



HAL
open science

Robust Sparse Analysis Regularization

Samuel Vaiter, Gabriel Peyré, Charles H Dossal, Jalal M. Fadili

► **To cite this version:**

Samuel Vaiter, Gabriel Peyré, Charles H Dossal, Jalal M. Fadili. Robust Sparse Analysis Regularization. 2011. hal-00627452v2

HAL Id: hal-00627452

<https://hal.science/hal-00627452v2>

Submitted on 21 Feb 2012 (v2), last revised 2 Oct 2012 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust Sparse Analysis Regularization

Samuel Vaite^{*}, Gabriel Peyré^{*}, Charles Dossal[†], and Jalal Fadili[‡]

February 22, 2012

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 and 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 theoretical 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.vaite^r, 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

$$y = \Phi x_0 + w, \quad (1)$$

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 $Q \leq N$. 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

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

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 $\lambda \rightarrow 0$ and solve the constrained problem

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

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}^Q and α a vector in \mathbb{R}^P .

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

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

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

$$\text{supp}(\alpha) = \{i \in \{1, \dots, P\} \mid \alpha_i \neq 0\}.$$

For a set I , $|I|$ will denote the cardinal of I .

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 $\{1, \dots, P\}$ 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 implicit. For any matrix M , M^+ is the Moore–Penrose pseudoinverse of M and M^* is the adjoint matrix 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

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

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 algorithms 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 $x \in \text{Span}(D) \subseteq \mathbb{R}^N$ 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 $R = R_S$ can be written as $x = D\alpha$ where α is a solution of

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

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

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

and is referred to as Basis Pursuit [11]. Taking $D = \text{Id}$ 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

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

As the objective in $(\mathcal{P}_\lambda(y))$ is proper, continuous and convex, it is a classical existence result that the set of (global) minimizers is nonempty and compact if and only if

$$\text{Ker } \Phi \cap \text{Ker } D^* = \{0\}. \quad (H_0)$$

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 $D = \text{Id}$ and $\Psi = \Phi$.

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

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

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 $D = D_{\text{DIF}}$ where

$$D_{\text{DIF}} = \begin{pmatrix} -1 & & & 0 \\ +1 & -1 & & \\ & +1 & \ddots & \\ 0 & & \ddots & -1 \\ & & & +1 \end{pmatrix}. \quad (6)$$

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 produce 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 favor 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 $I = \text{supp}(D^*x)$. 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*

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

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

$$\mathbb{R}^N = \bigcup_{k \in \{0, \dots, N\}} \Theta_k \quad \text{where } \Theta_k = \{\mathcal{G}_J \mid J \subseteq \{1, \dots, P\} \text{ and } \dim \mathcal{G}_J = k\}. \quad (7)$$

The union of subspaces associated to synthesis regularization ($D = \text{Id}$) defines Θ_k as the set of axis-aligned subspaces of dimension k . For the 1-D total variation prior, where $D = D_{\text{DIF}}$ 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 five 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 five contributions.

2 Contributions

This paper proves the following five results:

1. **Local affine parameterization:** a solution of $\mathcal{P}_\lambda(y)$ is a piecewise affine function of (y, λ) .
2. **Degrees of freedom:** the degrees of freedom (as defined in [18]) of the sparse analysis estimator is the dimension of \mathcal{G}_J .
3. **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.
4. **Noiseless identifiability:** the same condition ensures that x_0 is the unique solution of $\mathcal{P}_0(y)$ when $w = 0$.
5. **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

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

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

$$\text{Ker } \Phi \cap \mathcal{G}_J = \{0\}. \quad (H_J)$$

Note that there is always a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds as shown in Lemma 7.

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

$$A^{[J]} = U (U^* \Phi^* \Phi U)^{-1} U^*. \quad (8)$$

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

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

Definition 4. The transition space \mathcal{H} is defined as

$$\mathcal{H} = \bigcup_{\substack{J \subset \{1, \dots, P\} \\ (H_J) \text{ holds}}} \bigcup_{\substack{K \subset J \\ \tilde{\Pi}^{[J]} \not\subset \text{Im } D_{K^c \cap J}}} \bigcup_{s_{J^c} \in \{-1, 1\}^{|J^c|}} \bigcup_{s_K \in \{-1, 1\}^{|K|}} \mathcal{H}_{J, K, s_{J^c}, s_K},$$

where

$$\mathcal{H}_{J, K, s_{J^c}, s_K} = \left\{ (y, \lambda) \in \mathbb{R}^Q \times \mathbb{R} \setminus P_{\mathcal{G}_{K^c \cap J}} \tilde{\Pi}^{[J]} y = \lambda (\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \right\},$$

where $\tilde{\Pi}^{[J]} = \Phi^* (\Phi A^{[J]} \Phi^* - \text{Id})$, $\tilde{\Omega}^{[J]} = (\Phi^* \Phi A^{[J]} - \text{Id}) D_{J^c}$ and $P_{\mathcal{G}_{K^c \cap J}}$ is the orthogonal projection on $\mathcal{G}_{K^c \cap J}$.

We prove the following theorem

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

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

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

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.

2.2 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].

Definition 5. Let $\lambda \in \mathbb{R}_+^*$. We define the λ -restricted transition space as

$$\mathcal{H}_{\cdot,\lambda} = \{y \in \mathbb{R}^Q \mid (y, \lambda) \in \mathcal{H}\}.$$

We first notice that even if $\mathcal{P}_\lambda(y)$ admits several solutions, all of them share the same image under Φ , see Section 5.2 for proof of this point. Hence, we denote without ambiguity $\mu(y) = \Phi x^*$ where x^* is a solution of $\mathcal{P}_\lambda(y)$. We have the following result for the analysis regularization.

Theorem 2. Let $\lambda \in \mathbb{R}_+^*$. The mapping $y \mapsto \mu(y)$ is of class C^∞ on $\mathbb{R}^Q \setminus \mathcal{H}_{\cdot,\lambda}$. For $y \notin \mathcal{H}_{\cdot,\lambda}$, there exists x^* a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds with J the D -cosupport of x^* , and

$$\operatorname{div}(\mu(y)) = \dim(\mathcal{G}_J). \quad (9)$$

For $y \notin \mathcal{H}_{\cdot,\lambda}$, we define $d(y) = \dim(\mathcal{G}_J)$ where J is the D -cosupport of any solution x^* such that (H_J) holds. Let the noise $w \sim \mathcal{N}(0, \sigma^2 \operatorname{Id})$, and therefore $y \sim \mathcal{N}(\mu_0 = \Phi x_0, \sigma^2 \operatorname{Id})$.

From the seminal definition of Efron [18], and by the Stein Lemma [38], if μ is weakly differentiable, the degrees of freedom is given by

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

Corollary 1. Let $\lambda \in \mathbb{R}_+^*$. The λ -restricted transition space has a Lebesgue measure zero. Moreover, the mapping $y \mapsto \mu(y)$ is differentiable almost everywhere and

$$df(\mu) = \mathbb{E}_w(d(y)).$$

This result is known to hold in the special case of synthesis regularization ($D = \operatorname{Id}$). 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 $\mathcal{H}_{\cdot,\lambda}$. Our expression handles both over- and underdetermined measurements with a sharper characterization of the set $\mathcal{H}_{\cdot,\lambda}$.

2.3 Robustness to Small Noise

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

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 Identifiability Criterion **IC** of s is defined as

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

We have the following theorem.

Theorem 3. 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 $\mathbf{IC}(\text{sign}(D^* x_0)) < 1$. There exist two constants $c_J > 0$ and $\tilde{c}_J > 0$, such that if $y = \Phi x_0 + w$, where

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

and if λ satisfies

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

the vector defined by

$$\hat{x}^* = x_0 + A^{[J]} \Phi^* w - \lambda A^{[J]} D_I s_I, \tag{10}$$

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

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

Note that it is possible to choose λ proportional to the noise level $\|w\|_2$. Hence, for $\|w\|_2$ small enough, equation (10) gives

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

2.4 Noiseless Identifiability

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

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 4. 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.

2.5 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

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

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 5. 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

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

where c_J is defined as

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

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

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

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 \mathbf{IC} introduced in Definition 6 to the case where $D = \text{Id}$.

Definition 9. Let $s \in \{-1, 0, +1\}^P$, I its support and J its cosupport. We suppose Ψ_I has full rank. The Sign Criterion \mathbf{IC}_S of a sign vector s associated to a support I is defined as

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

Fuchs shows the following result.

Theorem ([22]). *Let $\alpha_0 \in \mathbb{R}^P$ be a fixed vector of support I . If Ψ_I has full rank and $\mathbf{IC}_S(\text{sign}(\alpha_0)) < 1$, 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 , there exists a unique solution α^* of (4) for $y = \Psi\alpha_0 + w$ of support included in I , verifying $\|\alpha_0 - \alpha^*\|_2 = O(\|w\|_2)$.*

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

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

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

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 to noise with overwhelming probability on the matrix Φ 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 $\text{Ker } \Phi$ 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 $\mathbf{IC}_0(\text{sign}(D^*x_0)) < 1$ and*

$$\mathbf{IC}_0(s) = \|\mathfrak{I}_I s_I\|_{\infty} \quad \text{where} \quad \mathfrak{I}_I = (MD_J)^+ MD_I,$$

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 \mathbf{IC} . 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 $s = \text{sign}(D^*x_0)$ 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^* 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 $\frac{\|x_0 - x^*\|_2}{\|w\|_2}$ is arbitrary large, whatever the amplitude $\|w\|_2$ of the noise.

4 Examples

This section details algorithms to compute identifiability criteria \mathbf{IC} and \mathbf{RC} , together with a study of total variation, shift invariant Haar transform and Fused Lasso regularizations.

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

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

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

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

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

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

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

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

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

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

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

Unfortunately, computing **RC** necessitates to solve a combinatorial optimization problem which is not convex. Recall that

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

A stronger criterion, which is easy to compute, is obtained by selecting $u = 0$ in $\text{Ker } D_J$

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

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

$$\mathbf{IC}(\text{sign}(D^*x_0)) \leq \mathbf{RC}(I) \leq \mathbf{wRC}(I).$$

4.3 Total Variation Denoising

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

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

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

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

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 sub-signal, then $\mathbf{IC}(\text{sign}(D^*x)) < 1$. Otherwise, $\mathbf{IC}(\text{sign}(D^*x)) = 1$.*

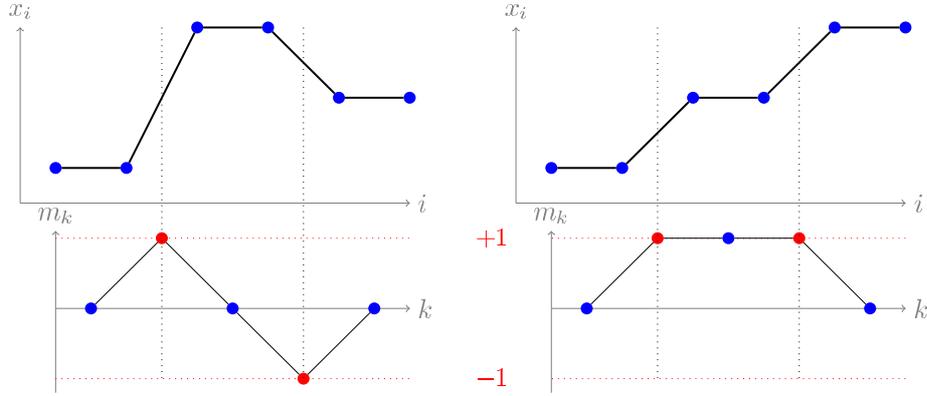


Figure 1: Top line: Signals x with 2 discontinuities. Bottom line: Associated dual vector m .

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

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

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 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 sub-signal, 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, i_2 are consecutive indexes of I , then for every $i_1 < j < i_2$, $m_j = s_{i_1} = \pm 1$ which implies $\mathbf{IC}(\text{sign}(D^*x)) = 1$. \square

This proposition together with Theorem 3 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 2. *If $|I| > 2$ such that $i \in I$ implies $i + 1 \notin I$, and $D = D_{DIF}$, then $\mathbf{RC}(I) = 1$.*

Proof. If $|I| > 2$, there exists a signal \tilde{x} which contain a staircase sub-signal, hence $1 = \mathbf{IC}(\text{sign}(D^*\tilde{x})) \leq \mathbf{RC}(I)$. Since there is no signal $x \in \mathbb{R}^N$ such that $\mathbf{IC}(\text{sign}(D^*x)) > 1$, we conclude. \square

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

4.4 Invariant Haar Transform

Sparse analysis regularization using a shift invariant Haar dictionary is efficient to recover piecewise constant signals. This dictionary is defined using a set of dilated Haar filters

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

We define the translation invariant Haar dictionary as

$$D_H^* x = \left(\psi^{(j)} \star x \right)_{0 \leq j < \log_2(N)}.$$

The analysis regularization $\|D_H^* x\|_1$ is a sum of the TV norm of filtered versions of the signal, it can thus be understood as some kind of multi-scale total variation. We consider the case where Φ is a realization from the uniform random distribution on the set of orthogonal projectors, that satisfy $\Phi^* \Phi = \text{Id}$. The inverse problem thus corresponds to a compressed sensing acquisition. Figure 2 shows the evolution of \mathbf{IC} as a function of the redundancy Q/N of the operator Φ for different box signals. More precisely, we consider the collection of box signals

$$x_\eta[i] = \begin{cases} 1 & \text{if } |i - \frac{n}{2}| \leq \eta N \\ 0 & \text{elsewhere.} \end{cases}$$

Figure 2 displays the average and standard deviation of \mathbf{IC} for three different values of η as a function of $Q/N \in [0.4, 1]$. They are estimated numerically using Monte-Carlo simulation, using 1000 samples for each redundancy. Remark that \mathbf{IC} increases when the signal converges to a single spike signal and the redundancy Q/N diminishes.

4.5 Fused Lasso

Fused Lasso is introduced in [40]. It is equivalent to $\mathcal{P}_\lambda(y)$ when using

$$D = [D_{\text{DIF}} \quad \varepsilon \text{Id}],$$

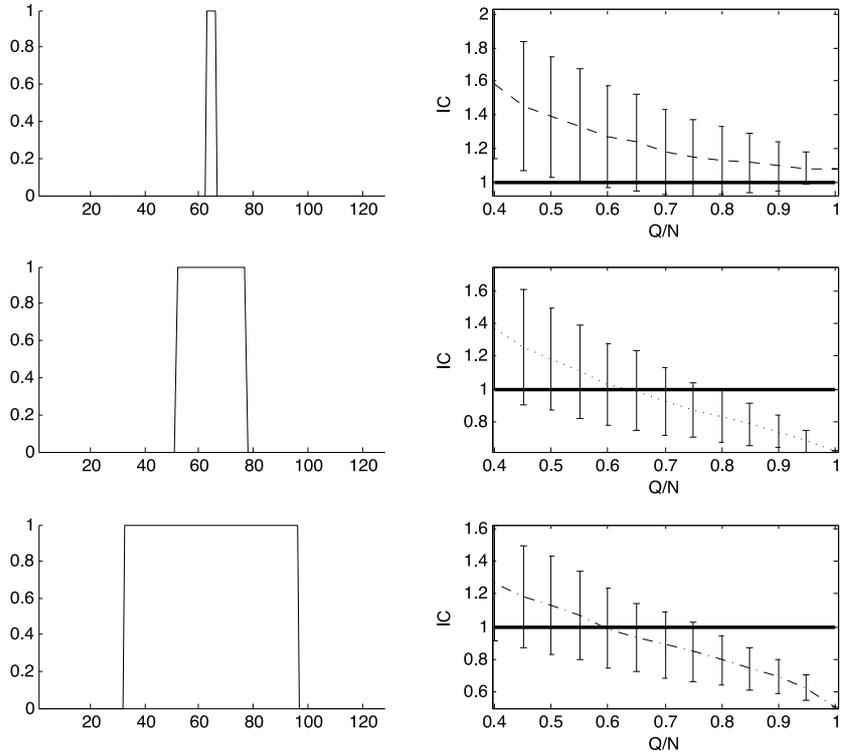


Figure 2: Evolution of \mathbf{IC} for a compressed sensing matrix with an invariant Haar dictionary. On the left side, a box signal. On the right, the dotted line represents the average \mathbf{IC} as a function of Q/N . The vertical lines represent the interval $[\text{mean}(\mathbf{IC}) - \text{std}(\mathbf{IC}), \text{mean}(\mathbf{IC}) + \text{std}(\mathbf{IC})]$. The horizontal line indicates the saturation level $\mathbf{IC} = 1$.

where ε is a positive real number. The associated union of subspaces (7) is $\bigcup_k \Theta_k$ where Θ_k is the set of sum of k interval indicators, 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 $a_i \leq b_i < a_{i+1}$.

We consider the case where Φ is a realization of the Gaussian matrix ensemble, which has i.i.d. entries distributed according to the normal law $\mathcal{N}(0, 1)$. We consider the collection of sum of two indicators

$$x_{\eta, \rho} = \mathbf{1}_{[(\frac{1}{2}-\eta-\rho)N, (\frac{1}{2}-\rho)N]} + \mathbf{1}_{[(\frac{1}{2}+\rho)N, (\frac{1}{2}+\eta+\rho)N]}. \quad (11)$$

We fixed $\rho = \frac{1}{4}N$, $\varepsilon = 50/N$. Figure 3 shows the evolution of the mean and standard deviation of \mathbf{IC} as a function of the redundancy $Q/N \in [0.5, 1]$ of Φ for different box signals. They are estimated numerically using Monte-Carlo simulation, using 1000 samples for each redundancy. Remark that \mathbf{IC} diminished when the signal converges to two spikes and when the redundancy Q/N

increases. An other choice of ε may lead to different results depending if D favors the ℓ^1 -sparsity or the total variation sparsity.

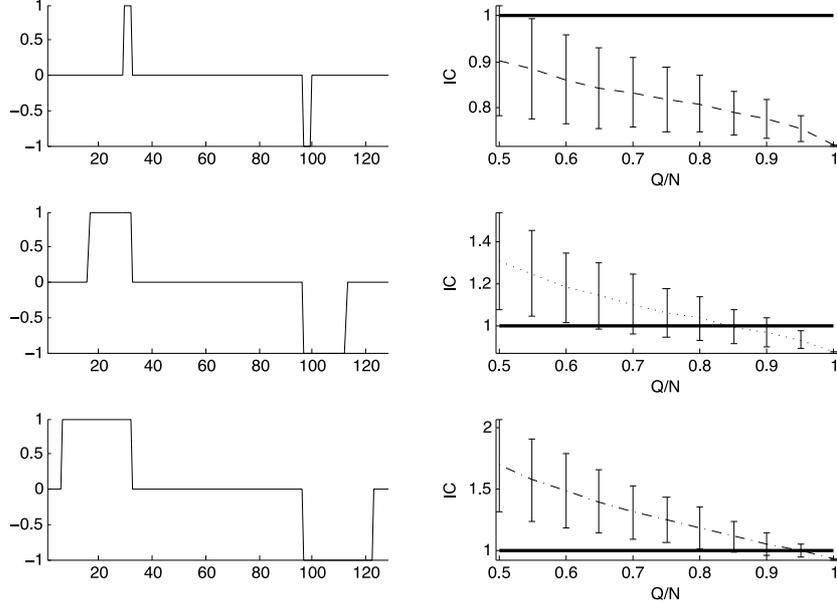


Figure 3: Evolution of IC for a compressed sensing matrix with a Fused Lasso dictionary. On the left side, a signal with a fixed interval size $\eta = 0.025N, 0.125N, 0.2N$. On the right, the average and the standard deviation of IC as a function of the redundancy Q/N of the random matrix.

5 Proofs

This section details the proofs of Theorems 1 – 5. The objective function $\mathcal{L}_{y,\lambda}$ minimized in $\mathcal{P}_\lambda(y)$ is

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

We recall that we suppose that condition (H_0) holds in every statements. 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^* 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^* , such that*

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

where $I = J^c$ the D -support,

$$\Sigma_{y,\lambda}(x^*) = \left\{ \sigma \in \mathbb{R}^{|J|} \mid \Phi^*(\Phi x^* - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leq 1 \right\}. \quad (13)$$

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

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

$$\partial F(x_0) = \{g \in \mathbb{R}^N \mid \forall x \in \mathbb{R}^N, f(x) \geq f(x_0) + \langle g, x - x_0 \rangle\}.$$

Note that x_0 is a minimum of F if, and only if, $0 \in \partial F(x_0)$. Indeed, if $0 \in \partial F(x_0)$, then for every $x \in \mathbb{R}^N$, $F(x) \geq F(x_0)$, meaning that x_0 is a minimum of F over \mathbb{R}^N . The subdifferential of $\mathcal{L}_{y,\lambda}(x)$ is

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

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

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

Defining $\sigma = u_J$, it is equivalent to the existence of $\sigma \in \Sigma_{y,\lambda}(x)$ with $\|\sigma\|_\infty \leq 1$. \square

The following lemma characterizes the normal cone at zero of the subdifferential of $\mathcal{L}_{y,\lambda}$ at a minimizer.

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

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

where $\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0)$ is the normal cone at zero of the subdifferential of $\mathcal{L}_{y,\lambda}$ in x^* defined by

$$\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0) = \{z \in \mathbb{R}^N \mid \langle z, d \rangle \leq 0 \text{ for every } d \in \partial \mathcal{L}_{y,\lambda}(x^*)\}.$$

Moreover, if J is the D -cosupport of x^* , then

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

Proof. Let $I = J^c$. We decompose I such that $I = I^* \cup J^*$. Since $\|\sigma_J\|_\infty < 1$, one remarks that \bar{u} defined by

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

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

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

We introduce $\varepsilon > 0$ such that $\|\sigma_J\|_\infty = 1 - \varepsilon$. Consider the set

$$\mathcal{U} = \{u \in \mathbb{R}^P \mid \|u_J - \bar{u}_J\|_\infty \leq \varepsilon \text{ and } u_I = \bar{u}_I\}.$$

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

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

and we denote

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

Remark that

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

Since $u_I = \bar{u}_I$, one has

$$d_u = \lambda D_J(u_J - \bar{u}_J).$$

Let $z \in \mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0)$ and let $u \in \mathcal{U}$. Note that

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

and

$$u_{J^*} = \text{sign}(D_{J^*}^* x)_{J^*} \text{ and } \|u_{J^*}\|_\infty \leq 1.$$

Hence, $d_u \in \partial \mathcal{L}_{y,\lambda}(x^*)$. By definition,

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

Particularly,

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

Remark that for every $u \in \mathcal{U}$, $2\bar{u} - u \in \mathcal{U}$ and $d_{2\bar{u}-u} = -d_u$. Indeed,

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

Moreover,

$$\|(2\bar{u} - u)_J - \bar{u}_J\|_\infty = \|\bar{u} - u_J\|_\infty \leq \varepsilon,$$

and $(2\bar{u} - u)_I = 2\bar{u}_I - u_I = \bar{u}_I$. Hence

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

Therefore,

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

Let $v \in \text{Im } D_J \setminus \{0\}$. Remark there exists $\mu_v \in \mathbb{R}^*$ such that

$$\mu_v v = D_J \sigma_v \quad \text{and} \quad \|\sigma_v\|_\infty \leq \varepsilon,$$

We define then the vector u as

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

Note that u is an element of \mathcal{U} since $\|u_J - \bar{u}_J\|_\infty = \|\sigma_v\|_\infty \leq \varepsilon$. Therefore,

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

is such that $\langle z, d_u \rangle = 0$ and $d_u \in \mathcal{D}$, i.e $\text{Im } D_J = \text{Span}(\mathcal{D})$. Finally,

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

We conclude that $\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0)$ is included in $(\text{Im } D_J)^\perp = \mathcal{G}_J$.

Suppose now that J is the D -cosupport of x^* . We prove that

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

Remark that $\partial \mathcal{L}_{y,\lambda}(x^*) \subseteq \text{Im } D_J$. Indeed, let $d \in \partial \mathcal{L}_{y,\lambda}(x^*)$. We write $d = \Phi^*(\Phi x^* - y) + \lambda D_J u$ with $u_I = \text{sign}(D_I^* x)_I$ and $\|u_J\|_\infty \leq 1$. Since $0 \in \partial \mathcal{L}_{y,\lambda}(x^*)$, one has

$$d = \lambda D(u - \bar{u}),$$

and since $u_I = \bar{u}_I$, one has

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

Hence, $(\text{Im } D_J)^\perp$ is included in the normal cone $\mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0)$. \square

The following lemma gives a sufficient condition to guarantee the uniqueness of the solution of $\mathcal{P}_\lambda(y)$.

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

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

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

Proof. We decompose $\mathcal{L}_{y,\lambda}$ in two functions:

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

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

1. If $h \notin \mathcal{G}_J$, then using Lemma 2, $h \notin \mathcal{N}_{\partial \mathcal{L}_{y,\lambda}(x^*)}(0)$ and there exists $d \in \partial \mathcal{L}_{y,\lambda}(x^*)$ such that $\langle d, h \rangle > 0$ and

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

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

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

where $v \in \partial_{\|D^*\|_1}(x^*)$ such that $\lambda v + \nabla q(x^*) = 0$. Then,

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

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

5.1 Proof of Theorem 1

The proof of Theorem 1 is done in three steps. First, we prove Lemma 4 which gives an implicit equation satisfied by a solution of $\mathcal{P}_\lambda(y)$. Then, we prove Lemma 5. Finally, we prove Theorem 1.

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

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

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

Proof. Using the first order condition (Lemma 1) there exists $\sigma \in \Sigma_{y,\lambda}(x^*)$ satisfying

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

By definition, one has $x^* \in \mathcal{G}_J$ so $x^* \in (\text{Im } D_J)^\perp$. Hence, we can write $x^* = U\alpha$. Since $U^* D_J = 0$, multiplying equation (15) on the left by U^* , we get

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

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

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}^* satisfies*

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

where $s = \text{sign}(D^ \hat{x}^*)$. Then, \hat{x}^* is a solution of $\mathcal{P}_\lambda(y)$ if, and only if, there exists σ satisfying one of the following conditions*

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

or equivalently,

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

where $\tilde{\Omega}^{[J]} = (\Phi^*\Phi A^{[J]} - \text{Id})D_I$, $\tilde{\Pi}^{[J]} = \Phi^*(\Phi A^{[J]}\Phi^* - \text{Id})$, $\Omega^{[J]} = D_J^+\tilde{\Omega}^{[J]}$ and $\Pi^{[J]} = D_J^+\tilde{\Pi}^{[J]}$. Moreover, if $\|\sigma\|_\infty < 1$ then \hat{x}^* is the unique solution of $\mathcal{P}_\lambda(y)$.

Proof. Remark that \hat{x}^* is an element of \mathcal{G}_J . According to Lemma 1, \hat{x}^* is a solution of $\mathcal{P}_\lambda(y)$ if, and only if, there exists $\sigma \in \Sigma_{y,\lambda}(\hat{x}^*)$ such that

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

Since (H_J) holds, one can define $A^{[J]}$. We use the implicit equation (14),

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

Factorizing the term in front of y and s_I , one has

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

which proves that

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

One has $U^*\tilde{\Omega} = 0$ and thus one remarks that $\Omega^{[J]} = D_J^+\tilde{\Omega}^{[J]}$. Similarly, we define $\tilde{\Pi}^{[J]}$ such that $\Pi^{[J]} = D_J^+\tilde{\Pi}^{[J]}$. Hence, the existence of $\sigma \in \Sigma_{y,\lambda}(\hat{x}^*)$ such that $\|\sigma\|_\infty \leq 1$ is equivalent to

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

which in turn is equivalent to

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

Replacing the inequality by a strict inequality condition gives the uniqueness of x^* using Lemma 3. \square

We now prove Theorem 1.

Proof of Theorem 1. Let $(y, \lambda) \notin \mathcal{H}$. By construction of the vector $\hat{x}_{\bar{\lambda}}(\bar{y})$ one has $D_J^*\hat{x}_{\bar{\lambda}}(\bar{y}) = 0$. So for $(\bar{y}, \bar{\lambda})$ close enough from (y, λ) , one has

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

Since x^* is a solution of $\mathcal{P}_\lambda(y)$, using Lemmas 4 and 5, there exists σ such that

$$\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_J\sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leq 1. \quad (18)$$

We split $J = K \cup L$, $K \cap L = \emptyset$ such that $\|\sigma_K\|_\infty = 1$ and $\|\sigma_L\|_\infty < 1$. We first suppose that $\text{Im } \tilde{\Pi}^{[J]} \subseteq \text{Im } D_L$. To prove that $\hat{x}_{\bar{\lambda}}(\bar{y})$ is solution to $\mathcal{P}_{\bar{\lambda}}(\bar{y})$ we show that there exists $\bar{\sigma}$ such that $\|\bar{\sigma}\|_\infty \leq 1$ and

$$\tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_K\bar{\sigma}_K + \bar{\lambda}D_L\bar{\sigma}_L = 0.$$

We impose that $\bar{\sigma}_K = \sigma_K$ and we introduce $\bar{\sigma}_L$ as

$$\bar{\sigma}_L = \sigma_L - \frac{1}{\lambda}D_L^+\tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right).$$

Hence,

$$\begin{aligned} & \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_J\bar{\sigma} \\ = & \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_K\sigma_K + \bar{\lambda}D_L\sigma_L \\ & - D_LD_L^+\frac{\bar{\lambda}}{\lambda}\tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right) \\ = & \underbrace{\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_K\sigma_K + \lambda D_L\sigma_L}_{=0} \\ & - \tilde{\Pi}^{[J]}(y - \bar{y}) + (\lambda - \bar{\lambda})\tilde{\Omega}^{[J]}s_I - (\lambda - \bar{\lambda})D_K\sigma_K - (\lambda - \bar{\lambda})D_L\sigma_L \\ & - D_LD_L^+\frac{\bar{\lambda}}{\lambda}\tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right) \end{aligned}$$

Since $\text{Im } \tilde{\Pi}^{[J]} \subseteq \text{Im } D_L$, there exists u such that

$$\tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right) = D_L u.$$

By property of Moore-Penrose pseudo-inverse,

$$D_LD_L^+D_L u = D_L u = \tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right).$$

Hence,

$$\begin{aligned} & \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_J\bar{\sigma} \\ = & \frac{\bar{\lambda} - \lambda}{\lambda}\left[\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_K\sigma_K + \lambda D_L\sigma_L\right] \\ = & 0. \end{aligned}$$

If $(\bar{y}, \bar{\lambda})$ is close enough from (y, λ) , one has

$$\|\bar{\sigma}_L\|_\infty = \left\|\sigma_L - \frac{1}{\lambda}D_L^+\tilde{\Pi}^{[J]}\left(\frac{\lambda - \bar{\lambda}}{\lambda}y + \frac{\lambda}{\lambda}(\bar{y} - y)\right)\right\|_\infty \leq 1,$$

i.e $\hat{x}_{\bar{\lambda}}(\bar{y})$ is solution of $\mathcal{P}_{\bar{\lambda}}(\bar{y})$. Suppose now that $\text{Im } \tilde{\Pi} \not\subseteq \text{Im } D_L$. Then remark that projecting (18) on \mathcal{G}_L shows that

$$P_{\mathcal{G}_L}\tilde{\Pi}^{[J]}y = \lambda(\tilde{\Omega}^{[J]}s_{J^c} + D_Ks_K),$$

which is a contradiction of $(y, \lambda) \notin \mathcal{H}$. \square

5.2 Proof of Theorem 2

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 $\text{div}(\mu)(y) = \dim \mathcal{G}_J$.

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

Lemma 6. *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

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

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 $\mathcal{P}_\lambda(y)$. \square

Lemma 7. *There exists x^* a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds, where J is the D -cosupport of x^* .*

Proof. Let x^* be a solution of $\mathcal{P}_\lambda(y)$. Suppose (H_J) does not hold. Our strategy is to prove that there exists a solution of D -support strictly included in $I = J^c$.

Since (H_J) does not hold, there exists $z \in \text{Ker } \Phi$ with $z \neq 0$ and $D_J^* z = 0$. We define for every $t \in \mathbb{R}$, the vector $v_t = x^* + tz$. Denote \mathcal{B} the subset of \mathbb{R} defined by

$$\mathcal{B} = \{t \in \mathbb{R} \mid \text{sign}(D^* v_t) = \text{sign}(D^* x^*)\},$$

The set \mathcal{B} is a non empty set, $0 \in \mathcal{B}$ and convex from its definition. Moreover for all $t \in \mathcal{B}$, $\partial \mathcal{L}_{y,\lambda}(v_t) = \partial \mathcal{L}_{y,\lambda}(x^*)$, it follows that for all $t \in \mathcal{B}$, v_t is a solution of $\mathcal{P}_\lambda(y)$. As a consequence,

$$\forall t \in \mathcal{B}, \quad \Phi v_t = \Phi x^* \quad \text{and} \quad \|D^* v_t\|_1 = \|D^* x^*\|_1.$$

Since $\lim_{|t| \rightarrow \infty} \|D^* v_t\|_1 = +\infty$, the set \mathcal{B} is bounded. Hence, \mathcal{B} is an open interval of \mathbb{R} which contain 0, i.e there exist $t_1, t_0 \in \mathbb{R}$ such that

$$\mathcal{B} =]t_1, t_0[\quad \text{where} \quad -\infty < t_1 < 0 \quad \text{and} \quad 0 < t_0 < +\infty.$$

Since $t_0 \notin \mathcal{B}$, the D -support of v_{t_0} is strictly included in I . Moreover by continuity,

$$\Phi v_{t_0} = \Phi x^* \quad \text{and} \quad \|D^* v_{t_0}\|_1 = \|D^* x^*\|_1.$$

Hence, v_{t_0} is a solution of $\mathcal{P}_\lambda(y)$ of D -support strictly included in I .

Iterating this argument for $x^* = v_{t_0}$ shows that there exists a solution such that (H_J) holds. \square

Next, we prove the theorem 2.

Proof of Theorem 2. Using Lemma 7, there exists a solution x^* of $\mathcal{P}_\lambda(y)$ such that (H_J) holds. We consider this solution. Using Theorem 1 for \bar{y} close enough from y one has

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

where J is the D -cosupport of x^* . Remark that $\mu(\bar{y})$ can be 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 $\dim(\operatorname{Im}(V)) = \dim(\mathcal{G}_J)$. \square

We now prove the Corollary 1 starting with a lemma on the measure of $\mathcal{H}_{\cdot, \lambda}$.

Lemma 8. *Let $J \subset \{1, \dots, P\}$ such that (H_J) holds, K a subset of J such that $\operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}$, $s_{J^c} \in \{-1, 1\}^{|J^c|}$ and $s_K \in \{-1, 1\}^{|K|}$. If $\operatorname{Im} \tilde{\Pi}^{[J]}$ is not included in $\operatorname{Im} D_{K^c}$ then $\mathcal{H}_{J, K, s_{J^c}, s_K}$ is an affine space of $\mathbb{R}^Q \times \mathbb{R}$ and different from $\mathbb{R}^Q \times \mathbb{R}$. Moreover, \mathcal{H} has a Lebesgue measure zero and for every $\lambda \in \mathbb{R}_+^*$ $\mathcal{H}_{\cdot, \lambda}$ has a Lebesgue measure zero.*

Proof. Consider $J \subset \{1, \dots, P\}$ such that (H_J) holds, K a subset of J such that $\operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}$, $s_{J^c} \in \{-1, 1\}^{|J^c|}$ and $s_K \in \{-1, 1\}^{|K|}$. The following set

$$\mathcal{H}_{J, K, s_{J^c}, s_K} = \left\{ (y, \lambda) \in \mathbb{R}^Q \times \mathbb{R} \setminus P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} y = \lambda (\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \right\},$$

is a vector subspace of $\mathbb{R}^Q \times \mathbb{R}$. Indeed, let $(y_1, \lambda_1), (y_2, \lambda_2) \in \mathcal{H}_{J, K, s_{J^c}, s_K}$ and $\mu \in \mathbb{R}$. Hence,

$$P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} (y_1 + \mu y_2) = (\lambda_1 + \mu \lambda_2) (\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K).$$

Moreover $(0_Q, 0) \in \mathcal{H}_{J, K, s_{J^c}, s_K}$. Each $\mathcal{H}_{J, K, s_{J^c}, s_K}$ is different from $\mathbb{R}^Q \times \mathbb{R}$. Indeed, $(y, \lambda) \in \mathcal{H}_{J, K, s_{J^c}, s_K}$ is equivalent to

$$\begin{pmatrix} P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} & 0 \\ 0 & -\lambda \operatorname{Id} \end{pmatrix} \begin{pmatrix} y \\ \lambda (\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \end{pmatrix} = 0.$$

Particularly, we fixed λ . If every $y \in \mathbb{R}^Q$ is solution of this system, the matrix $P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]}$ is invertible, which is impossible since $P_{\mathcal{G}_{K^c}}$ is an orthogonal projection on a strict subspace of \mathbb{R}^Q . Since \mathcal{H} is a finite union of $\mathcal{H}_{J, K, s_{J^c}, s_K}$ all different from $\mathbb{R}^Q \times \mathbb{R}$, \mathcal{H} has a Lebesgue measure zero. Remark that $\mathcal{H}_{\cdot, \lambda}$ is included in

$$\tilde{\mathcal{H}}^\lambda = \bigcup_{\substack{J \subset \{1, \dots, P\} \\ (H_J) \text{ holds}}} \bigcup_{\substack{K \subset J \\ \operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}}} \bigcup_{s_{J^c} \in \{-1, 1\}^{|J^c|}} \bigcup_{s_K \in \{-1, 1\}^{|K|}} \tilde{\mathcal{H}}_{J, K, s_{J^c}, s_K},$$

where

$$\tilde{\mathcal{H}}_{J,K,s_{J^c},s_K}^\lambda = \left\{ y \in \mathbb{R}^Q \setminus P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} y = \lambda(\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \right\},$$

Similarly to $\mathcal{H}_{J,K,s_{J^c},s_K}$, we prove that each $\tilde{\mathcal{H}}_{J,K,s_{J^c},s_K}^\lambda$ is a strict affine subspace of \mathbb{R}^Q . Hence, $\tilde{\mathcal{H}}^\lambda$ has a Lebesgue measure zero, and so does $\mathcal{H}_{\cdot,\lambda}$. \square

Proof of Corollary 1. Let $\lambda \in \mathbb{R}_+^*$. Using Lemma 8, $\mathcal{H}_{\cdot,\lambda}$ has a Lebesgue measure zero. Hence, $y \mapsto \mu(y)$ is differentiable almost everywhere and we can apply Stein's Lemma which gives

$$df(\mu_0) = \mathbb{E}_w(\operatorname{div}(\mu(y))) = \mathbb{E}_w(\operatorname{dim}(\mathcal{G}_J)).$$

\square

5.3 Proof of Theorem 3

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 Identifiability Criterion **IC** of a sign vector s associated to a D -support I is defined as

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

Proof of Theorem 3. The proof is done in three steps.

1. We give a condition on λ to have $\operatorname{sign}(D^* \hat{x}^*) = \operatorname{sign}(D^* x_0)$.
2. We give an other condition on $\frac{\|w\|_2}{\lambda}$ to ensure first-order condition on \hat{x}^* assuming $\mathbf{IC} < 1$.
3. We prove that the two conditions are compatible.

We consider the vector defined by

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

1. We first give a condition on λ to ensure signs equality

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

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

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

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

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

Using operator norm inequalities, one has

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

Introducing

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

the following condition

$$T > \|D_I^*A^{[J]}\|_{\infty,\infty} (\|\Phi^*\|_{2,\infty}\|w\|_2 + \lambda\|D_I\|_{\infty,\infty}), \quad (20)$$

ensures (19).

2. We now give a condition on $\frac{\|w\|_2}{\lambda}$ to ensure first-order condition (17) assuming $\mathbf{IC}(\text{sign}(D^*x_0)) < 1$. Remark that $\tilde{\Pi}y = \tilde{\Pi}w$ since $x_0 \in \mathcal{G}_J$. The minimum over $\text{Ker } D_J$ of $\|\Omega^{[J]}s_I - u\|_\infty$ is reached for a given $\bar{u} \in \text{Ker } D_J$. We consider the following σ defined by

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

Using operator norm inequality, one has

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

By definition of \bar{u} ,

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

Hence, under condition $\mathbf{IC}(\text{sign}(D^*x_0)) < 1$ and

$$\|\tilde{\Pi}^{[J]}\|_{2,\infty}\frac{\|w\|_2}{\lambda} < 1 - \mathbf{IC}(\text{sign}(D^*x_0)), \quad (21)$$

one has $\|\sigma\|_\infty < 1$ and using Lemma 5, the vector x^* is the unique solution of $\mathcal{P}_\lambda(y)$.

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

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

and

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

Suppose that

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

and

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

Then (20) and (21) are satisfied. \square

5.4 Proof of Theorem 4

The proof of Theorem 4 is done in three steps. First, we specialize Theorem 3 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 4 by considering an other potential solution of $\mathcal{P}_0(y)$.

Corollary 3. *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 $\lambda < \tilde{c}_J$,*

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

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

Proof. Take $w = 0$ in theorem 3. □

Lemma 9. *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 3, $\mathcal{P}_\lambda(y)$ has a unique solution for $\lambda < \tilde{c}_J$,

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

Let $x_{(1)} \neq x_0$ such that $\Phi x_{(1)} = y$. For every λ strictly positive, one has $\mathcal{L}_{y,\lambda}(x^*) < \mathcal{L}_{y,\lambda}(x_{(1)})$ by definition of x_λ . Then,

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

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

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

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

Proof of Theorem 4. Using Lemma 9, 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, one has $\text{sign}(D^*x_{(1)}) = \text{sign}(D^*x_0)$. Hence, if $\mathbf{IC}(\text{sign}(D^*x_0)) < 1$, then Corollary 3 holds and 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. □

5.5 Proof of Theorem 5

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

$$\mathbf{RC}(I) = \max_{\|p_I\|_\infty \leq 1} \min_{u \in \text{Ker } D_I} \|\Omega^{[J]} p_I - u\|_\infty.$$

Proof of Theorem 5. Consider the following restricted problem

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

Our strategy is to consider a solution of $\mathcal{P}_\lambda^J(y)$, and showing that it is the unique solution of $\mathcal{P}_\lambda(y)$. To achieve this goal, we use four steps:

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

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

2. We prove that x^* satisfies an equation of the form

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

3. We prove that x^* satisfies the first-order condition of Lemma 1 using the construction of p_I^* .

4. Finally, using operator norm inequalities we provide the bound announced in the statement.

We rewrite $\mathcal{P}_\lambda^J(y)$ without constraints

$$\underset{\alpha \in \mathbb{R}^{\dim \mathcal{G}_J}}{\text{argmin}} \frac{1}{2} \|y - \Phi U \alpha\|_2^2 + \lambda \|D_I^* U \alpha\|_1.$$

1. Using Lemma 1 with ΦU and $D_I^* U$ in place of Φ and D^* , if α^* is a solution of $\mathcal{P}_\lambda^J(y)$, then there exists σ^* with $\|\sigma^*\|_\infty \leq 1$ such that

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

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

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

which satisfies

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

First order conditions become

$$U^* [\Phi^* (\Phi U \alpha^* - y) + \lambda D_I p_I^*] = 0. \quad (22)$$

2. Moreover, the condition (H_J) holds, so the matrix $U^*\Phi^*\Phi U$ is invertible, so one has

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

Denoting $x^* = U\alpha^*$ and multiplying both side by U gives

$$x^* = A^{[J]}\Phi^*y - \lambda A^{[J]}D_I p_I^*. \quad (23)$$

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

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

Consider \bar{u} such that

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

and

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

We recall that

$$\begin{aligned} \tilde{\Omega}^{[J]} &= (\Phi^*\Phi A^{[J]} - \operatorname{Id})D_I, & \tilde{\Pi}^{[J]} &= \Phi^*(\Phi A^{[J]}\Phi^* - \operatorname{Id}), \\ \Omega^{[J]} &= D_J^+ \tilde{\Omega}^{[J]}, & \Pi^{[J]} &= D_J^+ \tilde{\Pi}^{[J]}. \end{aligned}$$

Remark that using equation (23), one has

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

Let denote $v = \Phi^*(\Phi x^* - y) + \lambda D_I p_I^*$. On one hand, multiplying the last expression by the pseudo-inverse D_J^+ and using the fact that $D_J^+ D_J D_J^+ = D_J^+$, one has

$$(\operatorname{Id} - D_J D_J^+) v \in \operatorname{Ker} D_J^+ = \mathcal{G}_J.$$

On another hand, using equation (22), we remark that $v \in \operatorname{Ker} U^*$. Since $\operatorname{Ker} U^* = (\operatorname{Im} U)^\perp = \mathcal{G}_J^\perp$, one has

$$v \in \mathcal{G}_J^\perp \quad \text{and} \quad (\operatorname{Id} - D_J D_J^+) v \in \mathcal{G}_J.$$

However, $D_J D_J^+ v \in \mathcal{G}_J^\perp$. Hence,

$$(\operatorname{Id} - D_J D_J^+) [\Phi^*(\Phi x^* - y) + \lambda D_I p_I^*] \in \mathcal{G}_J \cap \mathcal{G}_J^\perp = \{0\}.$$

Hence,

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

Using operator norm inequality, one has

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

By definition of \bar{u} ,

$$\|\bar{\sigma}\|_\infty \leq \min_{u \in \text{Ker } D_J} \|\Omega^{[J]} p_I^* - u\|_\infty + \frac{1}{\lambda} \|\Pi^{[J]}\|_{2,\infty} \|w\|_2.$$

Hence,

$$\|\bar{\sigma}\|_\infty \leq \mathbf{RC}(I) + \frac{1}{\lambda} \|\Pi^{[J]}\|_{2,\infty} \|w\|_2.$$

Hence, for $\mathbf{RC}(I) < 1$, σ defined by

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

and

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

one has $\|\bar{\sigma}\|_\infty < 1$ and $\|\sigma\|_\infty = \max(\|\bar{\sigma}\|_\infty, \|\sigma^*\|_\infty) \leq 1$. Using Lemma 1, the vector x^* is a solution of $\mathcal{P}_\lambda(y)$. Moreover, since $\|\bar{\sigma}\|_\infty < 1$ and (H_J) holds, x^* is the unique solution of $\mathcal{P}_\lambda(y)$ according to Lemma 3.

4. We now bound the distance between x_0 and x^* .

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

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

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

Using operator norm inequality, one has

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

□

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 regularizations.

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), however 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 regularization.

References