Robust Sparse Analysis Regularization

Samuel Vaiter*, Gabriel Peyré*, Charles Dossal†, and Jalal Fadili‡

September 28, 2011

Abstract

This paper studies the properties of ℓ^1-analysis regularization for the resolution of linear inverse problems. Most previous works consider sparse synthesis priors where the sparsity is measured as the ℓ^1 norm of the coefficients that synthesize the signal in 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. The corresponding variational problem includes several well-known regularizations such as the discrete total variation and the fused lasso.

We first prove that a solution of analysis regularization is a piecewise affine function of the observations. Similarly, it is a piecewise affine function of the regularization parameter. This allows us to compute the degrees of freedom associated to sparse analysis estimators. Another contribution gives a sufficient condition to ensure that a signal is the unique solution of the analysis regularization when there is no noise in the observations. The same criterion ensures the robustness of the sparse analysis solution to a small noise in the observations. Our last contribution defines a stronger sufficient condition that ensures robustness to an arbitrary bounded noise. In the special case of synthesis regularization, our contributions recover already known results, that are hence generalized to the analysis setting. We illustrate these theoritical results on practical examples to study the robustness of the total variation and the fused lasso regularizations.

Keywords: sparsity, analysis regularization, synthesis regularization, inverse problems, ℓ^1 minimization, union of subspaces, noise robustness, total variation, wavelets, Fused Lasso.

*{samuel.vaiter,gabriel.peyre}@ceremade.dauphine.fr, CEREMADE, CNRS, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France

†charles.dossal@math.u-bordeaux1.fr, IMB, Université Bordeaux 1, 351, Cours de la libération, 33405 Talence Cedex, France

‡Jalal.Fadili@greyc.ensicaen.fr, GREYC, CNRS-ENSICAEN-Université de Caen, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France

1 Introduction

1.1 Inverse Problems and Signal Priors

This paper considers the stability of inverse problems regularization using sparse priors. Many data acquisition systems are modeled using a linear mapping of some unknown source perturbed by an additive noise. This reads

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

where $y \in \mathbb{R}^Q$ are the observations, $x_0 \in \mathbb{R}^N$ the unknown signal to recover, w the noise and Φ a linear operator which maps the signal domain $\mathbb{R}^N$ into the observation domain $\mathbb{R}^Q$ where $Q \leqslant N$. The mapping Φ is in general ill-conditioned, which makes the recovery of an approximation of x_0 difficult, see for instance [24] for an introduction to inverse problems.

Regularization through variational analysis is a popular way to compute an approximation of x_0 from the measurements y as defined in (1). The general framework reads

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

This requires to define a prior R to enforce some regularity on the recovered signal. We restrict our attention in this paper to a ℓ^2 fidelity measure $\|y - \Phi x\|_2^2$ that reflects some Gaussian prior on the noise w. The regularization parameter $\lambda > 0$ should be adapted to match the noise level and the expected regularity of the data x_0.

For noiseless observations, $w = 0$, one has to take the limit $\lambda \to 0$ and solve 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 Hilbert norms of the form $R(x) = \langle x, Kx \rangle$ where K is some positive definite kernel. The minimizations (2) and (3) correspond to a Tikhonov regularization which typically enforces some kind of uniform smoothness in the recovered data. More advanced priors rely on non-quadratic functionals which enforce sparsity of the signal over some transformed domain (e.g. its wavelet transform or its gradient). These sparse priors are the subject of this article, and are described in the following section.

1.2 Notations

Our paper focus on real vector spaces. In all the following, the variable x will denote a vector in $\mathbb{R}^N$, y will be a vector in $\mathbb{R}^P$ and α a vector in $\mathbb{R}^N$.

The sign vector $\text{sign}(\alpha)$ of α is

$$\forall k \in \{1, \cdots, P\}, \quad \text{sign}(\alpha)_k = \begin{cases} +1 & \text{if} \quad \alpha_k > 0, \\ 0 & \text{if} \quad \alpha_k = 0, \\ -1 & \text{if} \quad \alpha_k < 0 \end{cases}.$$

The support of $\alpha \in \mathbb{R}^P$ is

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

In the following we make use of the matrix norms. The p,q-operator 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 reduced dimensional vector built upon the components of s indexed by J.

The matrix Id is the identity matrix, where the underlying space is implicited. For any matrix M, M^+ is the Moore–Penrose pseudoinverse of M.

1.3 Synthesis and Analysis Sparsity

Synthesis sparsity. Sparse regularization is a popular class of priors to model natural signals and images, see for instance [26]. In its simplest form, the sparsity of coefficients $\alpha \in \mathbb{R}^P$ is measured using the ℓ^0 pseudo-norm

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

Minimizing (2) or (3) with $R = R_0$ is however known to be in some sense NP-hard, see for instance [29]. Several workarounds have been proposed to alleviate this difficulty. A first class of methods uses greedy algorithms [30]. The most popular algortihms are Matching Pursuit [27] and Orthogonal Matching Pursuit [33, 13]. A second class of methods, which is the focus of this paper, replaces the ℓ^0 pseudo-norm by its ℓ^1 convex relaxation [15].

A dictionary $D = (d_i)_{i=1}^P$ is a (possibly redundant) 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 $x \in \operatorname{Span}(D) \subseteq \mathbb{R}^N$ as

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

In the redundant case ($P > N$) this decomposition is non-unique. The sparsest set of coefficients, according to the ℓ^1 norm, defines a prior

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

Any solution x of (2) using $R = R_A$ 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 $\Psi = \Phi D$, and $x = D\alpha$. It was first introduced in [39] in the statistical community and coined Lasso. It is also known in the signal processing community as Basis Pursuit Denoising [11]. Such problem corresponds to a so-called synthesis regularization because one assumes the sparsity of the coefficients α that synthesize the signal $x = D\alpha$. In the noiseless case, $w = 0$, one uses the constraint optimization (3), which reads

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

and is referred to as Basis Pursuit [11]. Taking $D = \mathrm{Id}$ to be the identity imposes sparsity of the signal itself, and is used for instance for sparse spikes deconvolution in seismic imaging [35]. Sparsity in orthogonal as well as redundant wavelet dictionaries are popular to model natural signals and images that exhibit sharp transitions [26]. Beside the regularization of inverse problems, a popular application of sparsity is blind source separation [44].

Analysis sparsity. *Analysis regularization* corresponds to using $R = R_A$ in (2) where

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

which leads to the following minimization problem

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

If

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

holds, this problem has a minimizer, and the set of minimizers is bounded by coercivity. All throughout this paper, we suppose that this condition holds. Note that the analysis problem ($\mathcal{P}_\lambda(y)$) is in some sense more general than the synthesis one (4) because the last one is recovered by setting $D = \mathrm{Id}$ and $\Psi = \Phi$.

In the noiseless case, $w = 0$, one uses the constrained optimization (3), which reads

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

The most popular analysis sparse regularization is the total variation, which was first introduced for denoising in [34]. It corresponds to using a derivative operator D^*. In the case of 1-D discrete signals, one can use forward finite differences $D = D_{\mathrm{DIF}}$ 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 review of total variation regularization can be found in [10].

The theoretical properties of total variation for denoising has been extensively studied. A distinctive feature of this regularization is that it tends to produces a staircasing effect, where discontinuities not present in the original data might be created by the regularization. This effect has been studied by Nikolova in [31] in 2-D. The stability of discontinuities for 2-D total variation denoising is the core of the work of [7]. Section 4.3 shows how our results also shed some light on this staircasing effect for 1-D signals.

It is also possible to use a dictionary D of translation invariant wavelets, so that the corresponding prior R_A can be interpreted as a sort of multi-scale total variation. Such a prior tends to favors piecewise regular signals and images. An extensive study of these redundant dictionaries highlighting differences between synthesis and analysis is done in [36].

As a last example of sparse analysis regularization, let us mention the Fused lasso [40], where D is the concatenation of a discrete derivative and a weighted identity. The corresponding prior R_A encourages both sparsity of the signal and its derivative, hence grouping block of non-zero coefficients together.

Synthesis versus analysis. In a synthesis prior, the generative vector α is sparse in the dictionary D whereas in analysis prior, the correlation between the signal x and the dictionary D is sparse. When D is orthogonal, $\mathcal{P}_\lambda(y)$ and Lasso define the same regularization. As highlighted in [19] synthesis and analysis regularizations however differ significantly when D is redundant. Some connections between total variation regularization and wavelet sparsity have been drawn in [37].

1.4 Union of Subspaces Model

Analysis regularization favors the sparsity of $D^* x$. It is thus natural to keep track of the support of this correlation vector, as done in the following definition.

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

A signal x such that $D^* x$ is sparse lives in a cospace $\mathcal{G}_J$ of small dimension where $\mathcal{G}_J$ is defined as follow.

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

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

where D_J *is the subdictionary whose columns are indexed by* J*.*

The signal space can thus be decomposed as a union of subspaces of increasing dimensions

$$\mathbb{R}^N = \bigcup_{k \in \{0,\dots,N\}} \Theta_k \quad \text{where} \quad \Theta_k = \{\mathcal{G}_J \setminus \dim \mathcal{G}_J = k\} . \tag{7}$$

The union of subspaces associated to synthesis regularization ($D = \mathrm{Id}$) defines Θ_k as the set of axis-aligned subspaces of dimension k. For the 1-D total variation prior, where $D = D_{\mathrm{DIF}}$ as defined in (6), Θ_k is the set of piecewise constant signals with $k-1$ steps. A detailed analysis of several sparse analysis subspaces, including translation invariant wavelets, can be found in [28].

More general unions of subspaces (not necessarily corresponding to analysis regularizations) have been introduced in sampling theory to model various kind of non-linear signal ensembles, see for instance [25]. Union of subspaces models have been extensively studied for the recovery from pointwise sampling measurements [25] and random measurements [20, 1, 2, 3].

1.5 Organization of this Paper

Section 2 details our six contributions. Section 3 draws some connexions with relevant previous works. Section 4 illustrates our results using concrete examples. Section 5 gives the proofs of the six contributions.

2 Contributions

This paper proves the following six results:

1. **Local affine parameterization w.r.t the observations:** a solution of $\mathcal{P}_\lambda(y)$ is a piecewise affine function of y.

2. **Local affine parameterization w.r.t the regularization parameter:** the mapping between λ and a solution of $\mathcal{P}_\lambda(y)$ is a piecewise linear path.

3. **Degrees of freedom:** the degrees of freedom (as defined in [18]) of the sparse analysis estimator is the dimension of $\mathcal{G}_J$.

4. **Robustness to small noise:** we give a sufficient condition on x_0 ensuring that the solution of $\mathcal{P}_\lambda(y)$ is close to x_0 when w is small enough.

5. **Noiseless identifiability:** the same condition ensures that x_0 is the unique solution of $\mathcal{P}_0(y)$ when $w = 0$.

6. **Robustness to bounded noise:** we give a sufficient condition on the D-cosupport of x_0 ensuring that the solution of $\mathcal{P}_\lambda(y)$ is close to x_0 when w is an arbitrary bounded noise and λ is large enough.

Each contribution is rigorously described in the following sub-sections.

Note that these contributions extend previously known results in the synthesis case, see for instance [17, 22, 43, 42, 6]. With the notable exception of the work of [28, 5] that studies analysis identifiability, to the best of our knowledge, it is the first time these questions are addressed in the analysis case.

2.1 Local Affine Parameterization w.r.t the Observations

Our first contribution gives a local affine parameterization of solutions of $\mathcal{P}_\lambda(y)$.

The transition space $\mathcal{Y}_\lambda$ defined below corresponds to observations y where the cospace $\mathcal{G}_J$ of the solution of $\mathcal{P}_\lambda(y)$ is not stable with respect to small perturbations of y.

Definition 3. *The* transition space $\mathcal{Y}_\lambda$ *is defined as*

$$\mathcal{Y}_\lambda = \left\{ y \in \mathbb{R}^Q \setminus \exists x \in \mathbb{R}^N : \min_{\sigma \in \Sigma_{y,\lambda}(x)} \|\sigma\|_\infty = 1 \right\},$$

where

$$\Sigma_{y,\lambda}(x) = \left\{ \sigma \in \mathbb{R}^{|J|} \setminus \Phi^*(\Phi x - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1 \right\}, \tag{8}$$

and $s = \operatorname{sign}(D^* x)$.

For some cosuport J, it is important to ensure the invertibility of Φ on $\mathcal{G}_J$. This is achieved by imposing

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

Note that there is always a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds as shown in Lemma 4.

Definition 4. *Let J be a D-cosupport. Suppose that (H_J) holds. We define the operator $A^{[J]}$ as*

$$A^{[J]} = U \left(U^* \Phi^* \Phi U \right)^{-1} U^*. \tag{9}$$

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

Except for observations in the transition space, the following theorem shows that a solution of $\mathcal{P}_\lambda(y)$ is a locally affine mapping of observations.

Theorem 1. *Let $y \notin \mathcal{Y}_\lambda$ and let $x^\star$ a solution of $\mathcal{P}_\lambda(y)$. Let I be the D-support and J the D-cosupport of $x^\star$ and $s = \operatorname{sign}(D^* x^\star)$. We suppose that (H_J) holds. We define*

$$\forall \bar{y} \in \mathbb{R}^P, \quad \hat{x}_\lambda(\bar{y}) = A^{[J]} \Phi^* \bar{y} - \lambda A^{[J]} D_I s_I.$$

There exists an open neighborhood $\mathcal{B} \subset \mathbb{R}^P$ of y such that $\hat{x}_\lambda(\bar{y}) \in \mathcal{B}$ is a solution of $\mathcal{P}_\lambda(\bar{y})$.

2.2 Local Affine Parameterization w.r.t the Regularization Parameter

For a given y, the λ-transition space Λ_y defined below corresponds to parameters λ where the cospace $\mathcal{G}_J$ of a solution of $\mathcal{P}_\lambda(y)$ is not stable with respect to small perturbations of λ.

Definition 5. *The λ-transition space Λ_y is defined as*

$$\Lambda_y = \left\{ \lambda \in \mathbb{R}_+^* \setminus \exists x \in \mathbb{R}^N : \min_{\sigma \in \Sigma_{y,\lambda}(x)} \|\sigma\|_\infty = 1 \right\},$$

where $\Sigma_{y,\lambda}$ is defined in (8)*.*

The following theorem proves that a solution of $\mathcal{P}_\lambda(y)$ is a piecewise affine function of $\lambda > 0$.

Theorem 2. *Let $y \in \mathbb{R}^P$, and let $\lambda \in \mathbb{R}_+^* \setminus \Lambda_y$. We denote $x^\star$ a solution of $\mathcal{P}_\lambda(y)$. Let I, J, s and $A^{[J]}$ be defined as in Theorem 1. We suppose that (H_J) holds. We define*

$$\forall \bar\lambda > 0, \quad \hat{x}_{\bar\lambda}(y) = A^{[J]} \Phi^* y - \bar\lambda A^{[J]} D_I s_I.$$

There exists an open neighborhood $\mathcal{C} \subset \mathbb{R}$ of λ such that $\hat{x}_{\bar\lambda}(y) \in \mathcal{C}$ is a solution of $\mathcal{P}_{\bar\lambda}(y)$.

If $\mathcal{P}_\lambda(y)$ admits a unique solution $x_\lambda(y)$ for each λ, this theorem shows that $\lambda \mapsto x_\lambda(y)$ is a polygonal path in $\mathbb{R}^N$. This results is already known in the synthesis case, see for instance [16, 32]. It also generalizes the work of [41] which studies the case of Φ overdetermined and develops an homotopy algorithm.

2.3 Degrees of Freedom

Degrees of freedom *df* is a familiar phrase in statistics. More generally, degrees of freedom is often used to quantify the complexity of a statistical modeling procedure. However, there is no exact correspondence between the degrees of freedom *df* and the number of parameters in the model. The concept of degrees of freedom plays an important role in model validation and selection, and its unbiased estimates provide unbiased estimates of the true risk, see e.g. [38].

Let the noise $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$, and therefore $y \sim \mathcal{N}(\mu_0 = \Phi x_0, \sigma^2 \mathrm{Id})$. We first notice that even if $\mathcal{P}_\lambda(y)$ admits several solutions, all of them share the same image under Φ, see Section 5.3 for proof of this point. Hence, we denote without ambiguity $\mu(y) = \Phi x^\star$ where $x^\star$ is a solution of $\mathcal{P}_\lambda(y)$.

From the seminal definition of Efron [18], and by the Stein Lemma [38], the degrees of freedom of a weakly differentiable estimator $y \mapsto \mu(y)$ is

$$df(\mu_0) = \mathbb{E}_w\left(\mathrm{div}(\mu(y))\right) = \sum_{i=1}^{Q} \mathbb{E}_w\left(\frac{\partial \mu(y)}{\partial y_i}\right).$$

We have the following result for the analysis regularization.

Theorem 3. *The mapping $y \mapsto \mu(y)$ is of class C^∞ on $\mathbb{R}^N \setminus \mathcal{Y}_\lambda$. For $y \notin \mathcal{Y}_\lambda$, there exists $x^\star$ a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds with J the D-cosupport of $x^\star$, and the associated degree of freedom is*

$$\mathrm{div}(\mu(y)) = \dim(\mathcal{G}_J). \tag{10}$$

This result is known to hold in the special case of synthesis regularization ($D = \mathrm{Id}$). It is proved in the overdetermined case in [45] and is extended to the general case in [23]. In the overdetermined case with an analysis regularization, [41] also proved a similar expression but with a completely different $\mathcal{Y}_\lambda$. Our expression handles both over- and underdetermined measurements with a sharper characterization of the set $\mathcal{Y}_\lambda$.

From (10), it is tempting to use $\dim(\mathcal{G}_J)$ as an estimator of degrees of freedom $df(\mu_0)$. However, we have no guarantee at this stage that this would be an unbiased estimator of $df(\mu_0)$. Indeed, for this to hold true, the set $\mathcal{Y}_\lambda$ should be of (Lebesgue) measure zero, which we have not established yet. This is an open question that we leave for a future research.

2.4 Robustness to Small Noise

Our next contribution shows that analysis regularization is robust to a small noise under a condition on $\mathrm{sign}(D^* x_0)$.

We now define our sign criterion.

Definition 6. *Let $s \in \{-1, 0, +1\}^P$, I its D-support and J its D-cosupport. We suppose (H_J) holds. The analysis* Identifiabiltiy Criterion **IC** *of s is defined as*

$$\mathbf{IC}(s) = \min_{u \in \mathrm{Ker}\, D_J} \|\Omega s_I - u\|_\infty \quad \textit{where} \quad \Omega = D_J^+(\Phi^*\Phi A^{[J]} - \mathrm{Id}) D_I.$$

We have the following theorem.

Theorem 4. *Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J, and of D-support $I = J^c$. Suppose (H_J) holds and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$. It exists two constants $c_J > 0$ and $\tilde{c}_J > 0$, such that if $y = \Phi x_0 + w$, where*

$$\|w\|_2 < \frac{\tilde{c}_J}{c_J},$$

and if λ satisfies

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

the vector defined by

$$\hat{x}^\star = x_0 + A^{[J]}\Phi^* w - \lambda A^{[J]} D_I s_I, \tag{11}$$

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

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

Note that it is possible to choose λ proportional to the noise level $\|w\|_2$. Hence, for $\|w\|_2$ small enough, equation (11) gives

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

2.5 Noiseless Identifiability

In the noiseless case, $w = 0$, the criterion **IC** can be used to test identifiability. Recall that

Definition 7. *A vector x_0 is said to be* identifiable *if x_0 is the unique solution of $\mathcal{P}_0(\Phi x_0)$.*

We prove the following theorem

Theorem 5. *Let $x_0 \in \mathbb{R}^N$ be a fixed vector of D-cosupport J. Suppose that (H_J) holds and $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$. Then x_0 is identifiable.*

2.6 Robustness to Bounded Noise

Our last contribution defines a stronger criterion that ensures robustness to an arbitrary bounded noise.

Definition 8. *The analysis* Recovery Criterion (RC) *of $I \subset \{1 \dots P\}$ is defined as*

$$\mathbf{RC}(I) = \max_{x \in \mathcal{G}_J} \mathbf{IC}\left(\text{sign}(D^* x)\right).$$

Note that if I is the D-support of x_0, $\mathbf{RC}(I) < 1$ implies $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$.

The following theorem shows that if the parameter λ is big enough, then $\mathcal{P}_\lambda(y)$ recovers a unique vector which is close enough in the ℓ^2 sense and lives in the same $\mathcal{G}_J$ as the unknown signal x_0.

Theorem 6. *Let I be a fixed D-support and J its associated D-cosupport $J = I^c$. Suppose that (H_J) holds. If $\mathbf{RC}(I) < 1$ and*

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

where c_J is defined as,

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

then for every x_0 of D-support I, it exists a unique solution $x^\star$ of D-support included in I, verifying $\|x_0 - x^\star\|_2 = O(\|w\|_2)$. More precisely,

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

3 Related Works

3.1 Previous Works on Synthesis Identifiability and Robustness

Several previous works have studied identifiability and noise robustness of sparse synthesis regularization. We recall that synthesis regularization (4) reads

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

where $\Psi = \Phi D$, and $x = D\alpha$. Fuchs defines [22] a criterion $\mathbf{IC}_S$ which is a specialization of our criterion **IC** introduced in Definition 6 to the case where $D = \mathrm{Id}$.

Definition 9. *Given some support I and cosupport J. Given some sign vector $s \in \{-1, +1\}^P$, the Sign Criterion $\mathbf{IC}_S$ of a sign vector s associated to a support I is defined as*

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

Fuchs shows the following result

Theorem ([22])**.** *Let $\alpha_0 \in \mathbb{R}^P$ be a fixed vector of support I. If Ψ_I is of 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 $y = \Psi\alpha_0$.*

The work of Tropp [43, 42] developed in the synthesis case a condition named Exact Recovery Condition (ERC) on the support.

Definition 10. *The* Exact Recovery Condition *(ERC) of $I \subset \{1 \dots P\}$ is defined as*

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

Tropp proves that $\mathbf{ERC}(I) < 1$ is a sufficient condition of identifiability and stability of the synthesis Lasso.

Theorem ([43])**.** *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, it exists a unique solution $\alpha^\star$ of* (4) *for $y = \Psi\alpha_0 + w$ of support included in I, verifying $\|\alpha_0 - \alpha^\star\|_2 = O(\|w\|_2)$.*

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

In the analysis case where $D = \mathrm{Id}$, the criterion of Tropp and our are equivalent. This is also true for the criterion of Fuchs and our.

Proposition 1. *If $D = \mathrm{Id}$, then $\mathbf{ERC}(I) = \mathbf{RC}(I)$ and $\mathbf{IC}(\mathrm{sign}(D^* x_0)) = \mathbf{IC}_S(\mathrm{sign}(D^* x_0))$.*

Let us mention that there exist several other criteria ensuring both identifiability and noise robustness in the synthesis cases. This includes criteria based on coherence (see [4] for a review) and RIP-based compressed sensing theory that requires that Φ is a realization of certain random matrices ensembles [6, 14].

3.2 Previous Works on Analysis Identifiability and Robustness

To the best of our knowledge, the only previous works that study the performance of sparse analysis regularization are the papers [5] and [28].

The work [5] proves a strong robustness with overwhelming probability to noise when D is tight frame and Φ a realization of certain random matrices ensembles satisfying a condition named D-RIP. This setting is thus quite far from our.

The work of Nam and al. is much closer to our results. It studies noiseless identifiability using ℓ^0 and ℓ^1 sparse analysis regularization. Their main result on ℓ^1 analysis identifiability is the following theorem.

Theorem ([28])**.** *Let M^* be a basis matrix of* $\mathrm{Ker}\,\Phi$ *and I a fixed D-support such that the matrix $D_J^* M^*$ has full rank. Let $x_0 \in \mathcal{G}_J$ be a fixed vector. If* $\mathbf{IC}_0(\mathrm{sign}(D^* x_0)) < 1$ *and*

$$\mathbf{IC}_0(s) = \|\daleth_I s_I\|_\infty \quad \textit{where} \quad \daleth_I = (MD_J)^+ MD_I,$$

then x_0 is identifiable.

Note that $\mathbf{IC}_0(s) < 1$ does not imply $\mathbf{IC}(s) < 1$ neither the opposite. Numerical results suggest that their criterion is most of the time sharper than RC. However, the condition $\mathbf{IC}_0(s) < 1$ does not imply in general a robustness to noise, even for a small one. Moreover, let x_0 be a fixed vector, and denote $s = \mathrm{sign}(D^* x_0)$ where I is its D-support and $y = \Phi x_0 + w$. If $\mathbf{IC}_0(s) < 1$ but $\mathbf{IC}(s) > 1$, then any solution $x^\star$ of $\mathcal{P}_\lambda(y)$, for λ close to zero, is such that the D-support of $x_\lambda(y)$ is not included in I. One can thus find vectors x_0 with $\mathbf{IC}_0(s) < 1$ but where $\frac{\|x_0 - x^\star\|_2}{\|w\|_2}$ is arbitrary large, whatever the amplitude $\|w\|_2$ of the noise.

4 Examples

This section provides algorithms to compute identifiability criteria IC and RC, along with an analysis of total variation and Fused Lasso denoising.

4.1 Computing Sparse Analysis Regularization

It is not the focus of this paper to give a full study of optimization schemes that can be used to solve the analysis regularization.

In the case where $\Phi = \mathrm{Id}$ (denoising), $\mathcal{P}_\lambda(y)$ is strictly convex, and one can compute its unique solution $x^\star$ by solving an equivalent dual problem [8]

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

In the general case, it is possible to use a primal-dual method such as the algorithm of Chambolle and Pock [9]. One way is to rewrite the optimization problem as follow

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

4.2 Computing the Criteria

In the case where $\mathrm{Ker}(D_J) \neq \{0\}$, computing $\mathbf{IC}(\mathrm{sign}(D^* x_0))$ necessitates the resolution of a convex problem. This optimization is re-written as

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

where $\iota_{\mathrm{Ker}(D_J)}$ is characteristic function of $\mathrm{Ker}(D_J)$

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

This requires the optimization of a sum of two simple functions, i.e. function whose proximal operators is easy to compute. The proximal operator Prox_f of a convex lower semicontinuous function f is defined as

$$\forall x \in \mathbb{R}^N, \quad \mathrm{Prox}_f(x) = \underset{z \in \mathbb{R}^N}{\mathrm{argmin}} \; \frac{1}{2}\|z - x\|_2^2 + f(z).$$

Such a minimization can hence be achieved using the Douglas-Rachford splitting algorithm [12]. Indeed, the proximity operator of $\iota_{\mathrm{Ker}(D_J)}$ is the orthogonal projector on $\mathrm{Ker}(D_J)$, and the proximal operator of $\|\cdot\|_\infty$ can be computed as

$$\mathrm{Prox}_{\gamma\|\cdot\|_\infty}(x) = x - P_{\|\cdot\|_1}\left(\frac{x}{\gamma}\right),$$

where $P_{\|\cdot\|_1}$ is the projection of the ℓ^1 ball $\left\{x \in \mathbb{R}^N \setminus \|x\|_1 \leqslant 1\right\}$. This projection is computed as explained for instance in [21].

Unfortunately, computing $\mathbf{RC}$ necessitates to solve a combinatorial optimization problem which is non convex. Recall that

$$\mathbf{RC}(I) = \max_{x \in \mathcal{G}_J} \mathbf{IC}\left(\mathrm{sign}(D^* x)\right).$$

It is possible to define a stronger criterion by extending the maximum to the set of all possible sign vectors.

$$\mathrm{s}\mathbf{RC}(I) = \max_{\substack{s \in \{-1,0,+1\}^P, \\ \mathrm{supp}(s) = I}} \mathbf{IC}(s).$$

This criterion however still necessitates the resolution of a combinatorial optimization problem. An even stronger criterion, which is easy to compute, is obtained by using $u = 0$ in $\mathrm{Ker}\, D_J$ for the computation of each $\mathbf{IC}(s)$.

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

Note that for every vector x_0 with D-support $I = \mathrm{supp}(D^* x)$, we have the following inequalities

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

4.3 Total Variation Denoising

Discrete total variation uses $D = D_{\text{DIF}}$ defined in (6). We recall that total variation model is formed by $\bigcup_k \Theta_k$ where Θ_k is the set of piecewise constant signals with $k-1$ steps. We define the following object

Definition 11. *A signal is said to* contain a staircase subsignal *if it exists* $i \in \{1 \dots |I|-1\}$ *such that*

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

Figure 1 shows examples of signals with and without staircase sub-signals with their dual vectors.

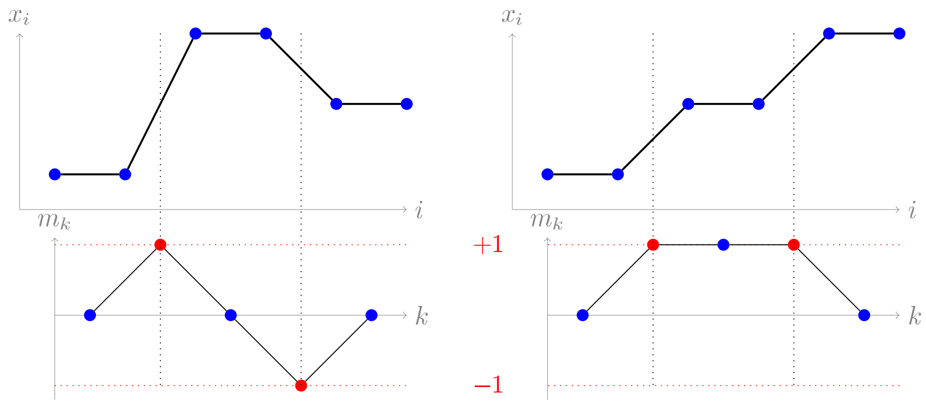

Figure 1: Top line: Signals x with 2 discontinuities. Bottom line: Associated dual vector m.

The following proposition studies the robustness of total variation denoising

Proposition 2. *We consider the denoising case,* $\Phi = \text{Id}$*. If* x *does not contain a staircase subsignal, then* $\mathbf{IC}(\operatorname{sign}(D^* x)) < 1$*. Otherwise,* $\mathbf{IC}(\operatorname{sign}(D^* x)) = 1$*.*

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

$$m : \begin{cases} m_I = s_I = \operatorname{sign}(D^* x)_I \\ m_J = \sigma = \Omega s_I \end{cases} .$$

The vector σ satisfies $(D_J^* D_J)\sigma = (D_J^* D_I) s_I$. One can shows 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 indexes of I, m is obtained by linearly interpolating (see Figure 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 does not contain a staircase subsignal, one has $\|\Omega s_I\|_\infty < 1$. On the contrary, if there is i_1 such that $s_{i_1} = s_{i_2}$, where i_1, i_2 are consecutive indexes of I, then for every $i_1 < j < i_2, m_j = s_{i_1} = \pm 1$ which implies $\mathbf{IC}(\text{sign}(D^* x)) = 1$. □

This proposition together with Theorem 4 shows that if a signal does not have a staircase sub-signal, TV denoising is robust to a small noise. This means that if w is small enough, for λ small, the TV denoising of $x_0 + w$ has the same discontinuities as x_0. However, the presence of a staircase in a signal implies that no robustness, even for a small one can be ensured.

Corollary 1. *If $|I| > 2$ such that $i \in I$ implies $i+1 \notin I$, and $D = D_{DIF}$, then* $\mathbf{RC}(I) = 1$.

Proof. If $|I| > 2$, it exists a signal $\tilde{x}$ which contain a staircase subsignal, hence $1 = \mathbf{IC}(\text{sign}(D^* \tilde{x})) \leqslant \mathbf{RC}(I)$. Since no signal is such that $\mathbf{IC}(\text{sign}(D^* x)) > 1$, we conclude. □

This corollary shows that in the case of total variation regularization, one cannot expect cospace robustness, i.e discontinuities conservation, for any class of bounded noise.

4.4 Fused Lasso

Fused Lasso is introduced in [40] as the following minimization problem

$$\min_{x \in \mathbb{R}^N} \|y - \Phi x\|_2^2 \quad \text{subject to} \quad \|x\|_1 \leqslant s_1 \text{ and } \|D_{\text{DIF}}^* x\|_1 \leqslant s_2, \tag{12}$$

where s_1, s_2 are two positive constants. The problem (12) is equivalent to $\mathcal{P}_\lambda(y)$ for

$$D = \begin{bmatrix} D_{\text{DIF}} & \varepsilon \text{Id} \end{bmatrix},$$

where ε is a function of s_1 and s_2. The associated union of subspaces (7) is $\bigcup_k \Theta_k$ where Θ_k is the set of k-sum of interval indicator signals, 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 $a_i < b_i < a_{i+1}$.

Definition 12. *The* minimum separation distance *of x is defined as*

$$d(x) = \min_{1 \leqslant i < k} |a_{i+1} - b_i|$$

Empirical observation. *Suppose* $\Phi = \mathrm{Id}$. *Let* $x_0 \in \Theta_k$. *If* $d(x_0) > \xi(\varepsilon)$ *where* ξ *is a decreasing real function with* $\lim_{\varepsilon \to 0} \xi(\varepsilon) = 0$, *then* $\mathbf{RC}(I) = \mathbf{IC}(\mathrm{sign}(D^* x_0)) < 1$, *where* J *is the* D*-support.*

This suggests that the Fused Lasso enables a stable recovery of sums of indicators of disjoint sets if these sets are separated enough.

5 Proofs

This section is dedicated to proofs of theorems 1 – 6. The objective function $\mathcal{L}_{y,\lambda}$ minimized in $\mathcal{P}_\lambda(y)$ is

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

The following lemma, which is at the heart of the proofs of our contributions, details 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*

$$\sigma \in \Sigma_{y,\lambda}(x^\star) \tag{13}$$

where $\Sigma_{y,\lambda}$ *is defined in* (8). *Moreover, if* $\|\sigma\|_\infty < 1$ *and* (H_J) *holds, then* $x^\star$ *is the unique solution of* $\mathcal{P}_\lambda(y)$.

Proof. The subdifferential ∂F of a real valued convex lower semicontinuous function $F : \mathbb{R}^N \to \mathbb{R}$ is the multifunction defined by

$$\partial F(x_0) = \left\{ g \in \mathbb{R}^N \setminus \forall x \in \mathbb{R}^N, \quad f(x) \geqslant f(x_0) + \langle g,\, x - x_0 \rangle \right\}$$

Note that x_0 is a minimum of F if, and only if, $0 \in \partial F(x_0)$. The subdifferential of $\mathcal{L}_{y,\lambda}(x)$ is

$$\partial \mathcal{L}_{y,\lambda}(x) = \left\{ \Phi^*(\Phi x - y) + \lambda D u \setminus u \in \mathbb{R}^N : u_I = \mathrm{sign}(D^* x)_I \text{ and } \|u_J\|_\infty \leqslant 1 \right\}.$$

Hence $0 \in \partial \mathcal{L}_{y,\lambda}(x)$ is equivalent to the existence of $u \in \mathbb{R}^N$ such that $u_I = \mathrm{sign}(D^* x)_I$ and $\|u_J\|_\infty \leqslant 1$ satisfiyng

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

Decomposing on the D-cosupport J, it is equivalent to the existence of $\sigma \in \Sigma_{y,\lambda}(x)$ with $\|\sigma\|_\infty \leqslant 1$.

Let $x^\star$ be a solution of $\mathcal{P}_\lambda(y)$. There exists $\sigma \in \Sigma_{y,\lambda}(x^\star)$ with $\|\sigma\|_\infty \leqslant 1$ Suppose also that $\|\sigma\|_\infty < 1$. We decompose $\mathcal{L}_{y,\lambda}$ in two functions :

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

Since $\|\sigma\|_\infty < 1$, one can prove there exists $\varepsilon > 0$ such that

$$\mathrm{Im} D_J \cap B(0, \varepsilon) \subseteq \partial \mathcal{L}_{y,\lambda}(x^\star).$$

Hence $\partial \mathcal{L}_{y,\lambda}(x^\star) \subseteq \mathrm{Im} D_J$. Indeed it cannot be in the affine plane, otherwise $0 \notin \partial \mathcal{L}_{y,\lambda}(x)$ which is a contradiction with $x^\star$ minimizer of $\mathcal{P}_\lambda(y)$. Hence, its normal cone, defined by

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

is $(\mathrm{Im} D_J)^\perp = \mathcal{G}_J$. Let $h \in \mathbb{R}^N \setminus \{0\}$. Two different cases occur

1. If $h \notin \mathcal{G}_J$, then there exists $d \in \partial \mathcal{L}_{y,\lambda}(x)$ such that $\langle d,\, h \rangle > 0$.

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

2. If $h \in \mathcal{G}_J$, observe that q is coercive on $\mathcal{G}_J$ since (H_J) holds. Hence,

$$\mathcal{L}_{y,\lambda}(x^\star + h) > q(x^\star) + \langle \nabla q(x^\star),\, h \rangle + \lambda \|x^\star\|_1 + \lambda \langle v,\, h \rangle.$$

where $v \in \partial_{\|D^* \cdot\|_\infty}(x^\star)$ such that $\lambda v + \nabla q(x^\star) = 0$. Then,

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

In summary, for every $h \in \mathbb{R}^N \setminus \{0\}$, $\mathcal{L}_{y,\lambda}(x^\star + h) > \mathcal{L}_{y,\lambda}(x^\star)$, and $x^\star$ is the unique minimizer of $\mathcal{P}_\lambda(y)$. □

We recall that we suppose condition (H_0) holds in every statements.

5.1 Proof of Theorem 1

The proof of Theorem 1 is done in two steps. First, we proves Temma 2 which gives an implicit solution of $\mathcal{P}_\lambda(y)$. Then, we proves Theorem 1.

The following lemma gives an implicit expression for a solution $x^\star$ of the problem $\mathcal{P}_\lambda(y)$. Note that $\mathcal{P}_\lambda(y)$ may have other solutions.

Lemma 2. *Let $x^\star$ a solution of $\mathcal{P}_\lambda(y)$. Let I be the D-support and J 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. Using the first order 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, one has $x^\star \in \mathcal{G}_J$ so $x^\star \in \mathrm{Im} D_J$. Hence, we can write $x^\star = Uz$. Since $U^* D_J = 0$, multiplying equation (15) on the left by U^*, we get

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

Since $U^* \Phi^* \Phi U$ is invertible, we conclude. □

We can now prove theorem 1.

Proof of Theorem 1. Let $y \notin \mathcal{Y}_\lambda$ and let $\sigma \in \Sigma_{y,\lambda}(x_\lambda(y))$ such that $\|\sigma\|_\infty < 1$. By construction of $\hat{x}_\lambda(\bar{y})$ one has $D_J^* \hat{x}_\lambda(\bar{y}) = 0$. So for $\bar{y}$ close enough from y, one has

$$\operatorname{sign}(D^* \hat{x}_\lambda(\bar{y})) = \operatorname{sign}(D^* x^\star).$$

One has that $\hat{x}_\lambda(\bar{y})$ is the solution of $\mathcal{P}_\lambda(\bar{y})$ if and only if there exists $\bar{\sigma} \in \Sigma_{y,\lambda}(\hat{x}_\lambda(\bar{y}))$. Using the property $\sigma \in \Sigma_{y,\lambda}(x_\lambda(y))$ and the expression (14), this is equivalent to requiring that

$$\Phi^*(\Phi A^{[J]} \Phi^* - \mathrm{Id})(\bar{y} - y) = -\lambda D_J(\sigma - \bar{\sigma}). \tag{16}$$

Since one has

$$U^* \Phi^*(\Phi A^{[J]} \Phi^* - \mathrm{Id}) = 0,$$

and since U is an orthogonal basis of $\mathrm{Im}(D_J)^\perp$, there exists a matrix B such that

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

Using the particular choice

$$\bar{\sigma} = \sigma + B(\bar{y} - y)$$

ensures that (16) is satisfied. Since $\|\sigma\|_\infty < 1$ and since $\sigma \mapsto \sigma + B(\bar{y} - y)$ is continuous, imposing $\bar{y}$ close enough to y ensures that $\|\bar{\sigma}\| \leqslant 1$. □

5.2 Proof of Theorem 2

The proof of Theorem 2 is similar to the proof of Theorem 1.

Proof of Theorem 2. Let $y \in \mathbb{R}^P$ and $\lambda \in \mathbb{R}_+^* \setminus \Delta_y$. Note that $x^\star$ is well defined according to Lemma 2. Since $\lambda \notin \Delta_y$, there exists $\sigma \in \Sigma_{y,\lambda}(x^\star)$ such that $\|\sigma\|_\infty < 1$. By construction of $\hat{x}_{\bar{\lambda}}(y)$ one has $D_J^* \hat{x}_{\bar{\lambda}}(y) = 0$. So for $\bar{\lambda}$ close enough from λ, one has

$$\operatorname{sign}(D^* \hat{x}_{\bar{\lambda}}(y)) = \operatorname{sign}(D^* x^\star) \overset{\text{def.}}{=} s.$$

Similarly to the proof of Theorem 1, $\hat{x}_{\bar{\lambda}}(y)$ is thus solution if and only if there exists $\bar{\sigma}$ with

$$\lambda D_I s_I + \lambda D_J \sigma = \bar{\lambda} D_I s_I + \bar{\lambda} D_J \bar{\sigma} \quad \text{and} \quad \|\bar{\sigma}\|_\infty \leqslant 1.$$

This is achievied for

$$\bar{\sigma} = \frac{\lambda}{\bar{\lambda}} \sigma - \frac{\lambda - \bar{\lambda}}{\bar{\lambda}} D_J^+ D_I s_I.$$

□

5.3 Proof of Theorem 3

The proof is done in four steps. First, we prove that $\mu(y)$ is well-defined. Then, we prove that there exists a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds. Finally, we prove that $df(\mu(y)) = \dim \mathcal{G}_J$.

We first proves that even if $\mathcal{P}_\lambda(y)$ admits several solutions, all of them share the same image under Φ.

Lemma 3. *If x_1 and x_2 are two solutions of $\mathcal{P}_\lambda(y)$, then $\Phi x_1 = \Phi x_2$.*

Proof. Let x_1, x_2 be two solutions of $\mathcal{P}_\lambda(y)$ and $\Phi x_1 \neq \Phi x_2$. We define $x_3 = \frac{1}{2}(x_1 + x_2)$. Since the function $u \mapsto \|y - u\|^2$ is strictly convex, one has the following inequality

$$\frac{1}{2}\|y - \Phi x_3\|^2 < \frac{1}{2}\left(\frac{1}{2}\|y - \Phi x_1\|^2 + \frac{1}{2}\|y - \Phi x_2\|^2\right).$$

Applying triangle inequality for the ℓ^1 norm gives

$$\|D^* x_3\|_1 \leqslant \|D^* x_1\|_1 + \|D^* x_2\|_1.$$

Hence, $\mathcal{L}_{y,\lambda}(x_3) < \mathcal{L}_{y,\lambda}(x_1)$ which is a contradiction with x_1 being a solution of the problem $\mathcal{P}_\lambda(y)$. □

Lemma 4. *There exists $x^\star$ a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds, where J is the D-cosupport of $x^\star$.*

Proof. Let $x^\star$ be a solution of $\mathcal{P}_\lambda(y)$. Suppose (H_J) does not hold. Then, there exists $z \in \operatorname{Ker} \Phi$ with $z \neq 0$ and $D_J^* z = 0$. We define $v_t = x^\star + tz$. For t small enough, let's say $t \leqslant t_0$

$$\operatorname{sign}(D^* v_t) = \operatorname{sign}(D^* x^\star).$$

Hence, the mapping $t \mapsto \|D^* v_t\|_1$ is locally affine until at least one of the component of $D^* v_t$ vanishes.

Two cases occur :

1. If for every $0 \leqslant t \leqslant t_0$, no component of $D^* v_t$ vanishes, then

   $$\forall t \leqslant t_0, \quad \|D^* v_t\|_1 = t\|D^* z\|_1 + \|D^* x^\star\|_1 \quad \text{where} \quad a > 0.$$

   Applying Minkowski inequality equality case, there exists some $\rho \in \mathbb{R}$

   $$\forall t \leqslant t_0, \quad t\|D^* z\|_1 = \rho\|D^* x^\star\|_1$$

   Thus, $D^* z = 0$. Then $z \in \operatorname{Ker} \Phi \cap \operatorname{Ker} D^*$ and $z \neq 0$, which is a contradiction of condition (H_0) .

2. Else, for some $0 < t_1 \leqslant t_0$, a component of $D^* v_t$ vanishes. Then, the D-cosupport of v_{t_1} is strictly included in J and $\mathcal{L}_{y,\lambda}(v_{t_1})$ is minimum. Iterating this argument shows that there exists a solution with D-cosupport strictly included in J such that (H_J) holds.

□

Next, we prove the theorem 3.

Proof of Theorem 3. Using Lemma 4, there exists a solution $x^\star$ of $\mathcal{P}_\lambda(y)$ such that (H_J) holds. We consider this solution. Since $\mu(\bar{y}) = \Phi x_\lambda(\bar{y})$, using Lemma 2 for $\bar{y}$ close enough from y, one has

$$\mu(\bar{y}) = \Phi A^{[J]}\Phi^*\bar{y} - \lambda \Phi A^{[J]} D_I s_I.$$

where J is the D-cosupport of $x^\star$. Remark that $\mu(\bar{y})$ can we written as $\mu(\bar{y}) = V\bar{y} + r$ where $V = \Phi A^{[J]}\Phi^*$ and $r \in \mathbb{R}^P$ is a constant vector. Hence,

$$\operatorname{div}(\mu(y)) = \operatorname{tr}(V)$$

Remark V is the orthogonal projector on $\operatorname{Im}(V) = \ker(V)^\perp$, so that $\operatorname{div}(\mu(y)) = \dim(\operatorname{Im}(V))$. Since Φ is injective on $\mathcal{G}_J$, one has $\dim(\operatorname{Im}(V)) = \dim(\mathcal{G}_J)$. □

5.4 Proof of Theorem 4

To prove Theorem 4, we start with the following lemma which gives a necessary and sufficient condition on $\hat{x}^\star$ to be the unique solution of $\mathcal{P}_\lambda(y)$.

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

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

where $s = \operatorname{sign}(D^*\hat{x}^\star)$. If there exists σ satisfying

$$\sigma - \Omega s + \frac{1}{\lambda}\tilde{B}y \in \operatorname{Ker} D_J \quad \textit{and} \quad \|\sigma\|_\infty < 1, \tag{17}$$

where $\tilde{B} = D_J^+\Phi^*(\Phi A^{[J]}\Phi^* - \operatorname{Id})$, then $\hat{x}^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$

Proof. According to Lemma 1, $\hat{x}^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$ if there exists $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$ such that

$$\Phi^*(\Phi\hat{x}^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty < 1.$$

Hence, with few algebraic manipulations, one has

$$By - \lambda C D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty < 1, \tag{18}$$

where $C = \Phi^*\Phi A^{[J]} - \operatorname{Id}$. One has $U^* C = 0$ and thus one can find $\tilde{C}$ such that $C = D_J \tilde{C}$. For instance we select $\tilde{C} = D_J^+ C = \Omega$. Similarly, we define $\tilde{B}$ such that $\tilde{B} = D_J^+ B$. Hence, the existence of $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$ such that $\|\sigma\|_\infty < 1$ is equivalent to

$$D_J \sigma = D_J \Omega s_I - \frac{1}{\lambda} D_J \tilde{B} y \quad \text{where} \quad \|\sigma\|_\infty < 1,$$

which in turn is equivalent to

$$\sigma - \Omega s_I + \frac{1}{\lambda}\tilde{B}y \in \operatorname{Ker} D_J \quad \text{where} \quad \|\sigma\|_\infty < 1,$$

□

We recall that, according to Definition 6, given some D-support I and D-cosupport $J = I^c$, we suppose that condition (H_J) holds. Given some sign vector $s \in \{-1, +1\}^P$, the analysis Identifiabiltiy Criterion **IC** of a sign vector s associated to a D-support I is defined as

$$\mathbf{IC}(s) = \min_{u \in \operatorname{Ker} D_J} \|\Omega s_I - u\|_\infty \quad \text{where} \quad \Omega = D_J^+(\Phi^*\Phi A^{[J]} - \mathrm{Id})D_I.$$

Proof of Theorem 4. The proof is done in three steps. First, we give a condition on λ to have signs equality between the proposed solution and x_0. Then, we give an other condition on $\frac{\|w\|_2}{\lambda}$ to ensure first-order condition on $\hat{x}^\star$ assuming $\mathbf{IC} < 1$. Finally, we prove that both condition are compatible.

We first give a condition on λ to ensure signs equality

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

Since $A^{[J]}\Phi^*y = x_0 + A^{[J]}\Phi^*w$, signs equality is achieved if

$$\forall i \in I, \quad |D_I^* x_0|_i > |D_I^*(\hat{x}^\star - x_0)|_i = |D_I^* A^{[J]}\Phi^* w - \lambda D_I^* A^{[J]} D_I s_I|_i. \tag{19}$$

We bound the term $\|D_I^*(\hat{x}^\star - x_0)\|_\infty$

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

Using operator norm inequalities, one has

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

Introducing

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

the following condition

$$\lambda < \frac{\|D_I^* A^{[J]}\|_{\infty,\infty} \|\Phi^*\|_{2,\infty}}{T} \frac{\|w\|_2}{\lambda} + \frac{\|D_I^* A^{[J]}\|_{\infty,\infty} \|D_I\|_{\infty,\infty}}{T}, \tag{20}$$

ensures (19), and so signs equality.

We give now a condition on $\frac{\|w\|_2}{\lambda}$ to ensure first-order condition assuming $\mathbf{IC}(\operatorname{sign}(D^*x_0)) < 1$. Remark that $\tilde{B}y = \tilde{B}w$ since $x_0 \in \mathcal{G}_J$. Using Lemma 5, the vector $\hat{x}^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$ if there exists σ satisfying

$$\hat{\sigma} = -u + \Omega s_I - \frac{1}{\lambda}\tilde{B}w \quad \text{and} \quad \|\sigma\|_\infty < 1,$$

for some $u \in \ker(D_J)$. Hence, under condition $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$, one has

$$\|\tilde{B}\|_{2,\infty} \frac{\|w\|_2}{\lambda} < 1 - \mathbf{IC}(\text{sign}(D^* x_0)). \tag{21}$$

implies that there exists $\hat{\sigma} \in \Sigma_{y,\lambda}(\hat{x}^\star)$ with $\|\hat{\sigma}\|_\infty < 1$.

Let show that (20) and (21) are compatible. We introduce constants c_J and $\tilde{c}_J$:

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

and

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

Suppose that

$$\|w\|_2 < \frac{\tilde{c}_J}{c_J},$$

and

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

Then (20) and (21) are satisfied. □

5.5 Proof of Theorem 5

The proof of Theorem 5 is done in three steps. First, we specializes Theorem 4 when $w = 0$. Then, we show that under the condition $\mathbf{IC}(\text{sign}(D^* x_0))$, the vector x_0 is a solution of $\mathcal{P}_0(y)$. Finally, we prove Theorem 5 by considering an other potential solution 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 (H_J) holds and $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$. Then for $\lambda < \tilde{c}_J$,*

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

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

Proof. Take $w = 0$ in theorem 4. □

Lemma 6. *Let $x_0 \in \mathbb{R}^N$ be a fixed vector, I be its D-support, and $y = \Phi x_0$. Suppose (H_J) holds and $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$. 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 $\lambda < \tilde{c}_J$,

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

Let $x_1 \neq x_0$ such that $\Phi x_1 = y$. For every λ strictly positive, one has $\mathcal{L}_{y,\lambda}(x^\star) < \mathcal{L}_{y,\lambda}(x_1)$ by definition of x_λ. Then,

$$\|D^* x_\lambda\|_1 < \|D^* x_1\|_1.$$

Using continuity of norms, taking the limit $\lambda \to 0$ in this equation gives

$$\|D^* x_0\|_1 \leqslant \|D^* x_1\|_1,$$

which proves that x_0 is a solution of $\mathcal{P}_0(y)$. □

Proof of Theorem 5. Using Lemma 6, x_0 is a solution of $\mathcal{P}_0(y)$. We shall prove that x_0 is the unique solution. Let denote

$$x_1 = x_0 + \lambda A^{[J]} D_I s_I.$$

Note that for λ small enough and $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$, one has $\text{sign}(D^* x_1) = \text{sign}(D^* x_0)$. Hence, since Corrolary 2 holds, x_0 is the unique solution of $\mathcal{P}_\lambda(y_1)$ where $y_1 = \Phi x_1$.

Let $x_2 \in \mathbb{R}^N$ such that $\Phi x_2 = y$ with $x_2 \neq x_0$. Then $\Phi x_0 = \Phi x_2$ and since x_0 is the unique solution of $\mathcal{P}_\lambda(y_1)$, one has

$$\frac{1}{2}\|y - \Phi x_0\|_2^2 + \lambda \|D^* x_0\|_1 < \frac{1}{2}\|y - \Phi x_2\|_2^2 + \lambda \|D^* x_2\|_1.$$

Then,

$$\|D^* x_0\|_1 < \|D^* x_2\|_1,$$

which gives uniqueness of the solution. □

5.6 Proof of Theorem 6

We split the proof in two parts. First, we show that a solution of the restricted minimization problem $\mathcal{P}_\lambda^J(y)$ over $\mathcal{G}_J$ has an implicit form as in (14).

Lemma 7. *Let $x^\star$ be a solution of*

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

Let J the D-cosupport of $x^\star$ and $s = \text{sign}(D^* x^\star)_I$. If (H_J) holds, then

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

Proof. Let U be a basis of $\mathcal{G}_J$. We rewrite $\mathcal{P}_\lambda^J(y)$ without constraints

$$\underset{\alpha \in \mathbb{R}^P}{\text{argmin}} \ \frac{1}{2}\|y - \Phi U \alpha\|_2^2 + \lambda \|D^* U \alpha\|_1.$$

Hence using Lemma 1 with ΦU and $D^* U$ in place of Φ and D^*, α_λ is a solution of $\mathcal{P}_\lambda^J(y)$ if, and only if, there exists σ such that

$$U^* \Phi^* (\Phi U \alpha_\lambda - y) + \lambda (U^* D)_I s + \lambda (U^* D)_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty < 1.$$

We conclude as in Lemma 2. By definition one has $U\alpha \in \mathcal{G}_J$. Hence, we can write

$$U^* \Phi^* (\Phi U \alpha_\lambda - y) + \lambda (U^* D)_I s = 0$$

Since (H_J) holds, the operator $U^* \Phi^* \Phi U$ is invertible, we conclude. □

We recall that the Recovery Criterion RC of $I \subset \{1 \ldots P\}$ is defined as

$$\mathbf{RC}(I) = \max_{x \in \mathcal{G}_J} \mathbf{IC}\left(\text{sign}(D^* x)\right).$$

Proof of Theorem 6. Let $x^\star$ be a solution of $\mathcal{P}_\lambda^J(y)$. Our strategy is to prove that $x^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$. By definition, $x^\star \in \mathcal{G}_J$, hence J is the D-cosupport of $x^\star$. Using Lemma 7, one has

$$x^\star = A^{[J]}\Phi^* y - \lambda A^{[J]} D_I s_I, \tag{22}$$

where $s = \text{sign}(D^* x^\star)_I$.

Remark that $\tilde{B}y = \tilde{B}w$ since $x^\star \in \mathcal{G}_J$. Using Lemma 5, the vector $x^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$ if there exists σ satisfying

$$\sigma = -u + \Omega s_I - \frac{1}{\lambda}\tilde{B}w \quad \text{where} \quad \|\sigma\|_\infty \leqslant 1,$$

for some $u \in \ker(D_J)$. If $\mathbf{RC}(I) < 1$, then

$$\|\sigma\|_\infty < \mathbf{RC}(I) + \frac{1}{\lambda}\|\tilde{B}\|_{2,\infty}\|w\|_2.$$

Hence, for

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

one has $\|\sigma\|_\infty < 1$. Hence, $x^\star$ is the unique solution of $\mathcal{P}_\lambda(y)$.

We now bound the distance between x_0 and $x^\star$.

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

We remark that $A^{[J]}\Phi^* y = x_0 + A^{[J]}\Phi^* w$. Hence,

$$\|x^\star - x_0\| = \|A^{[J]}(\Phi^* w - \lambda D_I s_I)\|.$$

Using operator norm inequality, one has

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

□

Conclusion

This paper has provided a theoretical analysis of the robustness of sparse analysis regularizations. We have studied both the local affine behavior of the solution, and the robustness to small and large noise. These contributions enable a better understanding of the behavior of this class of regularization.

Concrete examples illustrate our results. For discrete total variation, we show that staircasing induces an instability of the support, i.e discontinuities are not preserved. For Fused Lasso, our analysis shows that the support is stable and robust to an arbitrary bounded noise.

A distinctive feature of our approach is that we look for the robustness of the cospace associated to the original data. This approach often has a meaningful interpretation (such as the conservation of discontinuities for TV-like models) it also leads to quite restrictive conditions. A fascinating area for future work is to understand how to lift these restrictions to obtain sharper noise robustness of analysis methods

References

[1] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, 2010.

[2] T. Blumensath and M.E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory*, 55(4):1872–1882, 2009.

[3] P. Boufounos, G. Kutyniok, and H. Rauhut. Sparse recovery from combined fusion frame measurements. *IEEE Transactions on Information Theory*, 57 (6):3864–3876, 2011.

[4] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[5] E. Candès, Y.C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2010.

[6] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[7] V. Caselles, A. Chambolle, and M. Novaga. The discontinuity set of solutions of the tv denoising problem and some extensions. *Multiscale Modeling and Simulation*, 6(3):879–894, 2008.

[8] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.

[9] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[10] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock. *Theoretical Foundations and Numerical Methods for Sparse Recovery*, chapter An introduction to Total Variation for Image Analysis. De Gruyter, 2010.

[11] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.

[12] P.L. Combettes and J.C. Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[13] G. Davis, S.G. Mallat, and Z. Zhang. Adaptive time-frequency approximations with matching pursuits. Technical report, Courant Institute of Mathematical Sciences, 1994.

[14] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[15] D.L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[16] D.L. Donoho and Y. Tsaig. Fast solution of ℓ^1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008.

[17] C. Dossal. A necessary and sufficient condition for exact recovery by ℓ^1 minimization. Preprint HAL-00164738, 2007.

[18] B. Efron. How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470, 1986.

[19] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007.

[20] Y.C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory*, 55(11): 5302–5316, 2009.

[21] J. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 20(3):657–669, 2011.

[22] J.J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.

[23] M. Kachour, C. Dossal, J. Fadili, G. Peyré, and C. Chesneau. The degrees of freedom of the lasso in underdetermined linear regression models. In *Proc. SPARS 2011*, 2011.

[24] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Verlag, 2011.

[25] Y.M. Lu and M.N. Do. A theory for sampling signals from a union of subspaces. *IEEE Transactions on Signal Processing*, 56(6):2334–2345, 2008.

[26] S.G. Mallat. *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition, 2009.

[27] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[28] S. Nam, M.E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. Preprint arxiv-1106.4987v1, 2011.

[29] B.K. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[30] D. Needell, J. Tropp, and R. Vershynin. Greedy signal recovery review. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pages 1048–1050. IEEE, 2008.

[31] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658, 2000.

[32] M.R. Osborne, B. Presnell, and B.A. Turlach. A new approach to variable selection in least squares problems. *IMA journal of numerical analysis*, 20 (3):389, 2000.

[33] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, pages 40–44. IEEE, 1993.

[34] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[35] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4): 1307–1330, 1986.

[36] I.W. Selesnick and M.A.T. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *Proceedings of SPIE*, volume 7446, page 74460D, 2009.

[37] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM Journal on Numerical Analysis*, 42(2): 686–713, 2005.

[38] C.M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

[39] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

[40] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.

[41] R.J. Tibshirani and J. Taylor. The solution path of the generalized Lasso. *The Annals of Statistics*, 39(3):1335–1371, 2011.

[42] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[43] J.A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52 (3):1030–1051, 2006.

[44] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural computation*, 13(4):863–882, 2001.

[45] H. Zou, T. Hastie, and R. Tibshirani. On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5):2173–2192, 2007.