# Robust Multi-image Processing With Optimal Sparse Regularization

Yann Traonmilin, Saïd Ladjal, Andrés Almansa

**.... manuscript No.**
(will be inserted by the editor)

# Robust Multi-image Processing With Optimal Sparse Regularization

**Yann Traonmilin · Saïd Ladjal · Andrés Almansa**

the date of receipt and acceptance should be inserted later

**Abstract** Sparse modeling can be used to characterize outlier type noise. Thanks to sparse recovery theory, it was shown that 1-norm super-resolution is robust to outliers if enough images are captured. Moreover, sparse modeling of signals is a way to overcome ill-posedness of under-determined problems. This naturally leads to the question: does an added sparsity assumption on the signal will improve the robustness to outliers of the 1-norm super-resolution, and if yes, how strong should this assumption be? In this article, we review and extend results of the literature to the robustness to outliers of overdetermined signal recovery problems under sparse regularization, with a convex variational formulation. We then apply them to general random matrices, and show how the regularization parameter acts on the robustness to outliers. Finally, we show that in the case of multi-image processing, the structure of the support of signal and noise must be studied precisely. We show that the sparsity assumption improves robustness if outliers do not overlap with signal jumps, and determine how the regularization parameter can be chosen.

**Keywords** multi-image processing, super-resolution, outliers, sparse signal, regularization parameter

## 1 Introduction

Sparse signal approximation is a well known tool to deal with ill-conditioned inverse problems. A wide range of applications benefit from this tool. From compressive sensing (Candes et al. 2006) to image denoising (Elad and Aharon. 2006; Yu et al. 2012) and texture synthesis (Tartavel et al. 2013), the added knowledge that a signal is sparse allows for the recovery of said signal when the recovery of a non-sparse signal would have been impossible. It was shown that natural images can have sparse reprensations in some dictionaries (Mairal et al. 2009; Mallat and Yu. 2010; Benoit et al. 2011).

The sparse approximation has mainly been made on signals. It was shown that if an under-determined observation matrix has the required property (Null Space Property, Restricted Isometry Property), then perfect recovery from these observations is possible by solving a convex $L^1$ norm minimization problem. Candes and Tao (2005) show that the underdetermined sparse recovery problem is equivalent to an overdetermined outlier robustness problem. In this setting, the sparsity assumption is made on the noise which contaminates the data. Such noise is often consider as outlier, as it can represent observations unrelated to the signal. This was used to estimate the robustness to outliers of $L^1$ multi-image super-resolution (SR) in Traonmilin et al (2013). It was shown that, if enough images are available, then $L^1$-super-resolution can forgive a fraction of contaminated images in the acquired low resolution images. In Mitra et al (2013), similar theoretical developments are used and applied to the robustness to outliers of random uniform observations.

Knowing the benefits of sparsity priors, the question of imposing it on both the signal and the noise comes naturally. In this article, we will study the following problem: given an overdetermined outlier robustness problem, does sparse regularization improve the number of outliers that can be removed? This question has been studied in (Studer et al. 2012; Kuppinger et al.

Telecom Paristech LTCI
46, rue Barrault
75013 Paris
Tel.: +33145818080
E-mail: yann.traonmilin,said.ladjal,andres.almansa@telecom-paristech.fr

2012; Studer and Baraniuk. 2013) with a different (and underdetermined) observation model. It was shown that a compromise between signal and noise sparsity must be made to guarantee recovery for particular algorithms (approximate $L^0$, basis pursuit and $L^1 - L^2$) and observation matrices. However, in the variational formulation, the question of determining the optimal regularization parameter is not treated. Moreover the effect of such methods on multi-image processing is not studied.

Multi-image processing is a recent trend in computational photography. The use of information redundancy permits the design of image processing algorithms that produce images of higher quality. Multi-image denoising (or burst denoising) aims at lowering the noise in the final image. Super-resolution is the process of recovering a high resolution (HR) image from low resolution (LR) versions of it (Farsiu et al. 2004a;b; Milanfar. 2010; Tian and Ma. 2011). The $L^1$ super-resolution problem is already in use and recognized for its robustness (He et al. 2007; Yap et al. 2009). Moreover, sparse regularization is already in use for mono-image (Ning et al. 2013) super resolution. Even popular regularization methods such as total variation ($TV$) regularization (Marquina and Osher. 2008) can be interpreted as the convex version of a sparse gradient regularization. By extending results of sparse recovery theory to the case of multi-image processing, this work will give a better understanding of how these methods behave.

### 1.1 Outline and contributions

In this article, we begin with the introduction of the multi-image acquisition model in **Section 2**. This model is the justification for the more generic set-up that we use for the theoretical content of the paper. We recall the $L^1$ convex variational framework used to recover the higher quality image. In **Section 3**, we study the problem of outlier robustness (which we call forgiveness) using convex programming with sparse convex regularization. We show that forgiveness on sparse solutions is equivalent to an extension of a Non Concentration Property highlighting the distinct role of the supports of the outliers and the signal. In Studer et al (2012); Kuppinger et al (2012); Studer and Baraniuk (2013), coherency based sufficient conditions of convergence are given for variations of basis pursuit algorithm. We also show an extension of the popular Restricted Isometry Property (RIP) which respects the different roles of signal and noise sparsity. In **Section 4**, we apply our results to Gaussian random observation matrices, with 2 different regularization schemes. We show that the simplest case can be studied with existing results on compressive sensing. When the regularization parameter varies, checking our extended RIP gives a good tool to predict the qualitative behaviour of forgiveness with respect to sparse regularization. We also point out that the choice of the matrix on which we check the RIP is important, and illustrate that remark by using 2 different constructions. Finally, in **Section 5**, we study the benefits of sparse regularization for forgiveness in the area of multi-image processing. While the RIP was a good tool to study random matrices, we observe in two image processing problems (multi-image denoising and super resolution), that the convex sparse prior (the $TV$ in our case) adds some benefits only if precise requirements on the structure of the support of outliers and signal are met: signal jumps and outliers should not overlap. The structure of sparsity supports has been studied before (Jenatton et al. 2011; Bach et al. 2012) and has led to the creation of dedicated measures of sparsity (mixed $L^1, L^q$ norms) or dedicated algorithms (Yu et al. 2012). In this article, we show that, under a non-structured norm ($L^1$ norm), when the observation operator and sparsifying transform are structured, perfect recovery is possible for some structured sparsity supports. Such requirements can only be studied with characterizations more precise than the RIP such as the non concentration property. In the case of super-resolution, these requirements can be met only for particular acquisition models (finite support of the translation operator). These theoretical results are illustrated by experiments.

## 2 Multi-image acquisition Model

### 2.1 Forward model

We introduce an image generation model to describe the acquisition of a burst of images using a hand-held camera. In a finite-dimensional context, we suppose that images are generated by a linear map $A$:

$$\begin{aligned} A : \mathbb{R}^{Ml\times Ml} &\to \left(\mathbb{R}^{l\times l}\right)^N \\ u &\to (A_i u)_{i=1,N} = (SQ_i u)_{i=1,N} \end{aligned} \quad (1)$$

where $M$ is the super-resolution factor, $N$ is the number of acquired images, $l \times l$ is the size of acquired images, $u$ is a HR image of size $Ml \times Ml$, the $A_i$ are linear maps generating LR images, $S$ is the sub-sampling operator by a factor $M$ and $Q_i$ are the deformations associated with each image. SR is the process of recovering $u_0$ from $w = Au_0 + n$ ($n$ is the observation noise). In this paper, we suppose that the $Q_i$ are known. In this setting, for $M > 1$, the inversion of $A$ is called super-resolution interpolation. When $M = 1$, we will simply

talk of multi-image denoising.
It has been shown in Traonmilin et al (2012) that $A$ is almost surely full rank when motions are random compositions of translations and rotations and $N \geq M^2$.

### 2.2 Variational Formulation and previous results

When $A$ is full rank and $M^2 \leq N$, $L^2$-norm minimization guarantees that the energy of the reconstruction noise is bounded by the energy of observation noise times the operator norm of the pseudo-inverse $A^\dagger$ of $A$. This leads to useful results when observation noise has bounded energy. In the case of outliers, no assumption is made on the power of the noise and $L^2$ reconstruction does not guarantee a good result (unbounded reconstruction noise). It was shown that $L^1$-norm minimization removes outliers if the ratio of images contaminated by outliers is small enough. We write $L^1$-norm minimization of the data-fit :

$$\text{argmin}_u \|Au - w\|_1 \tag{2}$$

with $w = Au_0 + n_0$. The problem is to find conditions on $A$ ensuring that $u_0$ is the unique solution of (2) when $n_0$ is an outlying noise. Outliers have the form : $n_0 = n.T$ with $T$ a vector of 0 and 1 representing the support of the noise (the . represents the component-by-component vector product). We do not make any hypothesis on $n$. (Traonmilin et al. 2013) showed that if the number of noisy images $N_c$ fulfills the condition $N_c/N < C$ with $C$ a constant, then $u_0$ is the unique solution of Equation (2).

In Section 3, matrix $A$ will be a general full rank matrix of an over-determined system. In other sections, $A$ will be an over-determined full rank acquisition operator of size $Nl^2 \times (Ml)^2$ with $N > M^2$.

### 2.3 Sparse priors on natural images

Given an image $u$, we say that $u$ is sparse for the particular sparsifying transform $\Psi$ if $\Psi u$ is sparse. A wide category of sparse priors ($\Psi$) have been considered for natural images. Some of these are wavelet decompositions, projection in a dictionary of patches (Geiger et al. 1999; Elad and Aharon. 2006). Another frequent assumption on images is that they minimize the $L^1$-norm of the gradient (the total variation). Under the light of the sparse recovery theory, minimizing this $L^1$-norm is equivalent to imposing some sparsity on the gradient. In the part on super-resolution, we will use this prior as a practical example. We will study the minimization

$$\text{argmin}_u \|Au - w\|_1 + \lambda \|\Psi u\|_1 \tag{3}$$

where $\Psi$ is the sparsifying transform.

## 3 Forgiving Matrices and Sparse Regularization

### 3.1 Definitions

Let $\mathcal{T}$ be a family of supports, $A$ a $N \times M$ full rank matrix. From Traonmilin et al (2013), we define forgiving matrices :

**Definition 1 (Forgiving Matrix.)** Let $\mathcal{T}$ be a set of supports in $\mathbb{R}^N$ (subset of $\{0,1\}^N$).The matrix $A$ is $\mathcal{T}$-forgiving if for all $T \in \mathcal{T}, n \in \mathbb{R}^N, u_0 \in \mathbb{R}^M$, we have:

$$u_0 = \text{argmin}_u \|Au - (Au_0 + n.T)\|_1 \tag{4}$$

and $u_0$ is the unique minimizer. We say that $A$ forgives $\mathcal{T}$ errors.

Most of the time, in the literature, $\mathcal{T}$ is the set of supports with a certain cardinal $K$. They then describe $K$-sparse outliers. However, preserving arbitrary sets of supports will be crucial to the image processing part of this article. We extend this definition of forgiving matrix to the regularized case. Let the sparsifying transform $\Psi$ be a matrix of size $N' \times M$. We want to study the case when $A$ is $\mathcal{T}$-forgiving on regular solutions $u_0$ where $\Psi u_0$ is $\mathcal{L}$-sparse for some supports of signal sparsity $\mathcal{L}$. Ideally, we would like to extend the concept as

**Definition 2 ($(L_1, L_0)$ Forgiving Matrix Under a Sparse Hypothesis)** Let $\mathcal{T}$ be a set of supports in $\mathbb{R}^N$ and $\mathcal{L}$ a set of supports in $\mathbb{R}^{N'}$. The matrix $A$ is $\mathcal{T}$-forgiving on $\Psi, \mathcal{L}$-sparse solutions if for all $T \in \mathcal{T}, n \in \mathbb{R}^N, u_0 \in \mathbb{R}^M$ such that $\Psi u_0 = \alpha_0 \in \mathcal{L}.\mathbb{R}^{N'}$, we have:

$$u_0 = \text{argmin}_u \|Au - (Au_0 + n.T)\|_1 \text{ s.t. } \Psi u \in \mathcal{L}.\mathbb{R}^{N'} \tag{5}$$

and $u_0$ is the unique minimizer.

However, the lack of practical algorithm to solve exactly such a problem drives us to study a convexified version of it. In this paper we will consider the $L^1, L^1$ sparse regularization.

**Definition 3 ($(L_1, L_1)$ Forgiving Matrix Under a Sparse Hypothesis.)** Let $\lambda \in \mathbb{R}$. Let $\mathcal{T}$ be a set of supports in $\mathbb{R}^N$ and $\mathcal{L}$ a set of supports in $\mathbb{R}^{N'}$. We say that $A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse solutions by convex programming if for all $T \in \mathcal{T}, n \in \mathbb{R}^N, u_0 \in \mathbb{R}^M$ such that $\Psi u_0 = \alpha_0 \in \mathcal{L}.\mathbb{R}^{N'}$, we have:

$$u_0 = \text{argmin}_u \|Au - (Au_0 + n.T)\|_1 + \lambda \|\Psi u\|_1 \tag{6}$$

and $u_0$ is the unique minimizer.

*Remark 1* When performing, minimization (6). We call $\lambda$, the regularization parameter. Usually, this parameter is considered as the level of regularization, *i.e.* the amount of regularity we want to impose the on the image.

*Remark 2* Equation (6) is equivalent to: $u_0$ is the unique minimizer of

$$\min_u \|A_r^\lambda u - w_r\|_1 \tag{7}$$

where $A_r^\lambda = \begin{pmatrix} A \\ \lambda\Psi \end{pmatrix}$ and $w_r = \begin{pmatrix} Au_0 \\ \lambda\Psi u_0 \end{pmatrix} + \begin{pmatrix} n.T \\ -\lambda\Psi u_0 \end{pmatrix} = \begin{pmatrix} Au_0 + n.T \\ 0 \end{pmatrix}$

*Remark 3* From the previous remark, if there is a $\lambda$, such that $A_r^\lambda$ is forgiving outliers with supports in $\mathcal{T}' = \left\{ \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}, \tau_1 \in \mathcal{T}, \tau_2 \in \mathcal{L} \right\}$ (We say then that $A_r^\lambda$ is $\mathcal{T}'$-forgiving or $\mathcal{T}, \mathcal{L}$-forgiving), then $A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$ sparse solutions. In the next section, we will see that the converse is true.

**Definition 4 (Non-Concentration Property.)** Let $\mathcal{T}$ be a set of supports in $\mathbb{R}^N$ and $V$ a subspace of $\mathbb{R}^N$. We say that $V$ has the $\mathcal{T}$-Non-Concentration Property (NCP) if for all $v \in V \backslash \{0\}$ and all $T \in \mathcal{T}$,

$$\|v.T\|_1 < \|v.T^c\|_1 \tag{8}$$

where $T^c$ stands for the complement support of $T$.

We say that a matrix has the $\mathcal{T}$-Null Space Property ($\mathcal{T}$-NSP) if its null space has the $\mathcal{T}$-NCP.

Notice that, given the finite-dimensional setting, the NCP property implies the existence of a constant $\gamma < 1$ such that for all $v \in V$ and all $T \in \mathcal{T}$:

$$\|v.T\|_1 \leq \gamma \|v.T^c\|_1 \ . \tag{9}$$

This constant is called the NSP constant in the area of sparse recovery.

When $\mathcal{T}$ and $\mathcal{L}$ are families of supports of cardinal $K$ and $S$ respectively. We talk about $K, S$-forgiveness and $K, S$-NCP. As explained in Candes and Tao (2005), forgiveness of a matrix $A$ is equivalent to the sparse recovery property of matrices $B$ anihilating $A$.

**Definition 5 (Annihilating matrix.)** Let $A$ be a $N \times M$ matrix. We say that $B$ is an annihilator of $A$ or annihilates $A$ if $B$ has size $(N - M) \times N$ , is full rank and $BA = 0$.

**Definition 6 (Sparse Recovery Property.)** Let $\mathcal{T}$ be a set of supports in $\mathbb{R}^N$ (subset of $\{0, 1\}^N$).The matrix $B$ has the $\mathcal{T}$-sparse recovery property if for all $T \in \mathcal{T}, x_0 \in \mathbb{R}^N$, we have:

$$x_0 = \mathrm{argmin}_x \|x\|_1 \ s.t. \ Bx = Bx_0 \tag{10}$$

and $x_0$ is the unique minimizer. We say that $B$ is $\mathcal{T}$-sparse capable.

Moreocer forgiveness of $a$ is equivalent to the NCP of the space spanned by $A$. The NCP of a set is also equivalent to the sparse recovery property of matrices $B$ having this set for kernel:

**Theorem 1** *For $\lambda = 0$, the forgiveness of $A$ is equivalent to the sparse recovery propery of the annihilators of $A$.*

*Proof* See Candes and Tao (2005) or use Theorem 3. □

## 3.2 Exact Recovery Property

This property can be found in Candes and Tao (2006). It is useful to study a particular instance of a forgiveness problem. For completeness of the paper, we produce a direct proof for the forgiveness problem

**Theorem 2 Exact Reconstruction Property (ERP)** *Let $A$, $T$, $n$, $u_0$ be an instance of the problem (2). A sufficient condition for $u_0$ to be the unique minimizer is the existence of a vector $v \in \mathbb{R}^N$ such that :*

1. *$v \in \ker A^H$ (kernel of $A^H$)*
2. *$v.T = sign(n.T)$, $\|v.T^c\|_\infty < 1$*

*Proof* Because we can always make the change of variables $u' = u - u_0$, it suffices to verify that:

$$\|Au - n.T\|_1 \geq \|n.T\|_1 \tag{11}$$

Using the properties of the infinity norm of $v$, we have:

$$\|Au - n.T\|_1 \geq < v, n.T - Au > \tag{12}$$

which we can decompose:

$$\|Au - n.T\|_1 \geq < v, n.T > - < v, Au > \tag{13}$$

$$\|Au - n.T\|_1 \geq \|n.T\|_1 - < A^H v, u > \tag{14}$$

with $A^H v = 0$ which leads to the desired result. □

## 3.3 Characterization of Forgiveness by the Non-Concentration Property

We will extend the following theorem, (a direct demonstration can be found in Traonmilin et al (2013)).

**Theorem 3** *The two following propositions are equivalent:*

1. *$A$ is $\mathcal{T}$-forgiving*
2. Im$A$ *(the image of $A$) has the $\mathcal{T}$-Non Concentration Property.*

**Theorem 4** *If $N' = M$ and $\Psi$ is invertible, the three following propositions are equivalent (recall that $\mathcal{T}'$ is set of the concatenations of the supports from $\mathcal{T}$ and $\mathcal{L}$):*

1. *$A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse solutions.*
2. *$A_r^\lambda$ is the $\mathcal{T}'$-forgiving.*
3. *$\mathrm{Im}A_r^\lambda$ has the $\mathcal{T}'$-Non Concentration Property (or $\mathcal{T}, \mathcal{L}$-NCP).*

*Proof* 2 $\Leftrightarrow$ 3: We use directly Theorem 3.

3 $\Rightarrow$ 1: Using Theorem 3, $\mathrm{Im}A_r^\lambda$ has the $\mathcal{T}'$-NCP implies that $A_r^\lambda$ is $\mathcal{T}'$-forgiving. In particular, for every $w_r = \begin{pmatrix} Au_0 + n.T \\ 0 \end{pmatrix}$ defined as in Remark 2. Consequently, problem (6) has $u_0$ as a unique solution for each $u_0$,$\mathcal{T}'$ which is equivalent to the forgiveness on $\lambda\Psi, \mathcal{L}$-sparse solutions.

1 $\Rightarrow$ 2: Let $A$ be $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse solutions , $L \in \mathcal{L}$ and $T \in \mathcal{T}$. Let $u_0 \in \mathbb{R}^N$, let $n_1 \in \mathbb{R}^N, n_2 \in \mathbb{R}^{N'}$. Let $u_1 = \Psi^{-1}(n_2.L)$. Using the forgiveness of $A$ on $\lambda\Psi, \mathcal{L}$-sparse solutions, $u_1$ is the unique minimizer of

$$u_1 = \mathrm{argmin}_u \|Au - (Au_1 + n_1.T)\|_1 + \lambda\|\Psi u\|_1. \quad (15)$$

We make the change of variable $u = v + u_1 - u_0$. Then $u_0$ is the unique minimizer:

$$\begin{aligned} u_0 &= \\ &\mathrm{argmin}_v \|Av - (Au_0 + n_1.T)\|_1 + \lambda\|\Psi(v + u_1 - u_0)\|_1 \\ u_0 &= \\ &\mathrm{argmin}_v \|Av - (Au_0 + n_1.T)\|_1 + \lambda\|\Psi v - \Psi u_0 - n_2.L\|_1 \end{aligned} \quad (16)$$

Consequently, $A_r^\lambda$ is $\mathcal{T}'$-forgiving. □

*Remark 4* If $\Psi$ is a generic decomposition in an overdetermined dictionary, only the sufficient condition 3 $\Leftrightarrow$ 2 $\Rightarrow$ 1 holds.

*Remark 5* From this theorem, the forgiveness of $A_r^\lambda$ is guaranteed when $A$ is $\mathcal{T}$-forgiving and $\Psi$ is $\mathcal{L}$-forgiving.

A simple application of this theorem is the behaviour of the regularization when $\lambda \to 0$.

**Proposition 1** *Let us suppose that $A$ is $K$-forgiving. Then for $\lambda$ sufficiently small $A_r^\lambda$ is $K$ forgiving on sparse solutions for any sparsity.*

*Proof* Let $T' = \begin{pmatrix} T \\ L \end{pmatrix} \in \mathcal{T}'$. Let $\lambda > 0$ and $u$ with $\|u\|_1 = 1$ . Because $\mathrm{Im}A$ has the $K$-NCP with constant $\gamma$, we have:

$$\begin{aligned} \|(Au).T\|_1 + \lambda\|(\Psi u).L\|_1 &< \gamma\|(Au).T^c\|_1 + \lambda\|(\Psi u).L\|_1 \\ \|(A_r^\lambda u).T'\|_1 &< \gamma\|(Au).T^c\|_1 + \lambda(\|(\Psi u).L\|_1 \\ &\quad - \|(\Psi u).L^c\|_1) + \lambda\|(\Psi u).L^c\|_1 \end{aligned} \quad (17)$$

If $(\|(\Psi u).L\|_1 - \|(\Psi u).L^c\|_1) < 0$, $\mathrm{Im}A_r^\lambda$ has the $T, L$-NCP. If not, for $\lambda$ sufficiently small and independent from $u$ ($\Psi$ is a finite dimensional linear operator), we have

$$\gamma\|(Au).T^c\|_1 + \lambda(\|(\Psi u).L\|_1 - \|(\Psi u).L^c\|_1) < \|(Au).T^c\|_1 \quad (18)$$

and $\mathrm{Im}A_r^\lambda$ has the $T, L$-NCP. The conclusion follows from Theorem 4. □

With the non concentration property, we can reject cases where sparse regularization will not improve forgiveness.

**Proposition 2** *If $\Psi$ does not have the $S$-NCP. Then for $\lambda$ large enough, $A_r^\lambda$ does not have the $K, S$-NCP.*

*Proof* Let $T, L$ be supports of cardinal $K$ and $S$ respectively. There is a $u$ such that $\|(\Psi u).L\|_1 > \|\Psi u.L^c\|_1$. Let $0 < a < \|(\Psi u).L\|_1 - \|\Psi u.L^c\|_1$. Then,

$$\begin{aligned} \|(Au).T\|_1 + \lambda\|(\Psi u).L\|_1 >& \|(Au).T\|_1 + \lambda a \\ &+ \lambda\|(\Psi u).L^c\|_1. \end{aligned} \quad (19)$$

For $\lambda$ large enough, $\|(Au).T\|_1 + \lambda a > \|(Au).T^c\|_1$ and

$$\|(Au).T\|_1 + \lambda\|(\Psi u).L\|_1 > \|(Au).T^c\|_1 + \lambda\|\Psi u.L^c\|_1 \quad (20)$$

and $A_r^\lambda$ does not have the $K, S$-NCP. □

### 3.4 A sufficient condition of forgiveness: the Restricted Isometry Property

The Restricted Isometry Property introduced by (Candes and Tao. 2005; Candès. 2008) can be a convenient way to ensure forgiveness of some matrices. This sufficient property can be verified for a matrix $B$ annihilating $A$. It will then guarantee that $B$ is sparse capable, and give the forgiveness of $A$ with Theorem 1 The RIP is designed for supports with a particular cardinality. In order to study the regularized problem and its robustness to outliers, we need the RIP property for supports of the form $T = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}$ with cardinality $|\tau_1| = K$ on the first $N$ coordinates and $|\tau_2| = S$ on the last $N'$ coordinates. We say that such $T$ has cardinal $K, S$. Following definition 5, the matrix $B$ (annihilating $A$) has size $N \times (N + N')$. We will write $B_T$ the restriction of any matrix $B$ to the columns matching the support $T$ ($B_T x = B(x_T)$) where $x_T$ is $x$ zeroed outside of $T$.

*3.4.1 Extension of the RIP to the regularized case*

**Definition 7** $B$ has the restricted isometry property of order $K, S$ and constant $\delta > 0$ if for all $x \in \mathbb{R}^{N+N'}$, for all supports $T$ with cardinal $K, S$

$$(1-\delta)\|x_T\|_2 \leq \|B(x_T)\|_2 \leq (1+\delta)\|x_T\|_2 \ . \tag{21}$$

The usual demonstration (Candes and Tao. 2005; Cohen et al. 2009) proves that the RIP implies the non concentration property, which implies the forgiveness. Here, we show that it is important to distinguish the sparsity of the signal and of the outliers. We kept the simplest demonstration to obtain the qualitative result. Better constant can be obtained with slightly more complicated arguments.

**Theorem 5** *Let $B$ be an annihilator of $A$. If $B$ has RIP of order $3K, 3S$ with constant $\delta \leq \frac{\sqrt{2}-1}{\sqrt{2}+1}$, then* $\ker B$ *has NCP of order $K, S$ and $A$ is $K, S$ forgiving.*

*Proof* We adapt the proof from Cohen et al (2009). We prove the NCP of $\ker B$. Let $x \in \ker B$. Let $T_0$ be the support with cardinal $K, S$ selecting the $K$ biggest absolute values of $x$ on the first $N$ coordinates and $S$ biggest absolute values of $x$ on the last $N'$ coordinates. Let $T_1 ... T_r$ be the sequence of support of cardinal $2K, 2S$ having the next biggest elements. Let $x_0 = x_{T_0} + x_{T_1}$, then $Bx_0 = -B(\sum_{i=2}^r x_{T_i})$ because $x \in \ker B$. It is sufficient to verify the NCP for $T = T_0$ because it concentrates the most energy of $x$. Using Cauchy-Schwartz inequality (or norm equivalence) $\|x_T\|_1 \leq K^{1/2}\|x_{T,1}\|_2 + S^{1/2}\|x_{T,2}\|_2$ where $x_{T,1}$ is $x$ restricted to first $N$ coordinates, and $x_{T,2}$ is $x$ restricted to the last $N'$ coordinates

Because $x_{0,1}$ and $x_{0,2}$ are $3K, 3S$-sparse, we use the definition of $x_0$ followed by the RIP hypothesis and the triangle inequality :

$$\begin{aligned}
\|x_T\|_2 &\leq K^{1/2}\|x_{T,1}\|_2 + S^{1/2}\|x_{T,2}\|_2 \\
&\leq K^{1/2}\|x_{0,1}\|_2 + S^{1/2}\|x_{0,2}\|_2 \\
&\leq (1-\delta)^{-1}(K^{1/2}\|Bx_{0,1}\|_2 + S^{1/2}\|Bx_{0,2}\|_2) \\
&\leq (1-\delta)^{-1}\left(K^{1/2}\|B\left(\sum_{i=2}^r x_{T_i,1}\right)\|_2 \right. \\
&\qquad \left. + S^{1/2}\|B\left(\sum_{i=2}^r x_{T_i,2}\right)\|_2\right) \\
&\leq (1-\delta)^{-1}\left(\sum_{i=2}^r (K^{1/2}\|Bx_{T_i,1}\|_2 + S^{1/2}\|Bx_{T_i,2}\|_2)\right)
\end{aligned} \tag{22}$$

We use the RIP again:

$$\|x_T\|_2 \leq \frac{1+\delta}{1-\delta}\left(\sum_{i=2}^r (K^{1/2}\|x_{T_i,1}\|_2 + S^{1/2}\|x_{T_i,2}\|_2)\right) \tag{23}$$

Now, we bound the right side with $\|x\|_1$. Let $j \geq 1$ and $y$ (respectively $y'$) be one of the first $N$ coordinate of $x_{T_{j+1}}$ (respectively $x_{T_j}$). Then $|y| \leq |y'|$. Let $z$ (respectively $z'$) be one of the last $N'$ coordinate of $x_{T_{j+1}}$ (respectively $x_{T_j}$). Then $|z| \leq |z'|$. From this observation,

$$\begin{aligned}
2Ky &\leq \|x_{T_j,1}\|_1 \\
2Sz &\leq \|x_{T_j,2}\|_1 .
\end{aligned} \tag{24}$$

We square and sum over $y$ and $z$ :

$$\begin{aligned}
2K\|x_{T_{j+1},1}\|_2^2 &\leq \|x_{T_j,1}\|_1^2 \\
2S\|x_{T_{j+1},2}\|_2^2 &\leq \|x_{T_j,2}\|_1^2
\end{aligned} \tag{25}$$

We take the square root and add these inequalities:

$$\sqrt{2}(K^{1/2}\|x_{T_{j+1},1}\|_2 + S^{1/2}\|x_{T_{j+1},2}\|_2) \leq \|x_{T_j}\|_1 \tag{26}$$

We use this result with equation (23) :

$$\|x_T\|_1 \leq \frac{1+\delta}{1-\delta}\frac{1}{\sqrt{2}}\|x_{T^c}\|_1 \tag{27}$$

which is the NCP if $\frac{1}{\sqrt{2}}\frac{1+\delta}{1-\delta} \leq 1$. This is equivalent to $\delta \leq \frac{\sqrt{2}-1}{1+\sqrt{2}}$ . □

*Remark 6* This theorem could be easily extended to matrices $B$ having more than 2 blocks, *i.e.* for signal corrupted by sparse noise in a union of dictionaries $A_r$.

*3.4.2 The choice of the annihilating matrix*

The quality of the RIP is driven by the conditioning of sub-matrices of $B$. To our knowledge, the choice of the $B$ such that $\ker B = \mathrm{Im} A$ and $B$ has the RIP with best constant and/or order, is an open problem. This leads to some other questions. Given a matrix $C$, $C$ might not be the best candidate to verify the RIP for its own sparse recovery property. However, we can easily construct a $B$ from $A$ by taking the orthogonal projection on the orthogonal of the image of $A$ ($P_{(\mathrm{Im}A)^\perp}$) and restricting it to its image. Then $B$ is an annihilator of $A$. Also, we can construct a family of annihilators of $A_r$ using an annihilator of $A$.

**Proposition 3** *Let $\epsilon => 0$. Let $\Phi = (\Psi^H\Psi)^{-1}\Psi$. Let $B_r = \left(B_1, -\lambda^{-1}A\Phi\right)$, with $B_1 = \epsilon P_{(\mathrm{Im}A)^\perp} + I$ ($I$ is the identity matrix). Then $B_r$ is an annihilator of $A_r^\lambda$.*

*Proof* First, we remark that $B_2 = -\lambda^{-1}B_1 A\Phi$ implies $B_r A_r^\lambda = 0$. Thus $B_r = \left(B_1, -\lambda^{-1}B_1 A\Phi\right)$ is sufficient. Moreover $B_1$ is full rank and $B_1 A = A$. □

*Remark 7* We wanted to construct $B_r$ from $P_{(\mathrm{Im}A)^\perp}$. We added $I$ to ensure that $B_1$ is full rank.

In Section 4.3, we show with an example that different classes of annihilating matrices lead to different RIP constants.

*3.4.3 A sufficient condition for the RIP of the annihilating matrices*

Here we show that, with our particular construction of annihilating matrices, we can find sufficient conditions for the $K,S$-RIP of $A_r^\lambda$. Let $B_r = \begin{pmatrix} B_1 & B_2 \end{pmatrix}$ annihilating $A_r^\lambda$, with $B_1$ a $N \times N$ matrix and $B_2$ a $N \times N'$ matrix. We ask ourselves when $B_r$ has RIP.

**Definition 8 (Operator bounds on sparse vectors.)** $B$ is bounded by $\sigma_m, \sigma_M$ on $K$ sparse vectors if :

$$\sigma_M = \sup_{x,|T| \leq K} \frac{\|Bx_T\|_2}{\|x_T\|_2} \tag{28}$$

$$\sigma_m = \inf_{x,|T| \leq K} \frac{\|Bx_T\|_2}{\|x_T\|_2} \tag{29}$$

**Proposition 4 (Normalization.)** *If $\sigma_m < \frac{\|B_x\|_2}{\|x\|_2} < \sigma_M$ for $x$ $K,S$ sparse, there is an $\eta$ such that $\eta B$ has RIP with constant $\delta = \frac{\sigma_M - \sigma_m}{\sigma_M + \sigma_m}$. We call $\eta B$ normalized $B$.*

*Proof* If we let $\eta = 2/(\sigma_M + \sigma_m)$ and $\delta = \frac{\sigma_M - \sigma_m}{\sigma_M + \sigma_m}$, then $\eta\sigma_m = 1 - \delta$ and $\eta\sigma_M = 1 + \delta$. □

The mutual coherence between $B_1$ and $B_2$ (Similarly as in Candès and Romberg (2007)) is useful to produce a deterministic sufficient condition.

**Proposition 5 (Sufficient condition for the RIP of the annihilator of $A_r^\lambda$.)** *Let $B_1$ bounded by $\sigma_{m_1}, \sigma_{M_1}$ on $K$-sparse vectors and $B_2$ bounded by $\sigma_{m_2}, \sigma_{M_2}$ on $S$-sparse vectors. Let $\sigma_m = \min(\sigma_{m_1}, \sigma_{m_2})$ and $\sigma_M = \max(\sigma_{M_1}, \sigma_{M_2})$. Let $\mu = \sup_{|T| \leq K, |L| \leq S} \|B_{1,T}^H B_{2,L}\|_{op}$. If $\mu < \sigma_m^2$, then normalized $B_r$ has RIP of order $K,S$ and constant $\delta' = \frac{\sqrt{\sigma_M^2 + \mu} - \sqrt{\sigma_m^2 - \mu}}{\sqrt{\sigma_M^2 + \mu} + \sqrt{\sigma_m^2 - \mu}} < 1$.*

*Proof* Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ($x_1$ has size $N$) be a vector supported on $T$ with cardinal $K,S$.

$$\begin{aligned} \|Bx\|_2^2 &= \|B_1x_1 + B_2x_2\|_2^2 \\ &= \|B_1x_1\|_2^2 + \|B_2x_2\|_2^2 + 2 < B_1x_1, B_2x_2 > \end{aligned} \tag{30}$$

The mutual coherence $\mu$ between $B_1$ and $B_2$ plays an important role in this equation:

$$| < B_1x_1, B_2x_2 > | \leq \mu\|x_1\|_2\|x_2\|_2 \leq \frac{\mu}{2}\|x\|_2^2 \tag{31}$$

because $ab \leq (a^2 + b^2)/2$. The first two terms are controlled by the bounds of $B_1$ and $B_2$. We have :

$$\begin{aligned} \min(\sigma_{m_1}, \sigma_{m_2})^2 - \mu \leq & \frac{\|Bx\|_2^2}{\|x\|_2^2} \leq \max(\sigma_{M_1}, \sigma_{M_2})^2 + \mu \\ \sqrt{\sigma_m^2 - \mu} \leq & \frac{\|Bx\|_2}{\|x\|_2} \leq \sqrt{\sigma_M^2 + \mu} \end{aligned} \tag{32}$$

The conclusion follows from proposition 4. □

*3.4.4 On the choice of $\lambda$*

The parameter $\lambda$ represents the level of a priori knowledge on the signal that we want to use in the minimization of the considered problem. The question of chosing $\lambda$ in different settings (e.g. $L^2, L^2$ minimization $L^2, L^1$ minimization) is studied intensively in the area of image processing. A way to perform this choice is to set an objective function (generally a risk function) and find the $\lambda$ which minimize the risk (Vaiter et al. 2012). Here, the case of noiseless recovery is studied. Consequently, two questions come naturally:

– Given $\lambda$ , which couples $K,S$ lead to to a perfect recovery.
– Given $K,S$, can we find $\lambda$ leading to a perfect recovery (*i.e.* the problem is $K,S$-forgiving for some $\lambda$).

Ideally, we would need to choose $\lambda$ to optimize the NCP constant or to guarantee the ERP. Because the NCP is an equivalent characterization of forgiveness, this would associate each couple $K,S$ with an optimal NCP constant (when $\lambda$ varies). If this constant is lower than 1 then recovery would be possible. However, this is difficult to determine in general as $L^1$ conditioning of sub-matrices of $A_r^\lambda$ will intervene. In specific image processing problems, it might be possible to estimate directly this constant. In some simple set-up, we can study exactly the NCP and ERP with respect to $\lambda$ (Section 5).

A first step is to try to optimize the RIP, which will give a sufficient conditioning for $K,S$ forgiveness. We show that we can guarantee the RIP for some $\lambda$ with our construction of annihilating matrix for a certain class of $A_r^\lambda$. However, as said before, because we do not know the optimal annihilator (in terms of RIP), this result has limited practical use, but highlights how blocks of $B$ interact. Later experiments will show that testing directly the RIP of some annihilators lead to the qualitative behaviour of some random matrices.

Next, we study how, from $A_r^\lambda$ we can calculate the RIP constant $\delta'$ for a particular $B_r$. Which in turn, permits to derive sufficient condition on $\lambda$ to guarantee the $K,S$-forgiveness of $A_r^\lambda$.

**Theorem 6** *We consider a regularized $A_r^\lambda$ and the corresponding $B_r$ (constructed as in Proposition 3) with $\epsilon = 0$: $B_r = (I, -\lambda^{-1}A\Phi)$. Let $A' = A\Phi$. We suppose that $A'$ is bounded by $\sigma_m(A')$, $\sigma_M(A')$ on $S$-sparse vectors. Let $\mu_1 = \sup_{|T|<K,|L|<S} \|I_T^H A'_L\|_2$.*

*If $\mu_1^2 < \sigma_m(A')^2$, then we can find $\lambda$ sucht that $\frac{\mu_1}{\sigma_m(A')^2} < \lambda^{-1} < \frac{1}{\mu_1}$. With such $\lambda$, matrix $B_r$ (normalized) has RIP of order $K, S$ and constant $\delta' < 1$.*

*In particular for $\lambda^{-1} = 1/\sigma_m(A')$, we have $\delta' \leq \frac{\sqrt{\kappa^2+\mu_1}-\sqrt{1-\mu_1}}{\sqrt{\kappa^2+\mu_1}+\sqrt{1-\mu_1}}$ where $\kappa = \sigma_M(A')/\sigma_m(A')$*

*Proof* Let $\sigma_m = \min(\sigma_{m_1}, \sigma_{m_2})$, $\sigma_M = \max(\sigma_{M_1}, \sigma_{M_2})$ as defined in Proposition 5. Because $B_1 = I$, $\sigma_{m_1} = \sigma_{M_1} = 1$ :

$$\sigma_m = \min(1, \lambda^{-1}\sigma_m(A')) \tag{33}$$

and

$$\sigma_M = \max(1, \lambda^{-1}\sigma_M(A')) \tag{34}$$

To meet the hypothesis of Proposition 5, we need $\mu = \lambda^{-1}\mu_1 < \sigma_m^2$. This is true if the following inequalities are true :

$$\begin{aligned} &\lambda^{-1}\mu_1 < 1 \\ &\lambda^{-1}\mu_1 < \lambda^{-2}\sigma_m(A')^2 \end{aligned} \tag{35}$$

which are equivalent to

$$\begin{aligned} \lambda^{-1} &< \frac{1}{\mu_1} \\ \frac{\mu_1}{\sigma_m(A')^2} &< \lambda^{-1} \end{aligned} \tag{36}$$

Such $\lambda^{-1}$ exist if:

$$\mu_1^2 < \sigma_m(A')^2 \tag{37}$$

Then , with Proposition 5, $B_r$ has RIP of order $K, S$ and constant $\delta' < 1$.

If $\lambda^{-1} = 1/\sigma_m(A')$ , we have $\sigma_m = 1$ , $\sigma_M = \kappa$ and $\mu$. Using these values :

$$\delta' \leq \frac{\sqrt{\kappa^2+\mu_1}-\sqrt{1-\mu_1}}{\sqrt{\kappa^2+\mu_1}+\sqrt{1-\mu_1}} \tag{38}$$

□

*Remark 8* We showed that given a particular family of $B_r$, we can sometimes determine which one gives a RIP.

*Remark 9* The condition $\mu_1^2 < \sigma_m(A')^2$ is strong. Some matrices verify this : take for example the matrix $B_2 = \begin{pmatrix} I \\ I \end{pmatrix}$.

Using the explicit value of $\lambda$ from this theorem. Its hypotheses require that $A'$ has a well conditioned submatrices, and small norm (*i.e.* small mutual coherence between $A$ and $\Phi$ ) if $\kappa = 1.25$ and $\mu_1 = 0.1$, $\delta' \leq \frac{\sqrt{2}-1}{\sqrt{2}+1}$.

## 4 The case of random matrices

We study the case of convex sparse regularization and robustness for random matrices. First, we suppose that $A_r^\lambda = \begin{pmatrix} A \\ \lambda\Psi \end{pmatrix}$ where $A$ and $\Psi$ are a $N \times M$ and $N' \times M$ Gaussian random matrices. Let $\mathcal{T}$ be the set of supports $T \subset \mathbb{R}^N$ with $|T| = K$ and $\mathcal{L} \subset \mathbb{R}^{N'}$ the set of supports $L$ with $|L| = S$. In Section 4.3, the regularization part $\Psi$ will be identity.

### 4.1 Review of existing bounds for $\lambda = 1$

Let $B_r$ be an annihilator of $A_r^\lambda$ In Candes and Tao (2005), Candes and Tao argue that we can think of $B$ as a random Gaussian matrix. Matrix $B$ has size $N - M \times N$ ($\mathrm{dimIm}A = \mathrm{dimker}B = M$). Then $B$ has the $K$-sparse recovery capability and $A$ is $K$-forgiving with overwhelming probability if :

$$K < C\frac{N-M}{log\frac{N}{N-M}} \tag{39}$$

with $C$ a constant. We want to see how this condition evolves when we add regularization. We use the following proposition:

**Proposition 6** *If $A_r^\lambda$ is $K+S$-forgiving for any sparsity, then $A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse vectors.*

*Proof* We show the NCP. $\mathrm{Im}A_r^\lambda$ has the $K+S$-NCP. Moreover the set of supports of size $K+S$ contains $(\mathcal{T}, \mathcal{L})$. Consequently, $\mathrm{Im}A_r^\lambda$ has the $(\mathcal{T}, \mathcal{L})$-NCP and $A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse vectors. □.

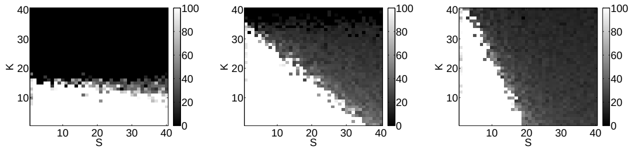
$A_r^\lambda$ has size $N + N' \times M$. We see how the bound for random matrix is changed by convex sparse regularization.

**Proposition 7** *$A$ is $\mathcal{T}$-forgiving on $\lambda\Psi, \mathcal{L}$-sparse vectors (or $A_r^\lambda$ is $K, S$ forgiving) with overwhelming probability if:*

$$K + S < C\frac{N+N'-M}{log\frac{N+N'}{N+N'-M}} \tag{40}$$

*Proof We use equation* (39) *with the dimensions of $A_r^\lambda$ and apply Proposition 6.* □

This study is only theoretical as taking the same kind of basis for observation and regularization is not realistic. However, it shows an example where adding regularization adds forgiveness power. The main point is that $K+S$ is what drives the forgiveness power, i.e. when there is more outliers, the signal needs to be the more sparse according to this bound. This is confirmed by the experiment from Figure 1, where we generated




(a) (b) (c)

**Fig. 1** *A* **random,** *Ψ* **random.** Worst reconstruction PSNR when $A_r^\lambda$ has size $N = 100$, $M = 50$ w.r.t to signal and noise sparsity $K, S$ for (a) $\lambda = 0.1$, (b) $\lambda = 1$ and (c) $\lambda = 2$.

random forgiveness experiments for different values of $K, S$. We show the worst reconstruction Peak Signal to Noise Ratio (PSNR) in decibels (dB) of the result of the minimization of problem (7) (using $\ell_1 - MAGIC$ , the software used in Candes and Tao (2006)).

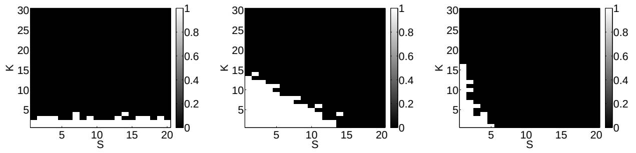
This review showed that $K+S$ is the limiting factor for $\lambda = 1$. However, each line of $A_r^\lambda$ in this case has the same energy. Thus checking $K+S$ in this non-uniform RIP leads to the same result as checking the $K, S$ RIP.

## 4.2 $K, S$-forgiveness with respect to $\lambda$

We generate $K, S$-forgiveness in Figure 1. 100 random forgiveness experiments using for each experiment a Gaussian random matrix $A$ of size $100 \times 50$ and a Gaussian random matrix $\Psi$ of size $50 \times 50$. The PSNR of the worst experiment for each $K, S$ is displayed. We observe that for each $\lambda$, we have a constraint of the type $K + c(\lambda)S < C_0$ where $c(\lambda)$ is a constant depending on $\lambda$. Qualitatively, for large $\lambda$ much more outliers are removed, but this works for very sparse signals. For small $\lambda$, improvement with respect to the non regularized case is less strong but works for signals that are less sparse.

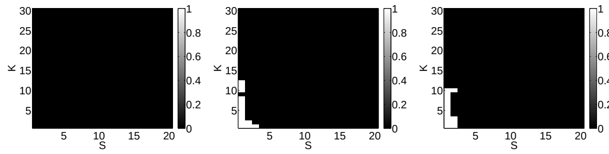
This is confirmed by our theoretical analysis with the $K, S$-RIP. We show experimentally that the $K, S$-RIP gives a qualitative way to check if the matrix is $K, S$-forgiving. Let $A$ and $\Psi$ be Gaussian random matrices. We show an estimation of the RIP of $P_{(\mathrm{Im} A_r^\lambda)^\perp}$. In Figure 2, we generate the same number of experiments, and check if the worst conditioning of one sub-matrix per experiment has the required RIP constant. We show when the RIP is verified for different $\lambda$. The union of all the couple $K, S$ shows where we can have a perfect reconstruction with sparse regularization (provided we can find $\lambda$). The shape of the zone of the $K, S$ couples leading to a perfect reconstruction follows the same behaviour as in Figure 1. As expected, a quantitative gap exists between the characterization by the RIP and the recovery experiments.

We studied the case of purely random matrices to make a link with results from the literature. Next, we



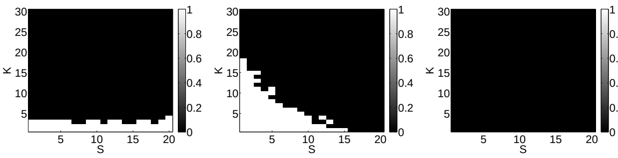

(a) (b) (c)

**Fig. 2** *A* **random,** *Ψ* **random.** Signal and noise sparsity $K, S$ leading to the RIP when $A_r^\lambda$ is a random observation with random regularization matrix with $N = 100$, $M = 50$ for (a) $\lambda = 0.1$, (b) $\lambda = 1$ and (c) $\lambda = 2$.




(a) (b) (c)

**Fig. 3** *A* **random,** $\Psi = I$**.** Signal and noise sparsity $K, S$ leading to the $K, S$-RIP of $B_r = (I, -\lambda^{-1}A)$ with $N = 100$, $M = 50$ for (a) $\lambda = 0.1\sqrt{M}$ ,(a)$\lambda = \sqrt{M}$ and (c) $\lambda = 2\sqrt{M}$.




(a) (b) (c)

**Fig. 4** *A* **random,** $\Psi = I$**.** Same experiment as Figure 3 using a different annihilator. Signal and noise sparsity $K, S$ leading to the $K, S$-RIP with $N = 100$, $M = 50$ for (a) $\lambda = 0.1\sqrt{M}$, (a) $\lambda = \sqrt{M}$ and (c) $\lambda = 2\sqrt{M}$.

study another type of random matrices where only the sparsifying transform of the signal is random.

## 4.3 The case of random observation from a sparse regularization space

Here we consider $A_r^\lambda = \begin{pmatrix} A \\ \lambda I \end{pmatrix}$. We perform the same experiments as in the previous section. Using the construction of $B_r$ from Section 3.4.3 for $\epsilon = 0$, we have $B_1 = I$ and $B_2 = -\lambda^{-1}A$. For each $\lambda$, we check the couples, $K, S$, for which, $B_r$ has RIP in Figure 3. In Figure 4, we make the same experiment using $P_{(\mathrm{Im} A_r^\lambda)^\perp}$. It illustrates the fact that different matrices having the same kernel lead to different RIP. In conclusion, the RIP should only be used as a qualitative measure of how the recovery behaves, and the annihilator should be chosen carefully.

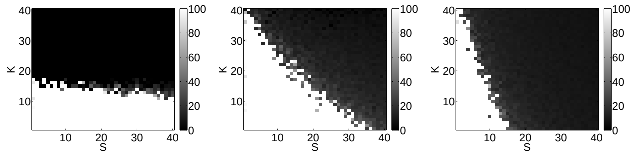



**Fig. 5** *A* **random,** $\Psi = I$**.** Worst PSNR of the reconstruction w.r.t. signal and noise sparsity $K, S$, for $A_r^\lambda$ with $N = 100$, $M = 50$ for (a) $\lambda = 0.1\sqrt{M}$ ,(a)$\lambda = \sqrt{M}$ (c) $\lambda = 2\sqrt{M}$

In Figure 5, the perfect recovery experiment is made using $A_r^\lambda$. Variations of $\lambda$ give the same behaviour of the shape of the zone leading to $K, S$-forgiveness.

## 5 Application to multi-image processing

In this section, we first show in the case of multi-image processing (multi image denoising), that for supports having a fixed value of sparsity $K, L$, $TV$ regularization does not improve forgiveness. However, if we make the assumption that supports of signal gradient and outliers are disjoint, we show that forgiveness is improved. We use this result to show that some particular SR problems can be better solved using $L^1 - TV$ minimization. These consideration lead to the determination of the value of the optimal regularization parameter for 1D $L^1 - TV$ multi-signal denoising and super-resolution problem.

### 5.1 Multi-image denoising

Multi-image denoising (also called burst denoising) is the process of using several images of the same scene to produce an image with reduced noise. We consider here the very simple problem of removing outliers from a collection of registered images. Using the multi-image acquisition model from Section 2 with sub-sampling factor $M = 1$ and no motion ($Q_i = Id$). The $L^1$ minimization from Equation 2 will lead to the selection of the median of the observations of each pixel. In this case, we can directly study the non-concentration property of $A$ and $A_r^\lambda$. Because it is a particular case of super-resolution (with $M = 1$), it will allow us to draw some limits for the behaviour of $L^1 - TV$ SR.

Let $N$ be odd. In 1D, $A$ is of size $Nl \times l$ and made of $l \times l$ diagonal blocks of 1. Let $T\ in \mathcal{T}$,

**Proposition 8** *$A$ is $\mathcal{T}$ forgiving if and only if each pixel is contaminated at most $N/2$ times*

*Proof* We check the equivalent $\mathcal{T}$ NCP for $A$. First, if a pixel $i$ is contaminated $K = \lfloor N/2 \rfloor + 1$ times. We consider the problem where only this pixel is not 0. Then

$$\|(Au).T\|_1 = K|u_i| > \lfloor N/2 \rfloor |u_i| = \|(Au).T^c\|_1 \quad (41)$$

and the NCP is not verified. It is necessary that any pixel is not contaminated more than $\lfloor N/2 \rfloor$ times for $A$ to be $\mathcal{T}$ forgiving.

Conversely, if every pixel is not contaminated more than $\lfloor N/2 \rfloor$, we look at the worst case when every pixel is contaminated exactly $K = \lfloor N/2 \rfloor$ times. We have

$$\begin{aligned}&\|(Au).T\|_1 = \\ &K\sum_i |u_i| < (\lfloor N/2 \rfloor + 1)\sum_i |u_i| = \|(Au).T^c\|_1\end{aligned} \quad (42)$$

Then Im$A$ has the $\mathcal{T}$-NCP. □

We now consider $\Psi$ the discrete gradient function : $(\Psi u)_1 = u_1$, $(\Psi u)_i = u_i - u_{i-1}$. Because $\Psi$ is invertible, for any support of sparsity $L$, we can find $u$ such that :

$$\|(\Psi u).L\|_1 > \|(\Psi u).L^c\|_1 \quad (43)$$

If we consider the worst case for the NCP of $A_r^\lambda$, we find that $\Psi$ cannot improve it. Given a support of sparsity $L$, we can always find a signal $u$ such that $\|(\Psi u).L\|_1 - \|(\Psi u).L^c\|_1 > 0$. With such $u$, we would need to ensure the NCP :

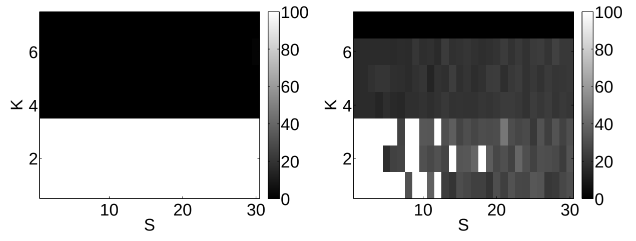
$$\begin{aligned}&\|(Au).T\|_1 + \lambda\|(\Psi u).L\|_1 = K\sum_{i=1,S}|u_i| + \lambda\|(\Psi u).L\|_1 \\ &< (N-K)\sum_{i=1,S}|u_i| + \lambda\|(\Psi u).L^c\|_1\end{aligned} \quad (44)$$

which is equivalent to

$$\begin{aligned}&K\sum_{i=1,S}|u_i| + \lambda(\|(\Psi u).L\|_1 - \|(\Psi u).L^c\|_1) \\ &< (N-K)\sum_{i=1,S}|u_i|\end{aligned} \quad (45)$$

The left side is minimized for $\lambda = 0$. This is the weakest inequality that we can obtain. Consequently, in the worst case, it is better to avoid $TV$ regularization. In practice, outliers might not contaminate jumps in the signal. Using the ERP, we can show that the system becomes very forgiving, we can even determine an optimal $\lambda$.

Let us suppose that no support of outlier coincides with jumps in the signal, *i.e.* if the gradient at a some position is non 0, the 2 samples used for its calculation are never contaminated by outliers.




(a) (b)

**Fig. 6** $A$ **multi-image observation,** $\Psi$ **gradient, overlap.** Worst PSNR of the reconstruction w.r.t. signal and noise sparsity $K, S$ when outliers can overlap with signal jumps with $N = 7$ and $l = 50$ for (a) $\lambda = 0$ ,(b)$\lambda = 0.95N/2$

**Proposition 9** *Let $\mathcal{T}'$ be the sparsity support of outliers and signal, such that outliers and signal jumps be disjoint. Then $\mathcal{T}'$ is not a concatenation of two sets of supports of signal and noise. Let us suppose that at most $K$ outliers contaminate each pixel and $\lambda > 0$. If $(2K - N)/2 < \lambda < N/2$, $A_r^\lambda$ is $\mathcal{T}'$-forgiving.*

*Proof* We look for a vector $v \in \mathbb{R}^{(N+1)l}$ which verifies properties of Theorem 2. We index the values of $v$ by indices $(v_{j,k})_{j=1..l,k=1..N+1}$. Index $j$ represents the pixel number. Index $k$ represents the acquired image number for $k \leq N$ and the image gradient for $k = N + 1$. Let $T' = (T, L)$ be a support in $\mathcal{T}'$. Let $v$ such that $v.T = sign(n).T$ $v.L = sign(\Psi u_0).L$. We need $(A_r^\lambda)^H v = 0$ which translates for each pixel (values indexed by $j$ of $(A_r^\lambda)^H v$):

$$\sum_{k=0,N-1} v_{j,k} + \lambda(v_{j,N+1} - v_{j+1,N+1}) = 0 \tag{46}$$

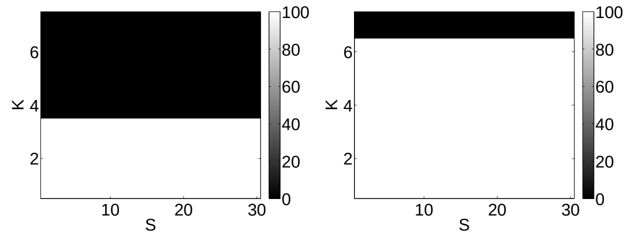
Given a pixel, we can distinguish two cases :

- pixel $j$ is not contaminated by outliers. Then, for $k = 1..N$ each variable $v_{j,k}$ is free and we can choose $v_{j,k} = -\lambda(v_{j,N+1} - v_{j+1,N+1})/N$. Because the maximum $L^1$ norm of the right side is 2, we guarantee that $\|v.T'^c\|_\infty < 1$ if $\lambda < N/2$.
- pixel $j$ is contaminated by outliers. Then we need $\sum_{k\in T} v_{j,k} = -\sum_{k\in T^c} v_{i,k} - \lambda(v_{j,N+1} - v_{j+1,N+1})$, with the right side having free variables. Because the maximum $L^1$ norm of the left side is $K$, we can find the right $v$ with $\|v.T'^c\|_\infty < 1$ if $K < N - K + 2\lambda$, which is rewritten $\lambda > (2K - N)/2$.

□

The same proposition is true in 2D with $(2K - N)/4 < \lambda < N/4$.

The result of experiments with overlapping outliers and signal jumps are displayed in Figure 6. They show that when we allow outliers and signal jumps to overlap, multi-image denoising with $L^1 - TV$ is not forgiving.




(a) (b)

**Fig. 7** $A =$ **multi-image observation,** $\Psi =$ **gradient. No overlap** Worst PSNR of the reconstruction w.r.t. signal and noise sparsity $K, S$ when outliers do not overlap with signal jumps with $N = 7$ and $l = 50$ for (a) $\lambda = 0$ ,(b)$\lambda = 0.95N/2$

In Figure 7, we show that with no overlap between jumps and outliers and $(N - 2)/2 < \lambda < N/2$, $A_r^\lambda$ is $N - 1$-forgiving., for $N = 7$. In other words, only one clean observation of each pixel is needed when there is no overlap.

## 5.2 Super-Resolution

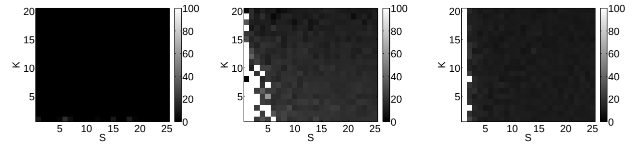
### *5.2.1 The limits of $L^1 - TV$ super-resolution for outlier removal*

We now consider $A$ as a 1D translational SR matrix and $\Psi$ is the same discrete gradient calculation as in the previous section. The study in the worst case scenario for multi-image denoising shows that without a precise hypothesis on the support of signal and outliers, sparse $L^1$ regularization does not enhance forgiveness.

For large number of images, $L^1$ norm super-resolution is forgiving $N_c$ contaminated images if the number of total images $N$ is greater than $CN$ with $C$ a constant. In Figure 8, we show experiments of the forgiveness of $A$ and $A_r^\lambda$ for an acquisition setting having poor forgiveness. We observe the trade-off between the number of outliers and the sparsity of the signal for a perfect reconstruction.

The main conclusion that we can draw from this experiment is that enforcing a regularity condition on the image allows to be robust to more outliers, but that the improvement is limited compared to the strength of the assumption on the signal.

Similarly to the multi-image denoising case, we would like to know if outliers not coinciding with signal jumps would be removed with $TV$ regularization. The main problem with exact super-resolution is that the lines of $A$ are sine cardinal. Consequently, the contamination by one outlier will concern all pixels in the desired HR image.



(a) (b) (c)

**Fig. 8** $A =$ **1D SR operator,** $\Psi =$ **gradient.** Worst PSNR of the reconstrucion for $TV$ regularized 1D super-resolution w.r.t to signal and noise sparsity with $N = 10$, $M = 2$ and $l = 25$ for (a) $\lambda = 0$,(b)$\lambda = 1$ and (c)$\lambda = 10$

*5.2.2 Finite impulse response $L^1$ SR algorithm benefits from $TV$ regularization*

In practice, some SR methods make the approximation that the shift operator is of finite length, such model can be found in Champagnat et al (2009). We propose to replace the shift by convolution by a sine cardinal by the convolution by a truncated Gaussian. Such model can be justified by the knowledge of real Point Spread Function estimation (which can be found in Delbracio et al (2012)). Here, for 1D SR , $A$ is defined by

$$\begin{aligned} A : \mathbb{R}^{Ml} &\to \left(\mathbb{R}^l\right)^N \\ u &\to (A_i u)_{i=1,N} = (SG_i u)_{i=1,N} \end{aligned} \tag{47}$$

where $G_i$ is the convolution by a shifted truncated Gaussian $(g_q)_{q=-p,p}$ of size $2p+1$. We choose the Gaussian parameter $a$ such that $|g_p|$ is small. We consequently suppose that $G_i$ is 0 outside of the first $p+3$ diagonals (the amplitude of translations in SR can be kept less than $M = 2$ without loss of generality). We keep the same definition for $\Psi$. To avoid complicated calculations, we also suppose that signal jumps are separated by the length of the filter.

**Proposition 10** *Let $\mathcal{T}'$ be the supports of sparsity such that outliers and signal jumps are separated by $p+1$ and that jumps are separated by at least $2p+1$ pixel. Let us suppose that at most $K$ outliers contaminate each pixel. let $G_+ = \max_{i,j} \|G_{i,j}\|_1$ and $G_- = \min_{i,j} \|G_{i,j}\|_1$ where $G_{i,j}$ is the jth column of $SG_i$. If $(K(G_+ + G_-) - NG_-)/2 < \lambda < NG_-/2$ , $A_r^\lambda$ is K-forgiving.*

*Proof* We look for a vector $v \in \mathbb{R}^{Nl+Ml}$ which verifies properties of theorem 2. Let $T' = (T, L)$ be a support in $\mathcal{T}'$. Let $v$ such that $v.T = sign(n).T$ $v.L = sign(\Psi u_0).L$. We write $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ where $v_1$ has size $Nl$ and $v_2$ has size $Ml$. We need $(A_r^\lambda)^H v = 0$ which, for each HR pixel (each column $j$ of $A_r^\lambda$), translates to:

$$A_j^H v_1 + \lambda \Psi_j^H v_2 = 0, \tag{48}$$

with $A_j$ the $jth$ column of $A$ and $\Psi_j$ the $jth$ column of $\Psi$. Given a pixel, we can distinguish two cases :

– pixel $j$ is not contaminated by outliers. Then each variable in $v_1$ is free. We need $A_j^H v_1 > 2\lambda$ because the maximum value of $\Psi_j^H v_2$ is 2. The restriction $|v_1|_\infty < 1$ imposes that $\lambda < \|A_j^H\|_1/2$.
– pixel $j$ is contaminated by outliers. Then we need

$$A_j^H(v_1.T) = -A_j^H(v_2.T^c) - \lambda \Psi_j^H v_2 \tag{49}$$

with the right side being free variables. Similarly as before, it is necessary that $\|A_{T,j}^H\|_1 < \|A_{T^c,j}^H\|_1 + 2\lambda$, which is equivalent to $\lambda > (\|A_{T,j}^H\|_1 - \|A_{T^c,j}^H\|)/2$.

These two conditions are sufficient, and using the definition of $G_+$ and $G_-$, the condition $(K(G_+ + G_-) - NG_-)/2 < \lambda < NG_-/2$ is sufficient for the ERP. □

*Remark 10* We can find a $\lambda$ if $K < K_{max} = N/(1 + G_+/G_-)/2$. Typically, $G_+/G_- \sim 2$.If consecutive jumps were allowed in the signal, the constants would change and would require the use of the $L^1$ norm of inverse matrices of restrictions of $A$. Because estimating the extremum of these norms is difficult, we restrict ourselves to this simplified version of the problem.

*Experiments for 1D super-resolution.* In Figure 9, we show the influence of $TV$ regularization on 1D finite length SR with zoom $M = 2$. The HR image has size $l = 210$. For each $K, S$, we generated 100 experiments meeting hypotheses of Proposition 10, contaminating the same pixel in the 10 LR images, in an area without jumps. We solved the regularized SR problem using our calculation of $\lambda$. We see that after regularization, the number of contaminated pixels can be greater than half the number of images.

*Experiments for 2D super-resolution.* We use the same model for 2D super-resolution (a 2D Gaussian is used instead). We use the image Shepp Logan phantom which is sparse in the gradient domain. $\Psi$ is the 2D discrete gradient. We perform a 2D $L^1$ super-resolution with total variation regularization with $M = 2$. We use an iteratively reweighted least squares algorithm (Daubechies et al. 2010). We contaminate the same region of the image on the different LR images. In Figure 10, we show the ideal image, the sparsity level of the image, a clean low resolution image and a contaminated low resolution image.

From Proposition 10, we can infer that a range of $\lambda$ will make the 2D $L^1 - TV$ SR problem forgiving. We show in Figure 11, the result of 2D $L^1$ SR without regularization, and with $TV$ regularization. The SR was performed using 6 LR images, with 2 contaminated images (recall that the minimum number of images for a perfect reconstruction is 4 for $M = 2$). While SR without regularization fails $TV$ regularization gives a

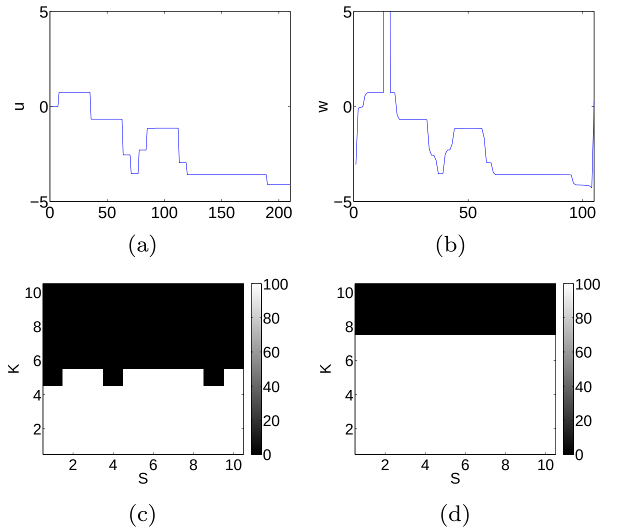



**Fig. 9** $A$ = **finite length 1D SR operator,** $\Psi$ = **gradient.** Worst reconstruction PSNR for different signal and noise sparsity $K, S$ with no overlap ($N = 10, M = 2, l = 20$). (a) Example of an HR image, (b) example of a contaminated LR image, (c) SR without regularization (d) SR with $TV$ regularization with optimal $\lambda$.

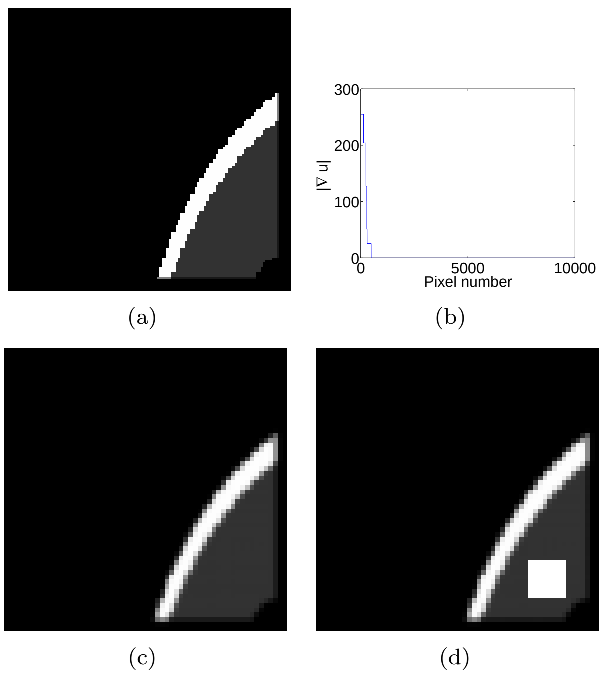



**Fig. 10 2D SR problem.** (a) HR image, (b) sorted values of the 2D discrete gradient (HR image gradient sparsity), (c) one clean LR image and (d) one contaminated LR image

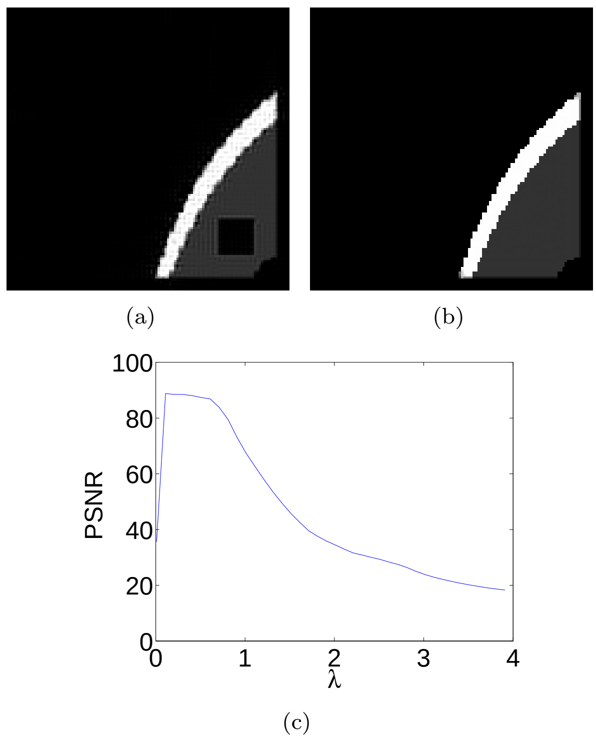



**Fig. 11 2D SR results, experiment 1** (a) reconstruction without regularization, PSNR=25.49dB, (b) reconstruction with regularization PSNR=88.27dB, (c) PSNR with respect to $\lambda$ .

perfect reconstruction. We also show how the PSNR behaves with respect to $\lambda$. A range of $\lambda$ values lead to a perfect reconstruction (from 0.1 to 0.8), outside of the range the effect of oversmooting (from 0.8 to 4) and failure (small $\lambda$) can be observed.

In Figure 12, we show another experiment with $N = 10$ and $M = 6$. Solving the unregularized problem is not forgiving the outliers and fails completely. $TV$ regularization leads to a perfect reconstruction. In this experiment, only the critical number of images $M = 4$ (for the system to be invertible) is necessary to recover the HR image and there is less clean images than contaminated images. The behaviour of the reconstruction with respect to $\lambda$ shows a similar behaviour as before. A range of $\lambda$ (from 0.9 to 1.05) yields a perfect reconstruction.

*Remark 11* In this article, we supposed that only sparse outlier noise was corrupting the signal. This hypothesis was made to keep the exposition of the main concepts clear. However, because the theoretical concepts are direct extensions of sparse recovery theory, stability with respect to additive noise (*e.g.* Gaussian noise) should be obtained by extending the corresponding results in the literature (Candès et al. 2006; Cohen et al. 2009).

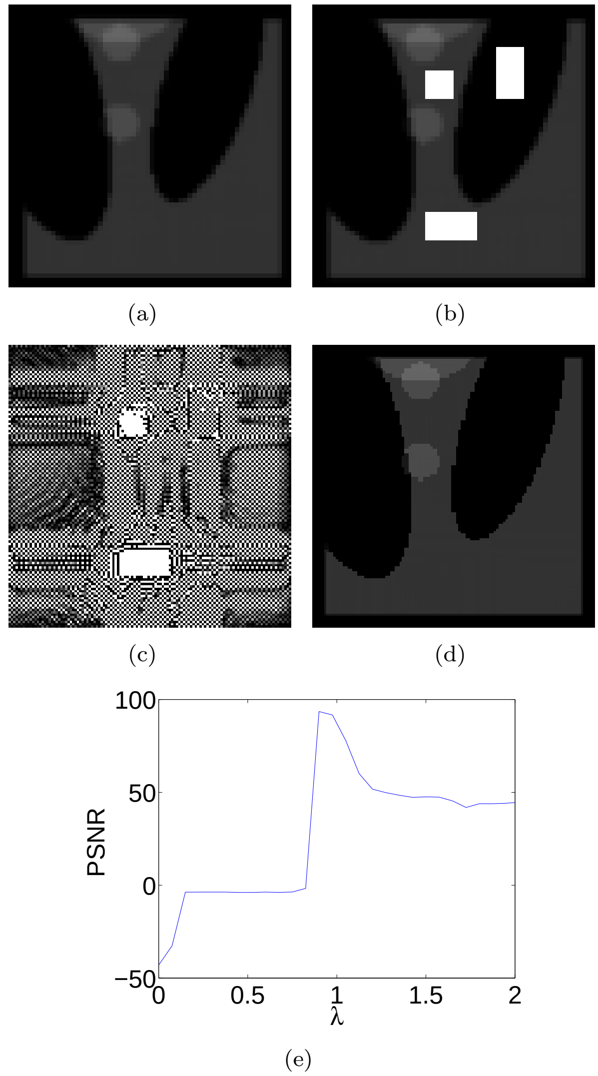



**Fig. 12 2D SR results, experiment 2** (a) HR image, (b) one contaminated LR image, (c) reconstruction without regularization, PSNR=-46.84dB, (d) reconstruction with regularization PSNR=93.31dB, (e) PSNR with respect to $\lambda$ .

*5.2.3 Other regularization schemes, $L_0$ and outlier detection*

As previously noticed, the precise *sinc* SR method does not benefit a lot from the $L^1-TV$ method. Two possible ways could be explored.

First, a way to ensure the robustness of the problem using sparse regularization is to make sure the overdetermined matrix $\Psi$ is $\mathcal{L}$-forgiving with good $NCP$ constant. Popular dictionary based or non-local approaches for image modeling can lead to a $\Psi$ being a projection of the signal into a high dimensional space where the signal is sparse. If this projection has the right NCP with respect to signal sparsity supports, outlier removal power will be increased with any type of observation (universal encoding).

Secondly, we can make the following observation : very energetic and concentrated outliers are easy to detect. While $L^1-TV$ minimization will not be able to remove them in critical cases (*e.g.* for 2D super-resolution $N = M^2+1$ ). Then detecting the outlier by thresholding, and removing contaminated equations will lead to an invertible system and the right solution. In this case this, process can be thought as an $L^0$ minimization, because the (informed) outlier suppression is a way to try to maximize the number of equations which are met by the data.

## 6 Conclusion

In this article, we showed how the theory of sparse recovery can be used and extended to the case of outlier robustness with sparsity priors on the signal for multi-image processing. We illustrated how some tools are more adapted to study matrices without too much structure (random matrices and the RIP) and others to the multi-image problems (Exact Recovery Property, Non Concentration Property). We showed that considering particular classes of supports allows for an understanding of how the $L^1-TV$ scheme behaves with outliers. In simple cases, we could determine the regularization parameter leading to the best robustness. The intuition gained was verified experimentally in the more theoretically challenging 2D SR case.

## References

Bach F, Jenatton R, Mairal J, Obozinski G (2012) Structured Sparsity through Convex Optimization. Statistical Science 27(4):450–468, DOI 10.1214/12-sts394

Benoit L, Mairal J, Bach F, Ponce J (2011) Sparse image representation with epitomes. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, pp 2913–2920, DOI 10.1109/cvpr.2011.5995636, URL http://dx.doi.org/10.1109/cvpr.2011.5995636

Candès E, Romberg J (2007) Sparsity and incoherence in compressive sampling. Inverse Problems 23(3):969–985, DOI 10.1088/0266-5611/23/3/008

Candès EJ (2008) The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique 346(9-10):589–592, DOI 10.1016/j.crma.2008.03.014

Candes EJ, Tao T (2005) Decoding by linear programming. Information Theory, IEEE Transactions on 51(12):4203–4215, DOI 10.1109/TIT.2005.858979

Candes EJ, Tao T (2006) Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? Information Theory, IEEE Transactions on 52(12):5406–5425, DOI 10.1109/TIT.2006.885507

Candes EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. Information Theory, IEEE Transactions on 52(2):489–509, DOI 10.1109/TIT.2005.862083

Candès EJ, Romberg JK, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. Comm Pure Appl Math 59(8):1207–1223, DOI 10.1002/cpa.20124

Champagnat F, Le Besnerais G, Kulcsár C (2009) Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework. J Opt Soc Am A 26(7):1730–1746, DOI 10.1364/JOSAA.26.001730

Cohen A, Dahmen W, DeVore R (2009) Compressed sensing and best k-term approximation. J Amer Math Soc 22(1)

Daubechies I, DeVore R, Fornasier M, Güntürk CS (2010) Iteratively reweighted least squares minimization for sparse recovery. Comm Pure Appl Math 63(1):1–38, DOI 10.1002/cpa.20303

Delbracio M, Almansa A, Morel JM, Musé P (2012) Subpixel Point Spread Function Estimation from Two Photographs at Different Distances. SIAM Journal on Imaging Sciences 5(4):1234–1260, DOI 10.1137/110848335

Elad M, Aharon M (2006) Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. Image Processing, IEEE Transactions on 15(12):3736–3745, DOI 10.1109/tip.2006.881969

Farsiu S, Robinson D, Elad M, Milanfar P (2004a) Advances and challenges in super-resolution. Int J Imaging Syst Technol 14(2):47–57, DOI 10.1002/ima.20007

Farsiu S, Robinson MD, Elad M, Milanfar P (2004b) Fast and robust multiframe super resolution. Image Processing, IEEE Transactions on 13(10):1327–1344, DOI 10.1109/TIP.2004.834669

Geiger D, Liu TL, Donahue M (1999) Sparse Representations for Image Decompositions. International Journal of Computer Vision 33(2):139–156, DOI 10.1023/a%253a1008146126392

He Y, Yap KH, Chen L, Chau LP (2007) A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution. Image Processing, IEEE Transactions on 16(11):2830–2841, DOI 10.1109/TIP.2007.908074

Jenatton R, Audibert JY, Bach F (2011) Structured Variable Selection with Sparsity-Inducing Norms. J Mach Learn Res 12:2777–2824

Kuppinger P, Durisi G, Bolcskei H (2012) Uncertainty Relations and Sparse Signal Recovery for Pairs of General Signal Sets. Information Theory, IEEE Transactions on 58(1):263–277, DOI 10.1109/tit.2011.2167215

Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2009) Non-local sparse models for image restoration. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE, pp 2272–2279, DOI 10.1109/ICCV.2009.5459452

Mallat S, Yu G (2010) Super-Resolution With Sparse Mixing Estimators. Image Processing, IEEE Transactions on 19(11):2889–2900, DOI 10.1109/TIP.2010.2049927

Marquina A, Osher S (2008) Image Super-Resolution by TV-Regularization and Bregman Iteration. J Sci Comput 37(3):367–382, DOI 10.1007/s10915-008-9214-8

Milanfar P (2010) Super-resolution imaging, vol 1. CRC Press

Mitra K, Veeraraghavan A, Chellappa R (2013) Analysis of Sparse Regularization Based Robust Regression Approaches. Signal Processing, IEEE Transactions on 61(5):1249–1257, DOI 10.1109/tsp.2012.2229992

Ning Q, Chen K, Yi L, Fan C, Lu Y, Wen J (2013) Image Super-Resolution Via Analysis Sparse Prior. Signal Processing Letters, IEEE 20(4):399–402, DOI 10.1109/lsp.2013.2242198

Studer C, Baraniuk RG (2013) Stable restoration and separation of approximately sparse signals. Applied and Computational Harmonic Analysis DOI 10.1016/j.acha.2013.08.006

Studer C, Kuppinger P, Pope G, Bolcskei H (2012) Recovery of Sparsely Corrupted Signals. Information Theory, IEEE Transactions on 58(5):3115–3130, DOI 10.1109/tit.2011.2179701

Tartavel G, Gousseau Y, Peyré G (2013) Constrained Sparse Texture Synthesis. In: Kuijper A, Bredies K, Pock T, Bischof H (eds) Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, vol 7893, Springer Berlin Heidelberg, pp 186–197, DOI 10.1007/978-3-642-38267-3_16

Tian J, Ma KK (2011) A survey on super-resolution imaging. Signal, Image and Video Processing 5(3):329–342, DOI 10.1007/s11760-010-0204-6

Traonmilin Y, Ladjal S, Almansa A (2012) On the amount of regularization for super-resolution interpolation. In: Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, IEEE, pp

380–384

Traonmilin Y, Ladjal S, Almansa A (2013) Outlier Removal Power of the L1-Norm Super-Resolution. In: Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, vol 7893, Springer Berlin Heidelberg, pp 198–209, DOI 10.1007/978-3-642-38267-3_17

Vaiter S, Deledalle CA, Peyré G, Dossal C, Fadili J (2012) Local behavior of sparse analysis regularization: Applications to risk estimation. Applied and Computational Harmonic Analysis DOI 10.1016/j.acha.2012.11.006

Yap KH, He Y, Tian Y, Chau LP (2009) A Nonlinear -Norm Approach for Joint Image Registration and Super-Resolution. Signal Processing Letters, IEEE 16(11):981–984, DOI 10.1109/LSP.2009.2028106

Yu G, Sapiro G, Mallat S (2012) Solving Inverse Problems With Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity. Image Processing, IEEE Transactions on 21(5):2481–2499, DOI 10.1109/tip.2011.2176743