname*{arg\,min}_{z \in \mathbb{R}^N} \frac{1}{2}\|z - x\|_2^2 + f(z).$$

Computing prox_F involves applying a soft thresholding (the ℓ^1-part) and a diagonal Wiener filtering (the separable quadratic part).

B. *Computing the Criteria*

In the case where $\operatorname{Ker} D_J \neq \{0\}$, computing $\mathbf{IC}(\operatorname{sign}(D^* x_0))$ entails solving a convex minimization problem. The latter can be cast as

$$\mathbf{IC}(\operatorname{sign}(D^* x_0)) = \min_{u \in \mathbb{R}^N} \|\Omega^{[J]} \operatorname{sign}(D^* x_0)_I - u\|_\infty + \iota_{\operatorname{Ker} D_J}(u)$$

where $\iota_{\operatorname{Ker} D_J}$ is the indicator function of $\operatorname{Ker} D_J$, i.e.,

$$\iota_{\operatorname{Ker} D_J}(u) = \begin{cases} 0 & \text{if } u \in \operatorname{Ker} D_J \\ +\infty & \text{otherwise.} \end{cases}$$

The previous objective is the sum of a translated ℓ^∞-norm and the indicator function of $\operatorname{Ker} D_J$. It can then be solved efficiently with the Douglas–Rachford splitting algorithm [48]. This will necessitate to compute the proximity operator of $\iota_{\operatorname{Ker} D_J}$ which is the orthogonal projector on $\operatorname{Ker} D_J$, and $\operatorname{prox}_{\|\Omega^{[J]} \operatorname{sign}(D^* x_0)_I - \cdot\|_\infty}$ can be computed with standard proximal calculus rules such as Moreau-identity

$$\operatorname{prox}_{\gamma \|\cdot\|_\infty}(x) = x - P_{\|\cdot\|_1}\left(\frac{x}{\gamma}\right), \quad \forall \gamma > 0$$

where $P_{\|\cdot\|_1}$ is the projection onto the unit ℓ^1 ball. This projector can be computed through sorting and soft thresholding; see [49] for details.

Unfortunately, computing $\mathbf{RC}$ (see Definition 5) is not as easy since it necessitates to solve a difficult maxi-minimization optimization problem which is nonsmooth, and convex in both u and p_I (while concavity in p_I would have been desirable). A stronger criterion, which is easy to compute, is obtained by taking $u = 0$ in $\operatorname{Ker} D_J$

$$\mathrm{w}\mathbf{RC}(I) = \|\Omega^{[J]}\|_{\infty,\infty}.$$

One can easily see that for every vector x_0 with D-support $I = \operatorname{supp}(D^* x)$, the following inequalities hold:

$$\mathbf{IC}(\operatorname{sign}(D^* x_0)) \leqslant \mathbf{RC}(I) \leqslant \mathrm{w}\mathbf{RC}(I).$$

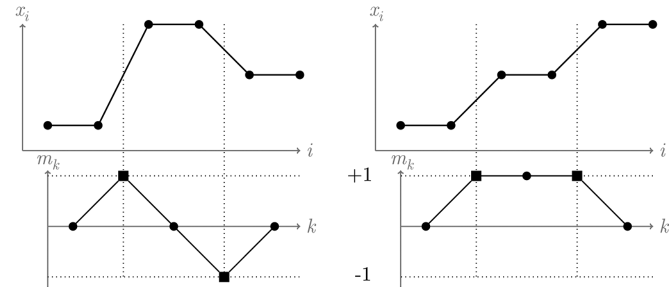

Fig. 1. (Top row) Two examples of signals x having two jumps. (Bottom row) Associated dual vector m.

For many cases, $\mathrm{w}\mathbf{RC}(I)$ might be strictly greater than 1. However, there are situations where $\mathrm{w}\mathbf{RC}(I) < 1$, such as when the associated cospace $\mathcal{G}_J$ is close to the whole space, i.e., high D-cosparsity or equivalently very small D-sparsity.

C. *TV Denoising*

The discrete 1-D TV corresponds to taking $D = D_{\mathrm{DIF}}$ as defined in (6). We recall that the TV union of subspaces model is formed by $\bigcup_k \Theta_k$ where Θ_k is the subspace of piecewise constant signals with $k - 1$ steps. We now define a subclass of such signals.

Definition 8: A signal is said to *contain a staircase subsignal* if there exists $i \in \{1 \ldots |I| - 1\}$ such that

$$\operatorname{sign}(D_I^* x)_i = \operatorname{sign}(D_I^* x)_{i+1} = \pm 1.$$

Fig. 1 shows examples of signals with (left) and without (right) staircase subsignals.

The following result will allow us to characterize robustness of TV regularization when $\Phi = \mathrm{Id}$, i.e., TV denoising.

Proposition 4: We consider the case where $\Phi = \mathrm{Id}$. If x_0 does not contain a staircase subsignal, then $\mathbf{IC}(\operatorname{sign}(D^* x_0)) < 1$. Otherwise, $\mathbf{IC}(\operatorname{sign}(D^* x_0)) = 1$.

Proof: Let $x^\star$ be the unique solution of $(\mathcal{P}_\lambda(y))$ with D-cosupport J and $I = J^c$. Using Lemma 1, there exists $\sigma \in \Sigma_{y,\lambda}(x^\star) \subset \mathbb{R}^{|J|}$. Since $D_J^+ A^{[J]} = 0$, we have $\Omega^{[J]} = -D_J^+ D_I$. We denote the vector m defined as

$$m : \begin{cases} m_I = s_I = \operatorname{sign}(D_I^* x) \\ m_J = \sigma = \Omega^{[J]} s_I. \end{cases}$$

The vector σ satisfies $(D_J^* D_J)\sigma = -(D_J^* D_I) s_I$. One can show that this implies that m is the solution of a discrete Poisson equation

$$\forall j \in J, \quad (\Delta m)_j = 0 \quad \text{and} \quad \begin{cases} \forall i \in I, \ m_i = s_i, \\ m_0 = m_N = 0 \end{cases}$$

where $\Delta = DD^*$ is a discrete Laplacian operator. This implies that for $i_1 < k < i_2$ where i_1, i_2 are consecutive indices of I, m is obtained by linearly interpolating (see Fig. 1) the values m_{i_1} and m_{i_2}, i.e.,

$$m_k = \rho m_{i_1} + (1 - \rho) m_{i_2} \quad \text{where} \quad \rho = \frac{k - i_1}{i_2 - i_1}.$$

Hence, if x_0 does not contain a staircase subsignal, one has $\|\Omega^{[J]} s_I\|_\infty < 1$. On the contrary, if there is i_1 such that $s_{i_1} = s_{i_2}$, where i_1 and i_2 are consecutive indices of I, then for every $i_1 < j < i_2, m_j = s_{i_1} = \pm 1$ which implies that $\mathbf{IC}(\text{sign}(D^* x_0)) = 1$. ■

This proposition together with Theorem 1 shows that if a signal x_0 does not have a staircase subsignal, then TV denoising from $y = x_0 + w$ is robust to a small noise. This means that if w is small enough, for λ proportional to the noise level, the TV denoised version of y contains the same jumps as x_0. However, the presence of a staircase in a signal, i.e., $\mathbf{IC}(\text{sign}(D^* x_0)) = 1$, does not comply with the assumptions of neither Theorem 1 nor Proposition 1. This prevents us from drawing positive or negative robustness conclusions.

To gain a better understanding of the latter situation, we build an instructive family of signals for which the $\mathbf{IC}$ criterion saturates at 1. It will turn out that depending on the structure of the noise w, the D-support of x_0 can be either stably identified or not.

For N a multiple of 4, we split $\{1, \ldots, N\}$ into four sets $l_k = \{(k-1)M+1, \ldots, kM\}$ of cardinality $M = N/4$. Let $\mathbf{1}_{l_k}$ be the boxcar signal whose support is l_k. Consider the staircase signal $x_0 = -\mathbf{1}_{l_1} + \mathbf{1}_{l_4}$ degraded by a deterministic noise w of the form $w = \epsilon(\mathbf{1}_{l_3} - \mathbf{1}_{l_2})$, where $\epsilon \in \mathbb{R}$. The observation vector $y = x_0 + w$ reads

$$y = -\mathbf{1}_{l_1} - \epsilon \mathbf{1}_{l_2} + \epsilon \mathbf{1}_{l_3} + \mathbf{1}_{l_4}.$$

Suppose that $\epsilon > 0$, then the solution $x^\star_\lambda$ of $(\mathcal{P}_\lambda(y))$ is

$$x^\star_\lambda = \left(-1 + \frac{\lambda}{M}\right) \mathbf{1}_{l_1} - \epsilon \mathbf{1}_{l_2} + \epsilon \mathbf{1}_{l_3} + \left(1 - \frac{\lambda}{M}\right) \mathbf{1}_{l_4}$$

if $0 \leqslant \lambda \leqslant \lambda_1 = M(1-\epsilon)$, and

$$x^\star_\lambda = \left(-\epsilon + \frac{\lambda - \lambda_1}{2M}\right)(\mathbf{1}_{l_1} + \mathbf{1}_{l_2}) + \left(\epsilon - \frac{\lambda - \lambda_1}{2M}\right)(\mathbf{1}_{l_3} + \mathbf{1}_{l_4})$$

if $\lambda_1 \leqslant \lambda \leqslant \lambda_2 = \lambda_1 + 2\epsilon M$, and 0 if $\lambda > \lambda_2$. Similarly, if $\epsilon < 0$, the solution $x^\star_\lambda$ reads

$$x^\star_\lambda = \left(-1 + \frac{\lambda}{M}\right) \mathbf{1}_{l_1} - \left(\epsilon + 2\frac{\lambda}{M}\right)(\mathbf{1}_{l_2} - \mathbf{1}_{l_3}) + \left(1 - \frac{\lambda}{M}\right) \mathbf{1}_{l_4}$$

if $0 \leqslant \lambda \leqslant \bar{\lambda}_1 = -\epsilon \frac{M}{2}$, and

$$x^\star_\lambda = \left(-1 + \frac{\lambda}{M}\right) \mathbf{1}_{l_1} + \left(1 - \frac{\lambda}{M}\right) \mathbf{1}_{l_4}$$

if $\bar{\lambda}_1 \leqslant \lambda \leqslant \bar{\lambda}_2 = M$, and 0 if $\lambda > \bar{\lambda}_2$. Fig. 2 displays plots of the coordinates' path for both cases. It is worth pointing out that when $\epsilon > 0$, the D-support of $x^\star_\lambda$ is always different from that of x_0 whatever the choice of λ, whereas in the case $\epsilon < 0$, for any $\bar{\lambda}_1 \leqslant \lambda \leqslant \bar{\lambda}_2$, the D-support of $x^\star_\lambda$ and sign of $D^* x^\star_\lambda$ are exactly those of x_0.

D. Shift-Invariant Haar Deconvolution

Sparse analysis regularization using a 1-D shift-invariant Haar dictionary is efficient to recover piecewise constant signals. This dictionary is built using a set of scaled and dilated Haar filters

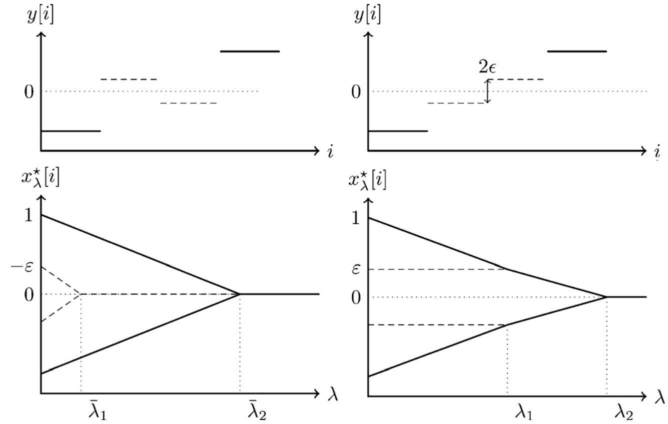

Fig. 2. (Top row) Signals y for (left) $\epsilon < 0$ and (right) $\epsilon > 0$. (Bottom row) Corresponding coordinates' path of $x^\star_\lambda$ as a function of λ. The solid lines correspond to the coordinates in l_1 and l_4, and the dashed ones to the coordinates in l_2 and l_3.

$$\psi_i^{(j)} = \frac{1}{2^{\tau(j+1)}} \begin{cases} +1 & \text{if } 0 \leqslant i < 2^j \\ -1 & \text{if } -2^j \leqslant i < 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\tau \geqslant 0$ is a normalization exponent. For $\tau = 1$, the dictionary is said to be *unit-normed*. For $\tau = 1/2$, it corresponds to a *Parseval tight frame*. The action on a signal x of the analysis operator corresponding to the translation-invariant Haar dictionary D_H is

$$D_H^* x = \left(\psi^{(j)} \star x\right)_{0 \leqslant j \leqslant J_{\max}},$$

where $\star$ stands for the discrete convolution (with appropriate boundary conditions) and $J_{\max} < \log_2(N)$, where N is the size of the signal. The analysis regularization $\|D_H^* x\|_1$ can also be written as the sum over scales of the TV semi-norms of filtered versions of the signal. As such, it can be understood as a sort of multiscale TV regularization. Apart from a multiplicative factor, one recovers TV when $J_{\max} = 0$.

We consider a noiseless convolution setting (for $N = 256$) where Φ is a circular convolution operator with a Gaussian kernel of standard deviation σ. We first study the impact of σ on the identifiability criterion $\mathbf{IC}$. The original signal x_η is a centered boxcar signal with a support of size $2\eta N$

$$x_\eta = \mathbf{1}_{\{\lfloor N/2 - \eta N \rfloor, \ldots, \lfloor N/2 + \eta N \rfloor\}}, \quad \eta \in (0, 1/2].$$

Fig. 3 displays the evolution of $\mathbf{IC}(\text{sign}(D_H^* x_0))$ as a function of σ for three dictionaries: the TV dictionary and the Haar wavelet dictionary with two normalization exponents $\tau = 1$ and $\tau = 0.5$. In this experiment, we chose $\eta = 0.2$. One can observe that the three curves pass through 1 for the same value of σ (near 1 here). In addition, in the identifiability regime, $\mathbf{IC}(\text{sign}(D_H^* x_0))$ appears smaller in the case of the unit-normed normalization (i.e., $\tau = 1$). However, one should avoid to infer stronger conclusions since a detailed computation of the constants involved in Theorem 1 would be necessary

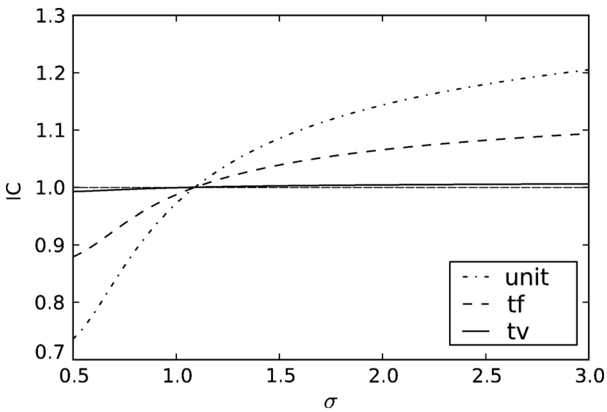

Fig. 3. Behavior of $\mathbf{IC}$ for a noiseless deconvolution scenario with a Gaussian blur and ℓ_1-analysis sparsity regularization in a shift-invariant Haar dictionary with $J_{\max} = 4$. $\mathbf{IC}$ is plotted as a function of the Gaussian blurring kernel size $\sigma \in [0.5, 3.0]$ for the TV dictionary and the Haar wavelet dictionary with two normalization exponents τ. Dash-dotted line: $\tau = 1$ (unit-normed). Dashed line: $\tau = 1/2$ (tight frame). Solid line: TV.

to completely and fairly compare the stability performance achieved with each of these three dictionaries.

E. Fused Lasso Compressed Sensing

Fused Lasso was introduced in [19]. It corresponds to taking

$$D = [\, D_{\mathrm{DIF}} \quad \epsilon \mathrm{Id} \,]$$

in $(\mathcal{P}_\lambda(y))$, where $\epsilon > 0$. The associated union of subspaces (7) is $\bigcup_k \Theta_k$, where Θ_k is the set of signals that are the sum of k boxcars of disjoint supports, i.e., a signal $x \in \Theta_k$ can be written as

$$x = \sum_{i=1}^{k} \gamma_i \mathbf{1}_{[a_i, b_i]}$$

where $\gamma_i \in \mathbb{R}$ and $a_i \leqslant b_i < a_{i+1}$.

We consider a noiseless compressed sensing setting (with the signal size $N = 256$) and examine the behavior of $\mathbf{IC}$ with respect to the undersampling ratio Q/N and the true signal properties. Φ is drawn from the standard Gaussian ensemble, i.e., $\Phi_{i,j} \sim_{i.i.d.} \mathcal{N}(0,1)$. The sampled signal $x_{\eta,\rho}$ is the superposition of two boxcars distant from each other by $2\rho N$ and each of support size ηN

$$x_{\eta,\rho} = \mathbf{1}_{\{\lfloor(\frac{1}{2}-\eta-\rho)N\rfloor,\dots,\lfloor(\frac{1}{2}-\rho)N\rfloor\}} + \mathbf{1}_{\{\lfloor(\frac{1}{2}+\rho)N\rfloor,\dots,\lfloor(\frac{1}{2}+\eta+\rho)N\rfloor\}}.$$

In our simulations, we fixed $\rho = 0.1$.

Fig. 4 depicts the evolution of the empirical probability with respect to the sampling of Φ of the event $\mathbf{IC} < 1$ as a function of the sampling ratio $Q/N \in [0.5, 1]$ and the boxcar support size $\eta \in [0.025, 0.15]$. This probability is computed from 1000 Monte Carlo replications of the sampling of Φ. With no surprise, one can clearly see that the probability increases as more measurements are collected. This probability profile also seems to be increasing as η decreases, but this is likely to be a consequence of the choice of the Fused Lasso parameter ϵ, and the conclusion may be different for other choices.

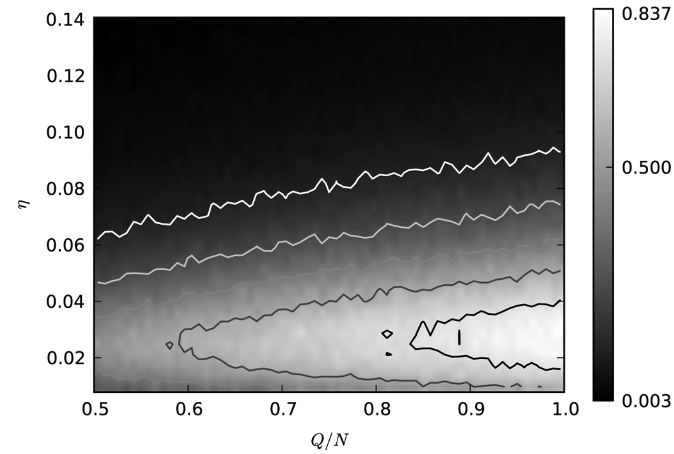

Fig. 4. Behavior of $\mathbf{IC}$ for a compressed sensing scenario matrix with a Gaussian measurement matrix and the Fused Lasso regularization. Empirical probability of the event $\mathbf{IC} < 1$ as a function of the sampling ratio $Q/N \in [0.5, 1]$ and the support size $\eta \in [0.025, 0.15]$ with $\epsilon = 50/N$.

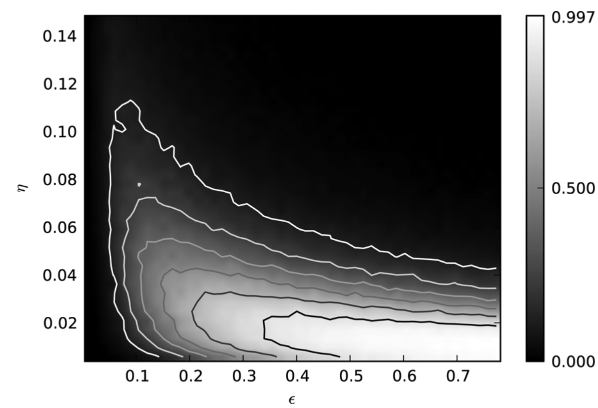

Fig. 5. Behavior of $\mathbf{IC}$ for a compressed sensing scenario matrix with a Gaussian measurement matrix and the fused Lasso regularization. Empirical probability of the event $\mathbf{IC} < 1$ as a function of the parameter $\epsilon \in [1/N, 200/N]$ and the support size $\eta \in [0.025, 0.15]$ with $Q/N = 0.8$.

This is indeed confirmed in our last experiment whose results are displayed in Fig. 5. It shows the evolution of the empirical probability of the event $\mathbf{IC} < 1$ as a function of the fused Lasso parameter $\epsilon \in [1/N, 200/N]$ and the support size $\eta \in [0.025, 0.15]$. This probability is again computed from 1000 Monte Carlo replications. Depending on the choice of ϵ, the probability profile does not necessarily exhibit a monotonic behavior as a function of η. For large values (more weight on Id in the fused Lasso dictionary), the probability decreases monotonically as η increases which can be explained by the fact that higher η corresponds to less sparse signals. As ϵ is lowered, higher weight is put on the TV regularization, and the behavior is not anymore monotonic. Now, the probability reaches a peak at intermediate values of η and then vanishes quickly. The peak probability also decreases with decreasing ϵ.

V. Proofs

This section details the proofs of our main results in Theorems 1–3. Throughout, we use the shorthand notation $\mathcal{L}_{y,\lambda}$ for the objective function in $(\mathcal{P}_\lambda(y))$

$$\mathcal{L}_{y,\lambda}(x) = \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \|D^* x\|_1.$$

We remind the reader that condition (H_0) is supposed to hold true in all our statements.

A. Preparatory Lemmata

We first need some key lemmata that will be central in our proofs.

The first one gives the first-order optimality conditions for the analysis variational problem $(\mathcal{P}_\lambda(y))$.

Lemma 1: A vector $x^\star$ is a solution of $(\mathcal{P}_\lambda(y))$ if, and only if, there exists $\sigma \in \mathbb{R}^{|J|}$, where J is the D-cosupport of $x^\star$, such that

$$\sigma \in \Sigma_{y,\lambda}(x^\star) \tag{12}$$

$$\Sigma_{y,\lambda}(x^\star) = \Big\{\sigma \in \mathbb{R}^{|J|} \setminus \Phi^*(\Phi x^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1\Big\} \tag{13}$$

where $I = J^c$ is the D-support of $x^\star$ and $s = \operatorname{sign}(D^* x^\star)$.

Proof: The subdifferential of a real-valued proper convex function $F : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is denoted ∂F. From standard convex analysis, we recall the definition of ∂F at a point x in the domain of F

$$\partial F(x) = \left\{g \in \mathbb{R}^N \setminus \forall z \in \mathbb{R}^N, F(z) \geqslant F(x) + \langle g,\, z - x\rangle\right\}.$$

It is clear from this definition that $x^\star$ is a (global) minimizer of F if, and only if, $0 \in \partial F(x)$. By classical subdifferential calculus, the subdifferential of $\mathcal{L}_{y,\lambda}$ at x is the nonempty convex compact set

$$\partial \mathcal{L}_{y,\lambda}(x) = \{\Phi^*(\Phi x - y) + \lambda D u \setminus u \in U(x)\}$$

where

$$\begin{aligned} U(x) &= (\partial \|\cdot\|_1)(D^* x) \\ &= \left\{u \in \mathbb{R}^N \setminus u_I = \operatorname{sign}(D^* x)_I \text{ and } \|u_J\|_\infty \leqslant 1\right\} \end{aligned}$$

where I and J are, respectively, the D-support and D-cosupport of x. Therefore, $0 \in \partial \mathcal{L}_{y,\lambda}(x^\star)$ is equivalent to the existence of $u \in \mathbb{R}^N$ such that $u_I = \operatorname{sign}(D^* x^\star)_I$ and $\|u_J\|_\infty \leqslant 1$ satisfying

$$\Phi^*(\Phi x^\star - y) + \lambda D u = 0.$$

Letting $\sigma = u_J$, this is equivalent to $\sigma \in \Sigma_{y,\lambda}(x^\star)$. ■

The following lemma is a key to prove uniqueness statements. It characterizes the normal cone at zero to the subdifferential of $\mathcal{L}_{y,\lambda}$ at a minimizer $x^\star$. By definition, this normal cone is

$$\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0) = \left\{z \in \mathbb{R}^N \setminus \langle z,\, d\rangle \leqslant 0, \forall d \in \partial \mathcal{L}_{y,\lambda}(x^\star)\right\}.$$

Lemma 2: Let $x^\star$ be a solution of $(\mathcal{P}_\lambda(y))$ whose D-support is $I^\star$. Suppose there exist $J \subseteq (I^\star)^c$ and $\sigma \in \Sigma_{y,\lambda}(x^\star)$ with $\|\sigma_J\|_\infty < 1$. Then

$$\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0) \subseteq (\operatorname{Im} D_J)^\perp = \mathcal{G}_J.$$

Moreover, if J is the D-cosupport of $x^\star$, then

$$\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0) = \mathcal{G}_J.$$

Proof: Let $I = J^c$. We decompose I such that $I = I^\star \cup J^\star$ where $J^\star = (I^\star)^c \cap I$. Since $\|\sigma_J\|_\infty < 1$, it follows that $\bar{u}$ defined by

$$\bar{u} : \begin{cases} \bar{u}_{I^\star} &= \operatorname{sign}(D^*_{I^\star} x^\star) \\ \bar{u}_{J^\star} &= \sigma_{J^\star} \\ \bar{u}_J &= \sigma_J \end{cases}$$

is such that $\|\bar{u}_J\|_\infty < 1$ and therefore from Lemma 1

$$\Phi^*(\Phi x^\star - y) + \lambda D \bar{u} = 0.$$

Let $0 < \epsilon < 1$ such that $\|\sigma_J\|_\infty = 1 - \epsilon$. Consider the set

$$\mathcal{U} = \left\{u \in \mathbb{R}^P \setminus \|u_J - \bar{u}_J\|_\infty \leqslant \epsilon \quad \text{and} \quad u_I = \bar{u}_I\right\}.$$

For every $u \in \mathcal{U}$, we define

$$d_u = \Phi^*(\Phi x^\star - y) + \lambda D u$$

and we denote

$$\mathcal{D} = \{d_u\}_{u \in \mathcal{U}}.$$

We therefore have

$$\begin{aligned} d_u &= \lambda D(u - \bar{u}) = \lambda D_I(u_I - \bar{u}_I) + \lambda D_J(u_J - \bar{u}_J) \\ &= \lambda D_J(u_J - \bar{u}_J) \end{aligned}$$

since $u_I = \bar{u}_I$.

Let $z \in \mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0)$ and $u \in \mathcal{U}$. By construction of u, we have that

$$\|u_J\|_\infty \leqslant \|u_J - \bar{u}_J\|_\infty + \|\bar{u}_J\|_\infty \leqslant 1$$

and

$$u_{I^\star} = \operatorname{sign}(D^*_{I^\star} x^\star) \quad \text{and} \quad \|u_{J^\star}\|_\infty \leqslant 1.$$

Clearly, $d_u \in \partial \mathcal{L}_{y,\lambda}(x^\star)$. In view of the definition of $\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0)$, we know that

$$\langle z,\, d\rangle \leqslant 0, \quad \forall d \in \partial \mathcal{L}_{y,\lambda}(x^\star).$$

In particular

$$\langle z,\, d_u\rangle \leqslant 0, \quad \forall u \in \mathcal{U}.$$

Now, observe that $\forall u \in \mathcal{U}$, $2\bar{u} - u \in \mathcal{U}$ and $d_{2\bar{u}-u} = -d_u$. Indeed, $(2\bar{u} - u)_I = 2\bar{u}_I - u_I = \bar{u}_I$ and

$$\|(2\bar{u} - u)_J - \bar{u}_J\|_\infty = \|\bar{u}_J - u_J\|_\infty \leqslant \epsilon.$$

Moreover

$$\begin{aligned} d_{2\bar{u}-u} &= \Phi^*(\Phi x^\star - y) + \lambda D(2\bar{u} - u) \\ &= \underbrace{\Phi^*(\Phi x^\star - y) + \lambda D\bar{u}}_{=0} - \lambda D(u - \bar{u}) \\ &= -d_u. \end{aligned}$$

This implies that

$$\forall u \in \mathcal{U}, \quad \langle z,\, d_u \rangle \leqslant 0 \quad \text{and} \quad \langle z,\, -d_u \rangle = \langle z,\, d_{2\bar{u}-u} \rangle \leqslant 0.$$

That is

$$\forall u \in \mathcal{U}, \quad \langle z,\, d_u \rangle = 0.$$

Let $v \in \mathrm{Im} D_J \setminus \{0\}$. Then, there exist $\mu_v \in \mathbb{R}^*$ and $\sigma_v \in \mathbb{R}^{|J|}$ such that

$$\mu_v v = D_J \sigma_v \quad \text{and} \quad \|\sigma_v\|_\infty \leqslant \epsilon.$$

Let the vector u defined as

$$u : \begin{cases} u_I = & \bar{u}_I \\ u_J = & \bar{u}_J + \sigma_v \end{cases}$$

u is by construction an element of $\mathcal{U}$ since $\|u_J - \bar{u}_J\|_\infty = \|\sigma_v\|_\infty \leqslant \epsilon$. Therefore, the associated vector d_u is

$$d_u = \lambda D_J(u_J - \bar{u}_J) = \lambda D_J \sigma_v = \frac{\lambda}{\mu_v} v$$

i.e., $\mathrm{Im}(D_J) \subseteq \mathrm{Im}(\mathcal{D})$. Since $\mathrm{Im}(\mathcal{D}) \subseteq \mathrm{Im}(D_J)$, we get $\mathrm{Im}(D_J) = \mathrm{Im}(\mathcal{D})$. This together with the fact that $\langle z,\, d_u \rangle = 0$ implies

$$\langle z,\, v \rangle = \frac{\mu_v}{\lambda} \langle z,\, d_u \rangle = 0.$$

We conclude that $\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0) \subseteq (\mathrm{Im} D_J)^\perp = \mathcal{G}_J$.

Suppose now that J is the D-cosupport of $x^\star$, i.e., $J = (I^\star)^c$. We prove that

$$\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0) = \mathcal{G}_J.$$

To this end, we show that $\partial \mathcal{L}_{y,\lambda}(x^\star) \subseteq \mathrm{Im} D_J$. Indeed, let $d \in \partial \mathcal{L}_{y,\lambda}(x^\star)$. We write $d = \Phi^*(\Phi x^\star - y) + \lambda D u$ with $u_I = \mathrm{sign}(D_I^* x)_I$ and $\|u_J\|_\infty \leqslant 1$. Since $0 \in \partial \mathcal{L}_{y,\lambda}(x^\star)$, one has

$$d = \lambda D(u - \bar{u}) = \lambda D_J(u_J - \bar{u}_J)$$

since $u_I = \bar{u}_I$. This implies that $(\mathrm{Im} D_J)^\perp \subseteq \mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0)$. In view of the assertion in the first part, we conclude. ■

The following lemma gives a sufficient condition, which guarantees that $(\mathcal{P}_\lambda(y))$ has exactly one minimizer.

Lemma 3: Let $x^\star$ be a vector of D-support $I^\star$. Suppose that there exist $J \subseteq (I^\star)^c$ such that (H_J) holds and $\sigma \in \mathbb{R}^{|(I^\star)^c|}$ such that

$$\sigma \in \Sigma_{y,\lambda}(x^\star) \quad \text{and} \quad \|\sigma_J\|_\infty < 1.$$

Then, $x^\star$ is the unique solution of $(\mathcal{P}_\lambda(y))$.

Proof: For notational convenience, we write $\mathcal{L}_{y,\lambda}$ as

$$\mathcal{L}_{y,\lambda}(x) = vq(x) + \lambda \|D^* x\|_1 \quad \text{where} \quad q(x) = \frac{1}{2}\|y - \Phi x\|_2^2.$$

Let $h \in \mathbb{R}^N \setminus \{0\}$. Two different cases occur.

1) If $h \notin \mathcal{G}_J$, then using Lemma 2, $h \notin \mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^\star)}(0)$. This negation means that $\langle d,\, h \rangle > 0$ for some $d \in \partial \mathcal{L}_{y,\lambda}(x^\star)$, whence it follows immediately that

$$\mathcal{L}_{y,\lambda}(x^\star + h) \geqslant \mathcal{L}_{y,\lambda}(x^\star) + \langle d,\, h \rangle > \mathcal{L}_{y,\lambda}(x^\star).$$

2) Let us turn to the case $h \in \mathcal{G}_J$. Since (H_J) holds, q, hence $\mathcal{L}_{y,\lambda}$, is strongly convex on $\mathcal{G}_J$ with some modulus $c > 0$. Consequently, for any $v \in \partial_{\|D^* \cdot\|_1}(x^\star)$, we have

$$\begin{aligned} \mathcal{L}_{y,\lambda}(x^\star + h) &\geqslant \mathcal{L}_{y,\lambda}(x^\star) + \langle \nabla q(x^\star) + \lambda v,\, h \rangle + \frac{c}{2}\|h\|_2^2 \\ &> \mathcal{L}_{y,\lambda}(x^\star) + \langle \nabla q(x^\star) + \lambda v,\, h \rangle \end{aligned}$$

$x^\star$ is a minimizer if, and only if, $\exists v \in \partial_{\|D^* \cdot\|_1}(x^\star)$ such that

$$\lambda v + \nabla q(x^\star) = 0.$$

This yields

$$\mathcal{L}_{y,\lambda}(x^\star + h) > \mathcal{L}_{y,\lambda}(x^\star).$$

Altogether, we have proved that for any $h \in \mathbb{R}^N \setminus \{0\}$, $\mathcal{L}_{y,\lambda}(x^\star + h) > \mathcal{L}_{y,\lambda}(x^\star)$, or equivalently that $x^\star$ is the unique minimizer of $(\mathcal{P}_\lambda(y))$. ■

The following lemma gives an implicit equation satisfied by any (nonnecessarily unique) minimizer $x^\star$ of $(\mathcal{P}_\lambda(y))$.

Lemma 4: Let $x^\star$ be a solution of $(\mathcal{P}_\lambda(y))$. Let I be the D-support and J be the D-cosupport of $x^\star$ and $s = \mathrm{sign}(D^* x^\star)$. We suppose that (H_J) holds. Then, $x^\star$ satisfies

$$x^\star = A^{[J]} \Phi^* y - \lambda A^{[J]} D_I s_I. \tag{14}$$

Proof: Owing to the first-order necessary and sufficient minimality condition (Lemma 1), there exists $\sigma \in \Sigma_{y,\lambda}(x^\star)$ satisfying

$$\Phi^*(\Phi x^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0. \tag{15}$$

By definition, $x^\star \in \mathcal{G}_J = (\mathrm{Im} D_J)^\perp$. We can then write $x^\star = U\alpha$ for some $\alpha \in \mathbb{R}^{\dim(\mathcal{G}_J)}$. Since $U^* D_J = 0$, multiplying both sides of (15) on the left-hand side by U^*, we get

$$U^* \Phi^* (\Phi U \alpha - y) + \lambda U^* D_I s_I = 0.$$

Since $U^* \Phi^* \Phi U$ is invertible, the implicit equation of $x^\star$ follows immediately. ■

Suppose now that a vector satisfies the aforementioned implicit equation. The next lemma derives two equivalent necessary and sufficient conditions to guarantee that this vector is actually a (possibly unique) solution to $(\mathcal{P}_\lambda(y))$.

Lemma 5: Let $y \in \mathbb{R}^Q$ and let J be a D-cosupport such that (H_J) holds, and $I = J^c$. Suppose that $\hat{x}^\star$ satisfies

$$\hat{x}^\star = A^{[J]} \Phi^* y - \lambda A^{[J]} D_I s_I$$

where $s = \operatorname{sign}(D^* \hat{x}^\star)$. Then, $\hat{x}^\star$ is a solution of $(\mathcal{P}_\lambda(y))$ if, and only if, there exists $\sigma \in \mathbb{R}^{|J|}$ satisfying one of the following equivalent conditions:

$$\sigma - \Omega^{[J]} s_I + \frac{1}{\lambda} \Pi^{[J]} y \in \operatorname{Ker} D_J \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1 \tag{16}$$

or

$$\tilde{\Pi}^{[J]} y - \lambda \tilde{\Omega}^{[J]} s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1 \tag{17}$$

where $\tilde{\Omega}^{[J]} = (\Phi^* \Phi A^{[J]} - \mathrm{Id}) D_I$, $\tilde{\Pi}^{[J]} = \Phi^*(\Phi A^{[J]} \Phi^* - \mathrm{Id})$, $\Omega^{[J]} = D_J^+ \tilde{\Omega}^{[J]}$, and $\Pi^{[J]} = D_J^+ \tilde{\Pi}^{[J]}$.

Moreover, if $\|\sigma\|_\infty < 1$ then $\hat{x}^\star$ is the unique solution of $(\mathcal{P}_\lambda(y))$.

Proof: First, we observe that $\hat{x}^\star \in \mathcal{G}_J$. According to Lemma 1, $\hat{x}^\star$ is a solution of $(\mathcal{P}_\lambda(y))$ if, and only if, there exists $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$. Since (H_J) holds, $A^{[J]}$ is properly defined. We can then plug the assumed implicit equation in (13) to get

$$\Phi^*(\Phi A^{[J]} \Phi^* y - \lambda \Phi A^{[J]} D_I s_I - y) + \lambda D_I s_I + \lambda D_J \sigma = 0.$$

Rearranging the terms multiplying y and s_I, we arrive at

$$\Phi^*(\Phi A^{[J]} \Phi^* - \mathrm{Id}) y - \lambda(\Phi^* \Phi A^{[J]} - \mathrm{Id}) D_I s_I + \lambda D_J \sigma = 0.$$

This shows that $x^\star$ is a minimizer of $(\mathcal{P}_\lambda(y))$ if and only if

$$\tilde{\Pi}^{[J]} y - \lambda \tilde{\Omega}^{[J]} s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1.$$

To prove the equivalence with (17), we first note that $U^* \tilde{\Omega}^{[J]} = 0$ implying that $\operatorname{Im}(\tilde{\Omega}^{[J]}) \subseteq \operatorname{Im}(D_J)$, and thus, $\tilde{\Omega}^{[J]} = D_J \Omega^{[J]}$. With a similar argument, we get $\tilde{\Pi}^{[J]} = D_J \Pi^{[J]}$. Hence, the existence of $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$ such that $\|\sigma\|_\infty \leqslant 1$ is equivalent to

$$D_J \sigma = D_J \Omega^{[J]} s_I - \frac{1}{\lambda} D_J \Pi^{[J]} y \quad \text{where} \quad \|\sigma\|_\infty \leqslant 1$$

which in turn is equivalent to

$$\sigma - \Omega^{[J]} s_I + \frac{1}{\lambda} \Pi^{[J]} y \in \operatorname{Ker} D_J \quad \text{where} \quad \|\sigma\|_\infty \leqslant 1.$$

Replacing the inequality by a strict inequality condition gives the uniqueness of $x^\star$ by virtue of Lemma 3. ■

B. Proof of Theorem 1

Recall the analysis identifiability criterion $\mathbf{IC}$ from Definition 4.

Proof: The proof is divided into three steps.

1) We give a first condition on λ to ensure $\operatorname{sign}(D^* \hat{x}^\star) = \operatorname{sign}(D^* x_0)$.
2) We then derive another condition on $\frac{\|w\|_2}{\lambda}$ to guarantee that the minimality conditions are satisfied at $\hat{x}^\star$, and assuming $\mathbf{IC}(\operatorname{sign}(D^* x_0)) < 1$ that $\hat{x}^\star$ is the unique solution to $(\mathcal{P}_\lambda(y))$.
3) We finally prove that these two conditions are compatible.

Let us consider the vector

$$\hat{x}^\star = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I$$

where $s = \operatorname{sign}(D^* x_0)$. Obviously, $\hat{x}^\star \in \mathcal{G}_J$.

1) We first give a condition on λ to ensure sign consistency, i.e.,

$$\operatorname{sign}(D^* \hat{x}^\star) = \operatorname{sign}(D^* x_0) \stackrel{\text{def.}}{=} s.$$

The two vectors have the same sign if

$$\begin{aligned} \forall i \in I, \quad |D_I^* x_0|_i &> |D_I^*(\hat{x}^\star - x_0)|_i \\ &= |D_I^* A^{[J]} \Phi^* w - \lambda D_I^* A^{[J]} D_I s_I|_i. \end{aligned} \tag{18}$$

Let us upper bound $\|D_I^*(\hat{x}^\star - x_0)\|_\infty$ as follows:

$$\begin{aligned} &\|D_I^*(\hat{x}^\star - x_0)\|_\infty \\ &\leqslant \|D_I^* A^{[J]}\|_{\infty,\infty} \left(\|\Phi^* w\|_\infty + \lambda \|D_I s_I\|_\infty\right) \\ &\leqslant \|D_I^* A^{[J]}\|_{\infty,\infty} \left(\|\Phi^*\|_{2,\infty} \|w\|_2 + \lambda \|D_I\|_{\infty,\infty}\right). \end{aligned}$$

Introducing

$$T = \min_{i \in \{1, \cdots, |I|\}} |D_I^* x_0|_i > 0$$

the condition

$$T > \|D_I^* A^{[J]}\|_{\infty,\infty} \left(\|\Phi^*\|_{2,\infty} \|w\|_2 + \lambda \|D_I\|_{\infty,\infty}\right) \tag{19}$$

is sufficient for (18) to hold true.

2) We now turn to the second step of the proof. Observe that $\tilde{\Pi}^{[J]} y = \tilde{\Pi}^{[J]} w$ since $x_0 \in \mathcal{G}_J$. Let $\bar{u} \in \operatorname{Ker} D_J$ a minimizer of $\|\Omega^{[J]} s_I - u\|_\infty$ over $\operatorname{Ker} D_J$. We consider the following candidate vector $\sigma \in \mathbb{R}^{|J|}$ defined by:

$$\sigma = -\bar{u} + \Omega^{[J]} s_I - \frac{1}{\lambda} \Pi^{[J]} w.$$

We have

$$\|\sigma\|_\infty \leqslant \|\Omega^{[J]} s_I - \bar{u}\|_\infty + \frac{1}{\lambda} \|\Pi^{[J]}\|_{2,\infty} \|w\|_2.$$

By definition of $\bar{u}$

$$\|\sigma\|_\infty \leqslant \mathbf{IC}(s) + \frac{1}{\lambda} \|\Pi^{[J]}\|_{2,\infty} \|w\|_2.$$

Thus, since $\mathbf{IC}(\operatorname{sign}(D^* x_0)) < 1$ and provided that

$$\|\Pi^{[J]}\|_{2,\infty} \frac{\|w\|_2}{\lambda} < 1 - \mathbf{IC}(\operatorname{sign}(D^* x_0)) \tag{20}$$

we have $\|\sigma\|_\infty < 1$. Appealing to Lemma 5, it follows that $\hat{x}^\star$ is the unique solution of $(\mathcal{P}_\lambda(y))$.

3) Let us show that (19) and (20) are in agreement. We introduce the constants c_J and $\tilde{c}_J$

$$c_J = \frac{\|\Pi^{[J]}\|_{2,\infty}}{1 - \mathbf{IC}(\operatorname{sign}(D^* x_0))}$$

and

$$\tilde{c}_J = \left[\|D_I^* A^{[J]}\|_{\infty,\infty} \left(\frac{\|\Phi^*\|_{2,\infty}}{c_J} + \|D_I\|_{\infty,\infty}\right)\right]^{-1}.$$

On the one hand, if

$$\lambda < T\tilde{c}_J$$

then

$$T > \lambda \|D_I^* A^{[J]}\|_{\infty,\infty} \left(\frac{\|\Phi^*\|_{2,\infty}}{c_J} + \|D_I\|_{\infty,\infty} \right).$$

On the other hand, if

$$c_J \|w\|_2 < \lambda$$

then

$$T > \|D_I^* A^{[J]}\|_{\infty,\infty} \left(\|\Phi^*\|_{2,\infty} \|w\|_2 + \lambda \|D_I\|_{\infty,\infty} \right)$$

which is condition (19). Moreover, $c_J \|w\|_2 < \lambda$ also implies that

$$\frac{\|\Pi^{[J]}\|_{2,\infty}}{1 - \mathbf{IC}(\mathrm{sign}(D^* x_0))} \frac{\|w\|_2}{\lambda} < 1$$

which is condition (20). ■

C. *Proof of Proposition 1*

Proposition 1 is a simple consequence of Lemmata 4 and 5.

Proof: Let $x^\star$ be a solution of $(\mathcal{P}_\lambda(y))$. Suppose that $\mathrm{sign}(D^* x_0) = \mathrm{sign}(D^* x^\star)$. As a consequence, J is the D-cosupport of $x^\star$. According to Lemmata 4 and 5, there exists σ such that $\|\sigma\|_\infty \leqslant 1$ and

$$\sigma - \Omega^{[J]} s_I + \frac{1}{\lambda} \Pi^{[J]} w \in \mathrm{Ker} D_J \quad \text{where} \quad s = \mathrm{sign}(D^* x_0)$$

or equivalently, there exists $-u \in \mathrm{Ker} D_J$ such that

$$\sigma = \Omega^{[J]} s_I - u - \frac{1}{\lambda} \Pi^{[J]} w.$$

It follows that

$$\|\sigma\|_\infty \geqslant \left| \|\Omega^{[J]} s_I - u\|_\infty - \frac{1}{\lambda} \|\Pi^{[J]} w\|_\infty \right|.$$

Since $\|\Omega^{[J]} s_I - u\|_\infty \geqslant \mathbf{IC}(s)$ and $\frac{1}{\lambda} \|\Pi^{[J]} w\|_\infty < \mathbf{IC}(s) - 1$ by assumption, we have

$$\|\Omega^{[J]} s_I - u\|_\infty - \frac{1}{\lambda} \|\Pi^{[J]} w\|_\infty \geqslant \mathbf{IC}(s) - \frac{1}{\lambda} \|\Pi^{[J]} w\|_\infty > 1.$$

This implies

$$\|\sigma\|_\infty > 1$$

which is a contradiction. ■

D. *Proof of Theorem 2*

Theorem 2 is proved in three steps.

1) First, we specialize Theorem 1 to the case $w = 0$.
2) Then, we show that under the condition $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$, the vector x_0 is a solution of $(\mathcal{P}_0(y))$.
3) Finally, we prove Theorem 2 by considering another feasible vector of $(\mathcal{P}_0(y))$.

Corollary 2: Let $x_0 \in \mathbb{R}^N$ be a fixed vector, I be its D-support, and $y = \Phi x_0$. Suppose that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Let $T = \min_{i \in \{1, \cdots, |I|\}} |D_I^* x_0|_i$. Then, for $\lambda < T\tilde{c}_J$

$$\hat{x}^\star = x_0 - \lambda A^{[J]} D_I s_I \quad \text{where} \quad s = \mathrm{sign}(D^* x_0)$$

is the unique solution of $(\mathcal{P}_\lambda(y))$.

Proof: Take $w = 0$ in Theorem 1. ■

Lemma 6: Let $x_0 \in \mathbb{R}^N$ be a fixed vector, I be its D-support, and $y = \Phi x_0$. Suppose that (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. Then, x_0 is a solution of $(\mathcal{P}_0(y))$.

Proof: According to Corollary 2, $(\mathcal{P}_\lambda(y))$ has a unique solution for $\lambda < T\tilde{c}_J$

$$x_\lambda^\star \overset{\text{def.}}{=} \hat{x}_\lambda^\star = x_0 - \lambda A^{[J]} D_I s_I$$

where $s = \mathrm{sign}(D^* x_0)$. Let $x_{(1)} \neq x_0$ such that $\Phi x_{(1)} = y$. For every $\lambda > 0$, one has $\mathcal{L}_{y,\lambda}(x_\lambda^\star) < \mathcal{L}_{y,\lambda}(x_{(1)})$ by definition of x_λ. Then

$$\|D^* x_\lambda^\star\|_1 < \|D^* x_{(1)}\|_1.$$

By continuity of the norm, and taking the limit as $\lambda \to 0$ in the last inequality yields

$$\|D^* x_0\|_1 \leqslant \|D^* x_{(1)}\|_1$$

whence it follows that x_0 is a solution of $(\mathcal{P}_0(y))$. ■

Proof of Theorem 2: Using Lemma 6, x_0 is a solution of $(\mathcal{P}_0(y))$. We shall prove that x_0 is actually unique. Let

$$x_{(1)} = x_0 + \lambda A^{[J]} D_I s_I.$$

For λ small enough, one has $\mathrm{sign}(D^* x_{(1)}) = \mathrm{sign}(D^* x_0)$. Then, if $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$, it follows from Corollary 2 that x_0 is the unique solution of $(\mathcal{P}_\lambda(y_1))$ where $y_1 = \Phi x_{(1)}$.

Let $x_{(2)} \in \mathbb{R}^N$ be another feasible point of $(\mathcal{P}_0(y))$, i.e., $\Phi x_{(2)} = y = \Phi x_0$ with $x_{(2)} \neq x_0$. Since x_0 is the unique solution of $(\mathcal{P}_\lambda(y_1))$, we obtain

$$\frac{1}{2} \|y_1 - \Phi x_0\|_2^2 + \lambda \|D^* x_0\|_1 < \frac{1}{2} \|y_1 - \Phi x_{(2)}\|_2^2 + \lambda \|D^* x_{(2)}\|_1$$

which implies that

$$\|D^* x_0\|_1 < \|D^* x_{(2)}\|_1.$$

This proves that indeed x_0 is the unique solution of $(\mathcal{P}_0(y))$. ■

E. *Proof of Theorem 3*

Recall the recovery criterion $\mathbf{RC}$ from Definition 5.

Proof: Consider the following restricted problem:

$$\min_{x\in\mathcal{G}_J} \frac{1}{2}\|y-\Phi x\|_2^2+\lambda\|D^*x\|_1. \qquad (\mathcal{P}_\lambda^J(y))$$

Our strategy is to construct a solution of $(\mathcal{P}_\lambda^J(y))$, and to show that it is the unique solution of $(\mathcal{P}_\lambda(y))$. To achieve this goal, we split the proof into four steps.

1) We exhibit $p_I^\star \in \mathbb{R}^{|I|}$ such that

$$U^*\left[\Phi^*(\Phi x^\star-y)+\lambda D_I p_I^\star\right]=0.$$

2) We prove that $x^\star$ satisfies an implicit equation of the form

$$x^\star = A^{[J]}\Phi^* y-\lambda A^{[J]}D_I p_I^\star.$$

3) We prove that $x^\star$ satisfies the first-order minimality conditions of Lemma 1 using the construction of $p_I^\star$.
4) Finally, we derive the ℓ^2-robustness bound.

By a simple change of variable $x=U\alpha$, we rewrite $(\mathcal{P}_\lambda^J(y))$ in an unconstrained form

$$\operatorname*{arg\,min}_{\alpha\in\mathbb{R}^{\dim\mathcal{G}_J}} \frac{1}{2}\|y-\Phi U\alpha\|_2^2+\lambda\|D_I^*U\alpha\|_1.$$

1) Applying Lemma 1 with ΦU and D_I^*U instead of Φ and D^*, $\alpha^\star$ is a solution of $(\mathcal{P}_\lambda^J(y))$ if, and only if, there exists $\sigma^\star$ with $\|\sigma^\star\|_\infty \leqslant 1$ such that

$$U^*\Phi^*(\Phi U\alpha^\star-y)+\lambda(U^*D_I)_{I^\star}s_{I^\star}+\lambda(U^*D_I)_{J^\star}\sigma^\star=0$$

where $I^\star\subseteq I$ is the (U^*D_I)-support of $U\alpha^\star$ and $J^\star=I\setminus I^\star$. We introduce $p_I^\star\in\mathbb{R}^{|I|}$ defined as

$$\forall i\in I,\quad (p_I^\star)_i=\begin{cases} s_i & \text{if } i\in I^\star\\ \sigma_i^\star & \text{if } i\in J^\star\end{cases}$$

which satisfies

$$D_I p_I^\star = D_{I^\star}s_{I^\star}+D_{J^\star}\sigma^\star.$$

The previous first-order optimality condition then takes the compact form

$$U^*\left[\Phi^*(\Phi U\alpha^\star-y)+\lambda D_I p_I^\star\right]=0. \qquad (21)$$

2) Owing to condition (H_J), $U^*\Phi^*\Phi U$ is invertible, and we obtain

$$\alpha^\star=(U^*\Phi^*\Phi U)^{-1}U^*\Phi^*y-\lambda(U^*\Phi^*\Phi U)^{-1}U^*D_I p_I^\star.$$

Multiplying both sides by U recovers $x^\star=U\alpha^\star$ as

$$x^\star = A^{[J]}\Phi^*y-\lambda A^{[J]}D_I p_I^\star. \qquad (22)$$

3) We now prove that $x^\star$ is a solution of $(\mathcal{P}_\lambda(y))$, i.e., there exists σ such that

$$\Phi^*(\Phi x^\star-y)+\lambda D_{I^\star}s_{I^\star}+\lambda D_{J\cup J^\star}\sigma=0 \quad\text{and}\quad \|\sigma\|_\infty\leqslant 1.$$

Take $\bar{u}$ such that

$$\bar{u}\in\operatorname*{arg\,min}_{u\in\operatorname{Ker}D_J}\|\Omega^{[J]}p_I^\star-u\|_\infty$$

and

$$\bar{\sigma}=\Omega^{[J]}p_I^\star-\bar{u}-\frac{1}{\lambda}\Pi^{[J]}w. \qquad (23)$$

We recall from Lemma 5 that

$$\tilde{\Omega}^{[J]}=(\Phi^*\Phi A^{[J]}-\mathrm{Id})D_I,\quad \tilde{\Pi}^{[J]}=\Phi^*(\Phi A^{[J]}\Phi^*-\mathrm{Id}),$$
$$\Omega^{[J]}=D_J^+\tilde{\Omega}^{[J]},\quad \Pi^{[J]}=D_J^+\tilde{\Pi}^{[J]}.$$

Plugging (22), we get

$$\begin{aligned}
&\Phi^*(\Phi x^\star-y)+\lambda D_I p_I^\star+\lambda D_J\bar{\sigma}\\
=&\Phi^*(\Phi(A^{[J]}\Phi^*y-\lambda A^{[J]}D_I p_I^\star)-y)\\
&+\lambda D_I p_I^\star+\lambda D_J D_J^+\tilde{\Omega}^{[J]}p_I^\star\\
&-\underbrace{\lambda D_J\bar{u}}_{=0}-D_J D_J^+\tilde{\Pi}^{[J]}y\\
=&(\mathrm{Id}-D_J D_J^+)(\tilde{\Pi}^{[J]}y-\lambda\tilde{\Omega}^{[J]}p_I^\star)\\
=&(\mathrm{Id}-D_J D_J^+)\left[\Phi^*(\Phi x^\star-y)+\lambda D_I p_I^\star\right].
\end{aligned}$$

Let us denote $v=\Phi^*(\Phi x^\star-y)+\lambda D_I p_I^\star$. From (21), we have $v\in\ker(U^*)=\operatorname{Im}(U)^\perp=\mathcal{G}_J{}^\perp$. Since $(\mathrm{Id}-D_J D_J^+)$ is the orthogonal projector on $\operatorname{Im}(D_J)^\perp=\mathcal{G}_J$, we conclude that $(\mathrm{Id}-D_J D_J^+)v=0$. It then follows that

$$\Phi^*(\Phi x^\star-y)+\lambda D_I p_I^\star+\lambda D_J\bar{\sigma}=0.$$

We can then write the bound

$$\|\bar{\sigma}\|_\infty\leqslant\|\Omega^{[J]}p_I^\star-\bar{u}\|_\infty+\frac{1}{\lambda}\|\Pi^{[J]}\|_{2,\infty}\|w\|_2.$$

From (23), and by definition of $\bar{u}$, we get the bound

$$\begin{aligned}
\|\bar{\sigma}\|_\infty&\leqslant\min_{u\in\operatorname{Ker}D_J}\|\Omega^{[J]}p_I^\star-u\|_\infty+\frac{1}{\lambda}\|\Pi^{[J]}\|_{2,\infty}\|w\|_2\\
&\leqslant\mathbf{RC}(I)+\frac{1}{\lambda}\|\Pi^{[J]}\|_{2,\infty}\|w\|_2.
\end{aligned}$$

Let σ defined by

$$\forall j\in\{1,\dots,P\}\setminus I,\quad \sigma_j=\begin{cases}\sigma_j^\star & \text{if } j\in J^\star\\ \bar{\sigma}_j & \text{if } j\in J.\end{cases}$$

Since by assumption $\mathbf{RC}(I)<1$ and

$$\lambda>\|w\|_2\frac{c_J}{1-\mathbf{RC}(I)}\quad\text{where}\quad c_J=\|\Pi^{[J]}\|_{2,\infty}$$

we get $\|\bar{\sigma}\|_\infty<1$ and $\|\sigma\|_\infty=\max(\|\bar{\sigma}\|_\infty,\|\sigma^\star\|_\infty)\leqslant 1$. Invoking Lemma 1, we conclude that $x^\star$ is a solution of $(\mathcal{P}_\lambda(y))$. Moreover, since $\|\bar{\sigma}\|_\infty<1$ and (H_J) holds, $x^\star$ is the unique solution of $(\mathcal{P}_\lambda(y))$ according to Lemma 3.

4) We now bound the ℓ^2-distance between x_0 and $x^\star$

$$\|x^\star-x_0\|_2=\|A^{[J]}\Phi^*y-\lambda A^{[J]}D_I p_I^\star-x_0\|_2.$$

Since $x_0 \in \mathcal{G}_J$, we have $A^{[J]}\Phi^* y = x_0 + A^{[J]}\Phi^* w$. Consequently

$$\begin{aligned}\|x^\star - x_0\|_2 &= \|A^{[J]}(\Phi^* w - \lambda D_I p_I^\star)\|_2\\ &\leqslant \|A^{[J]}\|_{2,2}\|w\|_2\left(\|\Phi^*\|_{2,2} + \frac{\rho c_J}{1-\mathbf{RC}(I)}\|D_I\|_{2,\infty}\right).\end{aligned}$$

This concludes the proof. ■

VI. Conclusion

In this paper, we provided theoretical guarantees for accurate and robust recovery with ℓ^1-analysis sparse regularization. We derived a sufficient condition under which the D-support and sign of the true signal can be exactly identified in presence of a small enough noise (and a fortiori without noise). We showed that this condition for support recovery is in some sense sharp. We proposed a stronger condition to ensure a partial support recovery for arbitrary noise if the regularization parameter is sufficiently large. As a byproduct, these conditions also guarantee robustness in ℓ^2-error. Some examples were provided and discussed to illustrate our results. For discrete 1-D TV regularization, we show that staircasing induces an instability of the D-support, i.e., jumps are not preserved. We believe that these contributions will allow us to gain a better understanding of the behavior of sparse analysis regularizations. We would like to emphasize that a distinctive feature of our approach with respect to the literature is that we have guarantees on the robustness of the cospace associated with the true signal. This approach often has a meaningful interpretation (such as the conservation of jumps for TV regularization).

Appendix
Proof of Proposition 3

Let $s = \operatorname{sign}(D^* x_0)$ and J be the D-cosupport of s. Let $\bar{u} \in \operatorname{Ker} D_J$ such that

$$\|\Omega^{[J]} s_I - \bar{u}\|_\infty = \mathbf{IC}(s).$$

Let α be such that $\alpha_I = s_I$ and $\alpha_J = \Omega^{[J]} s_I - \bar{u}$. Since $\mathbf{IC}(s) < 1$, we have $\|\alpha\|_\infty = \max(\|s_I\|_\infty, \|\alpha_J\|_\infty) \leqslant 1$, which shows that $\alpha \in \partial\|\cdot\|_1(D^* x_0)$, and therefore that $D\alpha \in \|D^* \cdot\|_1(x_0)$.

Now, as $\Omega^{[J]} = D_J^+ \tilde{\Omega}^{[J]}$ and $\operatorname{Im}\tilde{\Omega}^{[J]} \subseteq \operatorname{Im} D_J$, we have

$$D_J \Omega^{[J]} = D_J D_J^+ \tilde{\Omega}^{[J]} = P_{\operatorname{Im} D_J}\tilde{\Omega}^{[J]} = \tilde{\Omega}^{[J]} \tag{24}$$

where $P_{\operatorname{Im} D_J}$ is the orthogonal projection on $\operatorname{Im} D_J$. Since $\bar{u} \in \operatorname{Ker} D_J$ and owing to (24), we get

$$D_J \alpha_J = D_J(\Omega^{[J]} s_I - \bar{u}) = D_J \Omega^{[J]} s_I = \tilde{\Omega}^{[J]} s_I.$$

Using the expression of $\tilde{\Omega}^{[J]} = (\Phi^*\Phi A^{[J]} - \mathrm{Id}) D_I$, we obtain

$$D_J \alpha_J = \Phi^*\Phi A^{[J]} D_I s_I - D_I s_I = \Phi^*\Phi A^{[J]} D_I s_I - D_I \alpha_I.$$

Choosing $\eta = \Phi A^{[J]} D_I s_I$, and since $D\alpha = D_I \alpha_I + D_J \alpha_J$, we arrive at

$$\Phi^* \eta = D\alpha$$

or equivalently that $D\alpha \in \operatorname{Im}\Phi^*$. This concludes the proof.

Acknowledgment

We would like to thank the anonymous referees for their role in improving the original manuscript.

References

[1] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*. New York: Springer-Verlag, 2011, vol. 120.
[2] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2009.
[3] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.
[4] D. Needell, J. Tropp, and R. Vershynin, "Greedy signal recovery review," in *Proc. 42nd Asilomar Conf. IEEE Signals, Syst. Comput.*, 2008, pp. 1048–1050.
[5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[6] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. IEEE Signals, Syst. Comput. Conf. Rec.*, 1993, pp. 40–44.
[7] G. Davis, S. Mallat, and Z. Zhang, Adaptive time-frequency approximations with matching pursuits Courant Inst. Math. Sci., New York, Tech. Rep., 1994.
[8] D. 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.
[9] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Statist. Soc. Series B. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
[10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.
[11] F. Santosa and W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Statist. Comput.*, vol. 7, no. 4, pp. 1307–1330, 1986.
[12] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
[13] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Berlin, Germany: De Gruyter, 2010.
[14] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM J. Appl. Math.*, vol. 61, no. 2, pp. 633–658, 2000.
[15] W. Ring, "Structural properties of solutions to total variation regularization problems," *ESAIM Math. Modell. Numer. Anal.*, vol. 34, no. 4, pp. 799–810, 2000.
[16] V. Caselles, A. Chambolle, and M. Novaga, "The discontinuity set of solutions of the TV denoising problem and some extensions," *Multiscale Model. Simulat.*, vol. 6, no. 3, pp. 879–894, 2007.
[17] G. Steidl, J. Weickert, T. Brox, P. Mràzek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM J. Numer. Anal.*, vol. 42, no. 2, pp. 686–713, 2005.
[18] I. Selesnick and M. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *Proc. SPIE*, vol. 7446, p. 74460D, 2009.
[19] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused Lasso," *J. Roy. Statist. Soc.: Series B (Statist. Methodol.)*, vol. 67, no. 1, pp. 91–108, 2005.

[20] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Probl.*, vol. 23, no. 3, pp. 947–968, 2007.

[21] S. Nam, M. Davies, M. Elad, and R. Gribonval, The cosparse analysis model and algorithms 2011 [Online]. Available: arxiv-1106.4987v1

[22] Y. Lu and M. Do, “A theory for sampling signals from a union of subspaces,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.

[23] Y. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[24] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[25] T. Blumensath and M. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

[26] P. Boufounos, G. Kutyniok, and H. Rauhut, “Sparse recovery from combined fusion frame measurements,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3864–3876, Jun. 2011.

[27] C. Dossal, A necessary and sufficient condition for exact recovery by ℓ^1 minimization 2007 [Online]. Available: http://hal.archives-ouvertes.fr/hal-00164738

[28] J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[29] J. Tropp, “Just relax: Convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[30] J. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[31] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[32] E. Candès, Y. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2010.

[33] R. Giryes, S. Nam, R. Gribonval, and M. E. Davies, “Iterative cosparse projection algorithms for the recovery of cosparse vectors,” presented at the 19th Eur. Signal Process. Conf., Barcelona, Spain, 2011.

[34] T. Peleg and M. Elad, “Performance guarantees of the thresholding algorithm for the co-sparse analysis model,” *CoRR*, vol. abs/1203.2769, 2012.

[35] M. Grasmair, “Linear convergence rates for Tikhonov regularization with positively homogeneous functionals,” *Inverse Probl.*, vol. 27, no. 7, p. 075014, 2011.

[36] M. Burger and S. Osher, “Convergence rates of convex variational regularization,” *Inverse Probl.*, vol. 20, no. 5, p. 1411, 2004.

[37] D. Needell and R. Ward, “Stable image reconstruction using total variation minimization,” *CoRR*, vol. abs/1202.6429, 2012.

[38] A. Bruckstein, D. 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.

[39] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[40] D. A. Lorenz, “Convergence rates and source conditions for Tikhonov regularization with sparsity constraints,” *J. Inverse Ill-Posed Probl.*, vol. 16, no. 5, pp. 463–478, 2008.

[41] M. Grasmair, M. Haltmeier, and O. Scherzer, “Sparse regularization with ℓ_q penalty term,” *Inverse Probl.*, vol. 24, no. 5, p. 055020, 2008.

[42] M. Grasmair, “The residual method for regularizing ill-posed problems,” *Appl. Math. Comput.*, vol. 218, no. 6, pp. 2693–2710, 2011.

[43] M. Grasmair, O. Scherzer, and M. Haltmeier, “Necessary and sufficient conditions for linear convergence of ℓ_1-regularization,” *Commun. Pure Appl. Math.*, vol. 64, no. 2, pp. 161–182, 2011.

[44] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, ser. Applied Mathematical Sciences Series. New York: Springer-Verlag, 2009, vol. 167.

[45] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[46] A. Chambolle, “An algorithm for total variation minimization and applications,” *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[47] A. Chambolle and T. Pock, “First-order primal-dual algorithm for convex problems with applications to imaging,” *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[48] P. Combettes and J.-C. Pesquet, “A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE J. Sel. Top. Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[49] M. J. Fadili and G. Peyré, “Total variation projection with first order schemes,” *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 657–669, Mar. 2011.

Samuel Vaiter received the M.Sc. in mathematics, vision and learning from Université Paris-Descartes/École Normale Supérieure de Cachan, France, in 2011. Since Oct. 2011, he is a Ph.D. student in applied mathematics at CNRS-Ceremade, Université Paris-Dauphine, France. His research interests include image processing, inverse problems and sparse representations.

Gabriel Peyré graduated from Ecole Normale Supérieure de Cachan, France, in 2003 and received his Ph.D. in applied mathematics from École Polytechnique, Paris, France, in 2005. Since 2006, he has been a researcher at the Centre Nationale de Recherche Scientifique (CNRS), working in Ceremade, University Paris-Dauphine. He his head of the research group SIGMA-Vision, which is funded by the European Research Council (ERC). SIGMA-Vision activity is focused on sparse and adaptive representations with application in computer vision, computer graphics and neurosciences. Since 2005 Gabriel Peyré has co-authored 30 papers in international journals, 40 conference proceedings in top vision and image processing conferences, and two books. He was the main contributor of the 3rd edition of the book *Wavelet Tour of Signal Processing* of S. Mallat. He is the creator of the “Numerical tour of signal processing” (www.numerical-tours.com), a popular online repository of MATLAB/Scilab resources to teach modern signal and image processing.

Charles Dossal graduated from Ecole Normale Supérieure de Cachan, France, in 2002 and received his Ph.D. in applied mathematics from École Polytechnique, Paris, France, in 2005. Since 2006, he has been a assistant professor at Institut Mathématiques de Bordeaux in the University Bordeaux 1. His activity is focused on optimization, inverse problems, sparsity especially ℓ^1 minimization.

Jalal Fadili received the M.Sc. and Ph.D. degrees in signal and image processing from the University of Caen. He was a Research Associate with the University of Cambridge, U.K., 1999–2000 and an Associate Professor Sept. 2001–2011. Since Oct. 2011, he is a Full Professor. He is a project leader in a CNRS laboratory, France, from 2006. He also holds several scientific management positions (editorial activities, national excellence research networks), and serves as an expert of several national and international funding agencies. As of January 2013, he will be the director of the French national network on Mathematical Imaging and Applications. He also held several visiting positions at several universities (QUT-Australia, Stanford, CalTech, EPFL-Switzerland, MIT). In the last decade, he has been an invited or plenary speaker at various international events. He is also part of the scientific or program committees of several major international and national conferences, and he organized many important international or national events (conferences, workshops, symposia, summer schools). His research interests include mathematical signal and image processing, statistical estimation and detection theory, inverse problems, computational harmonic analysis, sparse representations, non-smooth optimization. His areas of application include medical and astronomical imaging. From 1998 to 2012, he has published 55 papers in the leading journals of these fields, 5 book chapters and co-authored one of the first books on the highly growing field of sparsity in 2010 (Cambridge University Press).