

# Lipschitz bounds for noise robustness in compressive sensing: two algorithms

Marc Nicodème, Charles Dossal, Flavius Turcu, Yannick Berthoumieu

► **To cite this version:**

Marc Nicodème, Charles Dossal, Flavius Turcu, Yannick Berthoumieu. Lipschitz bounds for noise robustness in compressive sensing: two algorithms. International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Sep 2014, Timisoara, Romania. 2014. <hal-01063235>

**HAL Id: hal-01063235**

**<https://hal.archives-ouvertes.fr/hal-01063235>**

Submitted on 11 Sep 2014

**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.

# Lipschitz bounds for noise robustness in compressive sensing: two algorithms

Marc Nicodème<sup>\*+</sup>, Charles Dossal<sup>+</sup>, Flavius Turcu<sup>\*</sup> and Yannick Berthoumiou<sup>\*</sup>

<sup>\*</sup>: IMS, CNRS UMR 5218, *Université de Bordeaux*

<sup>+</sup>: IMB, CNRS UMR 5251, *Université de Bordeaux*

## ABSTRACT

The paper deals with numerical estimations of Lipschitz bounds relating locally the reconstruction error to the measurement error in the compressive sensing framework. Most recent theoretical papers in the field parametrize such bounds relatively to certain families of vectors called dual certificates, which are fundamental to several reconstruction criteria. The paper provides two algorithms for computing dual certificates that optimize their related reconstruction error bounds. We give a greedy algorithm that provides a fast approximate solution, and a convex-projection algorithm that computes the exact optimum.

## I. INTRODUCTION

An important topic in compressive sensing, that comes alongside with the central problem of the recoverability of sparse signals, is the robustness to measurement noise for signals that are normally recoverable in the absence of noise.

Suppose  $x^0 \in \mathbb{R}^n$  is an unknown sparse signal that has to be recovered from an ill-posed linear problem

$$y = \mathbf{A}x^0 + b \quad (1)$$

in which  $y$  is a known vector, called *observation* or *measure data*,  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is a known linear operator with  $m < n$ , called *sensing* or *measuring operator*, and  $b \in \mathbb{R}^m$  is an unknown vector called *sensing* or *measuring noise*. Since the sensing matrix always has a non-trivial kernel, "solving" (1) undergoes a *regularization*, which classically means finding the solution(s) of (1) that minimize a certain functional. In the noiseless case  $b = 0$ , a vector  $x_0$  that coincides with the unique solution of the regularized problem is called *recoverable* or *identifiable* with respect to the chosen functional.

If one is only interested in sparse signals  $x^0$ , the natural choice for such a functional would be the  $\ell^0$ -pseudo-norm  $\|x\|_0 = \#(\text{supp}(x))$ , i.e. the number of non-zero coordinates of  $x$ . Unfortunately solving the  $\ell^0$ -regularized problem is considered untractable, hence a popular choice is rather to consider a convex  $\ell^1$  regularization. In the noiseless case  $b = 0$  this amounts to solving the *basis pursuit* [4]:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } \mathbf{A}x = y \quad (2)$$

In the noisy case several variants have been developed, such as the popular Lagrangian formulation, but here we will rather focus on the (equivalent) constrained optimization problem

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \text{ s.t. } \|\mathbf{A}x - y\|_2 \leq \delta. \quad (3)$$

This model has been explored by many papers and applied in a different applications in signal and image processing. Here the identifiability of the original vector is evaluated by sign consistency and noise robustness.

More precisely,  $x^0$  is considered identifiable with sign consistency if the solution of (3) has the same support than  $x^0$  i.e. the positions of

the non zero components. The recovery of the support generally needs strong assumptions and might hold only for small enough noises.

On the other side, noise robustness involves the  $\ell^2$  recovery error regardless of the support, which in our case is  $\|x^0 - x^\delta\|_2$ , for some minimizer  $x^\delta$  of the optimization problem (3). Noise robustness is usually expressed in terms of some Lipschitz bound  $C$  relating the  $\ell^2$  recovery error to the  $\ell^2$  norm of the noise, such as

$$\|x^\delta - x^0\|_2 \leq C\delta, \quad \|b\|_2 \leq \delta \quad (4)$$

in the case of the problem (3).

Classical theoretical estimations for such constants were given in the well-studied context of random matrices  $\mathbf{A}$  with various prescribed column distributions. Such matrices are shown to satisfy the restricted isometry property (RIP), hence noise-robustness bounds can be derived in terms of RIP constants.

Unfortunately these RIP assumptions can only be applied for random matrices and are unsuitable in general for some given deterministic matrix. Deterministic matrices occur in various applications, as for instance in tomography, which was our initial target application. Here sensing matrices can be viewed as discrete versions of Radon transforms, whose behavior does not match the RIP context.

Other known results applying to deterministic matrices give errors bounds estimates using a semi implicit formula of the solution of the noiseless  $\ell^1$ -minimization problem, but are only valid under small-noise assumptions.

As an alternative, more recent papers focus on the local behavior of the reconstruction error. One such approach theoretically consists in emphasizing an error bound for each member of a certain family of vectors associated to a local noiseless solution. These vectors are called *dual certificates* and play a fundamental role in identifiability criteria based on the so-called *source condition* [1], [11]. As such, the theoretical bound estimates based on dual certificates are more versatile, since they are local - hence potentially more accurate - and need no a priori assumptions on the magnitude of the measurement noise. However, dual certificates are not unique, and finding them is not straightforward and might have a significant computational cost. The common strategy consists in building "candidates" which can be either generic or specific to a certain context. In the context of super-resolution, Candès et al [2] [3] build a trigonometric polynomial which is a certificate under certain conditions. In the same context, Duval et al [8] constructed candidates by minimizing their  $\ell^2$  norm and imposing a zero-derivative whenever  $x^0$  has a non-zero component. Also, Vaiter et al [14] compute a generic candidate.

In this context involving dual certificates, theoretical robustness Lipschitz bound were given throughout [1], [12] and [7], as

$$\|x^\delta - x^0\|_2 \leq C(\eta)\delta, \quad \|b\|_2 \leq \delta, \quad (5)$$

where the constant  $C(\eta)$  depends, besides on the matrix  $\mathbf{A}$  and the support of  $x^0$ , on some dual certificate  $\eta \in \mathbf{R}^m$  for  $x^0$ .

In the paper we provide two algorithms for computing dual certificates  $\eta$  which minimize their related Lipschitz error bounds  $C(\eta)$  in (5), based on two different approaches:

The first is a greedy algorithm that provides a fast approximate solution, by iteratively minimizing the error bound with respect to a current dual certificate candidate. This produces a fast approximation of the optimal bound, since all the numerical computations only rely on scalar products.

The second algorithm is based on a theoretical geometric result that describes the exact optimal solution. More precisely we showed that the problem of finding the dual certificate minimizing the error bound can be reduced to computing the convex projection of a suitable vector over a certain convex set. As such, it gives the best error bound - which is attained for a unique dual certificate  $\eta_{opt}$ , but has a higher computational cost than the greedy approximation.

The two algorithms are illustrated and numerically compared for compressive sensing matrices issued from a tomography application framework. The relationship between theoretical bounds and numerical observed bounds is also discussed.

*Notations:* Throughout the paper,  $(\mathbf{e}_i)_{i=1,\dots,n}$  stands for the canonical base in  $\mathbb{R}^n$ ,  $\mathbf{A} \in \mathbb{R}^{m \times n}$  denotes a linear operator from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ , and  $\mathbf{a}_i = \mathbf{A}\mathbf{e}_i$ , ( $i = 1, \dots, n$ ) the columns of  $\mathbf{A}$ .

For  $x = [x_1, \dots, x_n] \in \mathbb{R}^n$ , its support is denoted

$$\text{supp}(x) = \{i \in \{1, \dots, n\} \mid x_i \neq 0\},$$

and its vector sign by  $\text{sign}(x) = [\text{sign}(x_i)]_{i=1,\dots,n}$ .

Also, for some subset  $I \subset \{1, \dots, n\}$ , we denote  $\#(I)$  the cardinal of  $I$ , we write  $x_I$  for the restriction of  $x$  to its components indexed by  $I$ , and  $\mathbf{A}_I$  for the submatrix of  $\mathbf{A}$  with the columns  $\mathbf{a}_i$ ,  $i \in I$ .

The notation  $\mathbf{A}^*$  stands for the adjoint of  $\mathbf{A}$ , while  $\mathbf{A}^+$  denotes the Moore-Penrose pseudoinverse of  $\mathbf{A}$ , i.e.  $\mathbf{A}^+ = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*$ . The  $p, q$ -operator norm of a matrix  $\mathbf{A}$  is

$$\|\mathbf{A}\|_{p,q} = \max_{x \neq 0} \frac{\|\mathbf{A}x\|_q}{\|x\|_p} \text{ and } \|\mathbf{A}\|_{p,p} = \|\mathbf{A}\|_p \quad (6)$$

Note that  $\|\mathbf{A}\|_{1,2} = \max_i \|\mathbf{a}_i\|_2$  and  $\|\mathbf{A}^+\|_2 = \sqrt{\|(\mathbf{A}^* \mathbf{A})^{-1}\|_2}$ .

Finally, for any subset  $M$  of  $\mathbb{R}^n$ , we write  $\text{Aff}(M)$  and  $\text{Conv}(M)$  for the affine, respectively convex envelope of  $M$ . In particular,  $B_1 = \text{Conv}(\{\pm \mathbf{e}_i\}_{i=1,\dots,n})$  stands for the closed unit ball of  $\mathbb{R}^n$  under the  $\ell^1$ -norm.

## II. STATE OF THE ART

Let us begin by recalling the notion of *dual certificate*, its related *source* (or *range*) *condition* from regularization theory [1] and their connection with identifiability and noise robustness in the framework of  $\ell^1$ -minimization.

Fix a linear operator  $\mathbf{A} \in \mathbb{R}^{m \times n}$  from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ , a vector  $x^0$  in  $\mathbb{R}^n$  and let  $I = \text{supp}(x^0)$ . A vector  $\eta$  in  $\mathbb{R}^m$  is called a *dual certificate* for  $x^0$  if the vector  $\mathbf{A}^* \eta$  belongs to the subgradient of the  $\ell^1$  norm at  $x^0$ . This latter set can be viewed as

$$\partial \|x^0\|_1 = \{u \in \mathbb{R}^n : u_I = \text{sign}(x_I) \text{ and } \|u\|_\infty \leq 1\} \quad (7)$$

If  $\eta$  satisfies only the first condition above, i.e.

$$(\mathbf{A}^* \eta)_I = \text{sign}(x_I^0) \quad (8)$$

then  $\eta$  is usually called a *candidate* or a *pre-certificate*. Also,  $x^0$  is said to satisfy the *source condition* if it admits at least one dual certificate.

Let us first discuss these notions in connection to the identifiability of  $x^0$  in the noiseless case, which means, as mentioned in the introduction, the property of  $x^0$  to be the unique minimizer of the  $\ell^1$ -regularized problem (2).

One can easily check that the source condition on  $x^0$  is a sufficient condition for  $x^0$  to be a minimizer of (2), but it does not guarantee its uniqueness.

One crucial step towards uniqueness is to turn the second condition in (7) into a strict inequality outside the support of  $x^0$ , i.e. requesting that

$$\|(\mathbf{A}^* \eta)_J\|_\infty < 1, \quad J = \{1, \dots, n\} \setminus I \quad (9)$$

This strengthened condition insures that no minimizer of (2) can be supported anywhere but inside  $I$ . Searching for dual certificates with this strengthened condition has been proposed by several authors. It has been first exploited by Fuchs in [9], where the inequality (9) is tested for the generic pre-certificate

$$d(x^0) = \mathbf{A}_I^{+*} \text{sign}(x^0) \quad (10)$$

in order to derive a sufficient identifiability criterion. This latter condition is also known as the irrepresentability condition in statistics, see [7] for more details. Rather than testing a particular candidate, a more general sufficient criterion was given by Vaiteer et al in [14], where convex optimization is used to solve the problem

$$IC(x^0) = \min_{\eta \in \mathbb{R}^m} \|\mathbf{A}_J^* \eta\|_\infty \text{ s. t. } \mathbf{A}_I^* \eta = \text{sign}(x_I^0) \quad (11)$$

whose solutions are dual certificates satisfying (9) whenever  $IC(x^0) < 1$ .

Finally, even if  $x^0$  admits a dual certificate satisfying (9), in order for  $x^0$  to be the unique minimizer of (2) one has to make sure that there are no other minimizers supported inside  $I$ . One can see that this happens if and only if  $\mathbf{A}_I$  is injective.

Summing up, this latter injectivity condition, together with (8) and (9) represent a strengthened version of the source condition that was considered by Grasmair et al in [11][13][12] to describe the noiseless  $\ell^1$ -identifiability.

Dual certificates have a deep geometric interpretation in  $\ell^1$ -identifiability regarded in the context of polytope theory, as considered by Donoho in [6]. In fact, from the main result there it follows that  $x^0$  is identifiable if and only if the convex set

$$F(x^0) := \text{Conv}(\{\text{sgn}(x_i^0) \mathbf{a}_i : i \in I\}) \quad (12)$$

is a face of affine dimension  $\#(I) - 1$  of the image polytope

$$P := \mathbf{A}B_1 = \text{Conv}(\{\pm \mathbf{a}_i : i = 1, \dots, n\}) \quad (13)$$

Now the fact that  $F(x^0)$  is a face of  $P$  means by definition that there is a linear functional on  $\mathbb{R}^m$ , thus a vector  $\eta \in \mathbb{R}^m$  (via Riesz representation theorem), that separates  $F(x^0)$  from the rest of the polytope  $P$ . This translates by

$$\langle \eta, \text{sgn}(x_i^0) \mathbf{a}_i \rangle = 1 \quad i \in I, \quad (14)$$

$$\langle \eta, \pm \mathbf{a}_j \rangle < 1 \quad j \in J, \quad (15)$$

which are nothing but equivalent forms of (8) and (9) respectively. Moreover the injectivity of  $\mathbf{A}_I$  is equivalent to the fact that  $F(x^0)$  has affine dimension  $\#(I) - 1$ .

This means precisely that dual certificates, in the sense of the strengthened source condition, can be regarded as separating functionals for faces of the image polytope  $P$ . This geometric interpretation is also very useful in the matter of noise robustness, that will be discussed next.

Let us suppose now that  $x^0$  is noiseless-identifiable, thus satisfies the strengthened source condition for some dual certificate  $\eta$ . Assume that  $x^0$  is measured in (1) with some noise  $0 < \|b\|_2 \leq \delta$  and denote  $x^\delta$  a minimizer of the regularized problem (3). Estimating the noise robustness of the reconstruction deals with measuring the reconstruction error  $x^0 - x^\delta$ , which unfortunately has no implicit formula in general. It is much instead easier to measure this error along various directions, each direction producing an error estimate, with more or less accuracy. Among all possible directions, the most "fiable" are those given by the dual certificates, in the sense of scalar products of the type  $\langle \mathbf{A}^* \eta, x^0 - x^\delta \rangle$ . Such quantities characterize the Bregman distance between  $x^0$  and  $x^\delta$ , which was used by Burger et al. [1] in order to measure the recovery error with a very simple Lipschitz constant. However, the Bregman distance, as a pseudo-distance, is a rather weak measure, and one would prefer an  $\ell^2$ -estimation instead. This was done by Grasmair et al. [13] by relating the Bregman distance to the  $\ell^2$  norm. The resulting Lipschitz constant was afterwards reduced by Dossal et al [7], leading to the following result:

**Theorem 1** ([7]). *For any minimizer  $x^\delta$  of (3) and any dual certificate  $\eta$  satisfying the strengthened source condition (9) one has*

$$\|x^\delta - x^0\| \leq C(\eta)\delta \quad (16)$$

with

$$C(\eta) = 2 (\|\mathbf{A}_I^+\|_2 + (\|\mathbf{A}_I^+\|_2 \|\mathbf{A}_J\|_{1,2} + 1)Q(\eta)) \quad (17)$$

$$Q(\eta) = \frac{\|\eta\|_2}{1 - \|\mathbf{A}_J^* \eta\|_\infty} \quad (18)$$

Clearly the optimal Lipschitz constant is obtained for the dual certificate that minimizes the factor  $Q(\eta)$ , which solely depends on  $\eta$  in (17). This leads to the optimization problem

$$\min Q(\eta) : \quad (\mathbf{A}^* \eta)_I = \text{sign}(x_I^0), \quad \|(\mathbf{A}^* \eta)_J\|_\infty < 1 \quad (19)$$

We give in the next section the explicit formula for the solution of this problem and the corresponding algorithm that computes the optimal dual certificate and its related Lipschitz bound.

### III. A CONVEX-PROJECTION ALGORITHM FOR THE OPTIMAL LIPSCHITZ BOUND

We saw in the last section that a dual certificate  $\eta$  for an identifiable vector  $x^0$  supported on  $I$  has a geometric interpretation as a separating functional for the face  $F(x^0)$  of the image polytope  $P$  (see (12),(13)). This means that the dual affine subspace of  $\eta$

$$H_\eta := \{u \in \mathbf{R}^m : \langle \eta, u \rangle = 1\} \quad (20)$$

is a *support hyperplane* for the face  $F(x^0)$ , i.e.  $H_\eta \cap P = F(x^0)$ .

The key fact in solving the problem (19) is that the quantity  $Q(\eta)$  also has a strong geometric interpretation. More precisely, consider the "reduced" polytope

$$P_J = \text{Conv}(\{\pm \mathbf{a}_j : j \in J\}) \quad (J = \{1, \dots, n\} \setminus I) \quad (21)$$

One can show that:

**Proposition 2.** *For any dual certificate  $\eta$  of  $x^0$  one has*

$$Q(\eta) = \frac{1}{\mathcal{D}(H_\eta, P_J)} \quad (22)$$

where  $\mathcal{D}$  stands for the affine euclidian distance.

Therefore in order to minimize  $Q(\eta)$  one has to maximize the distance from the reduced polytope  $P_J$  to a generic affine hyperplane

$H_\eta$  passing through the face  $F(x^0)$ . Obviously the intersection of all these hyperplanes is the affine closure of the face  $F(x^0)$ , denoted

$$\mathcal{L}(x^0) := \text{Aff}(F(x^0)) = \text{Aff}(\{\text{sgn}(x_i^0) \mathbf{a}_i : i \in I\}) \quad (23)$$

It is more convenient to consider  $K$  the orthogonal complement of  $\mathcal{L}(x^0)$  in  $\mathbf{R}^m$  and project the whole problem onto  $K$ . Denote  $\mathbf{P}_K$  the orthogonal projector onto  $K$ . More precisely, suppose if  $I = \{i_1, \dots, i_p\}$  and  $\bar{\mathbf{A}}_I = (\bar{a}_i)_{i \in \{i_2, \dots, i_p\}}$  is the matrix with the columns  $\{a_i - a_{i_1}, i = \{i_2, \dots, i_p\}\}$ , then

$$P_K = \mathbf{I}_{\mathbf{R}^m} - \bar{\mathbf{A}}_I \bar{\mathbf{A}}_I^+ \quad (24)$$

Put  $\tilde{\mathbf{a}}_j = \mathbf{P}_K \mathbf{a}_j$  and consider the projected polytope

$$\tilde{P}_J = \mathbf{P}_K P_J = \text{Conv}(\{\pm \tilde{\mathbf{a}}_j : j \in J\}) \text{ and } \tilde{H}_\eta = \mathbf{P}_K H_\eta. \quad (25)$$

Now the entire affine subspace  $\mathcal{L}(x^0)$  projects onto a singleton, which can be shown to be precisely the vector

$$d(x^0)^+ = \frac{d(x^0)}{\|d(x^0)\|_2^2}, \quad (26)$$

the dual vector of the generic pre-certificate  $d(x^0)$  given in (10). It is also straightforward that

$$Q(\eta) = \frac{1}{\mathcal{D}(H_\eta, P_J)} = \frac{1}{\mathcal{D}(\tilde{H}_\eta, \tilde{P}_J)}. \quad (27)$$

Note that, in contrast to (22),  $\tilde{H}_\eta$  above is now a generic affine hyperplane in  $K$  which passes just through the point  $d(x^0)^+$ , and whose distance to the projected reduced polytope  $\tilde{P}_J$  needs to be maximized. One can use then a couple of standard convex analysis arguments to show the main result:

**Theorem 3.** *Denote  $w^0$  the point in  $\tilde{P}_J$  at minimal distance from  $d(x^0)^+$ . Then*

$$\begin{aligned} \min_\eta Q(\eta) &= \frac{1}{\mathcal{D}(\mathcal{L}(x^0), P_J)} = \frac{1}{\mathcal{D}(d(x^0)^+, \tilde{P}_J)} = \\ &= \frac{1}{\|d(x^0)^+ - w^0\|_2} \end{aligned} \quad (28)$$

Moreover, the minimal value is attained for the unique dual certificate

$$\eta_{opt} = \frac{\|d(x^0)\|_2^2}{1 - \langle d(x^0), w^0 \rangle} (d(x^0)^+ - w^0). \quad (29)$$

The algorithmic implementation is straightforward from the last theorem:

---

**Algorithm 1** Convex projection algorithm for optimal Lipschitz bound

---

- 1: Put  $d(x^0) = \mathbf{A}_I^{+*} \text{sign}(x_I^0)$  and  $d(x^0)^+ = d(x^0) / \|d(x^0)\|_2^2$ ;
  - 2: Compute the projection matrix onto  $K$ :  $\mathbf{P}_K = \mathbf{I}_{\mathbf{R}^m} - \bar{\mathbf{A}}_I \bar{\mathbf{A}}_I^+$ ;
  - 3: Put  $\tilde{\mathbf{a}}_j = \mathbf{P}_K \mathbf{a}_j$ ,  $j \in J$ ;
  - 4: Compute the convex projection  $w_0$  of  $d(x^0)^+$  onto the convex closure of  $\pm \tilde{\mathbf{a}}_j$ ,  $j \in J$ ;
  - 5: Compute  $Q_{opt} = \|d(x^0)^+ - w_0\|_2^{-1}$
- 

One can use any algorithm to compute the convex projection in Step 4., which clearly is the most costly in terms of complexity. In our implementation we used an  $m$ -dimensional version of the Gilbert-Johnson-Keerthi (GJK) algorithm [10]. For this algorithm for instance, each iteration comprises two steps, one step which performs a support hyperplane computation, and a second step containing a simplex convex projection. The first step is rather fast, since it relies

solely on scalar products. On the contrary, the simplex projection is the costly part, since it is itself recursive.

In the next section we show how one can roughly eliminate the type of costly computation above, at the expense of some precision in the bound estimation.

#### IV. A GREEDY ALGORITHM FOR FAST LIPSCHITZ BOUNDS ESTIMATES

In this section we construct an accelerated greedy algorithm 2 which only deals with scalar products, thus fast to compute, for producing approximative Lipschitz bound estimations.

The main idea is roughly to iteratively select a sequence of candidates  $\eta$  which decrease the denominator in (18) until a dual certificate is found. This is done by decreasing the scalar product between columns of  $\mathbf{A}$  indexed by the cosupport  $J$  of the signal and the candidate vector  $\eta$  at each step.

More precisely, as for  $IC$  criterion (11), we aim to construct a candidate  $\eta$  with a minimal  $\|\mathbf{A}_{\mathbf{J}}^* \eta\|_{\infty}$ .

The algorithm is initiated with the same generic candidate  $d(x_0)$  in (10) which also appears in the previous algorithm.

At each step  $k$ ,  $\|\mathbf{A}_{\mathbf{J}}^* \eta\|_{\infty}$  is reduced by adding a new component  $a_{j_k}$  to the vector  $\eta_{k+1} = \eta_k - t_k \mathbf{A}_{\mathbf{I}_k}^{+*} \text{sign}_k$  such that  $J_k = J_{k-1} \cup \{j_k\}$  and  $\mathbf{A}_{\mathbf{J}_k}$  injective. The vector  $\text{sign}_k$  is the appropriated sign for effectively reducing the scalar products with the selected columns. The maximum of the scalar products between these selected columns is scaled down together proportionally to  $t_k$  and, moreover,

$$|\langle a_s, \eta_k \rangle| = \|\mathbf{A}_{\mathbf{J}}^* \eta_k\|_{\infty} \quad (s \in S)$$

where  $S$  denotes the set of selected columns. When the vector  $\eta$  is modified, its scalar products with  $\mathbf{A}_{\mathbf{J}}^*$  are also affected and there are some scalar products which can increase. The descent factor  $t_k$  is computed to have the best trade-off between the reduced scalar products and those which increase.

Empirically, the  $\text{sign}_k$  is stable by  $(\mathbf{A}_{\mathbf{J}_k}^* \mathbf{A}_{\mathbf{J}_k})^{-1}$  i.e

$$\text{sign}((\mathbf{A}_{\mathbf{J}_k}^* \mathbf{A}_{\mathbf{J}_k})^{-1} \text{sign}_k) = \text{sign}_k$$

Indeed, if the sign of the scalar product of an already selected column is inverted, the scalar product scalar will increase instead of decreasing. Consequently, the factor  $t_k$  will be smaller and the maximum of the scalar products will be limited. This stability can be an indicator of a column which is no more needed.

Finally, an optional elimination step can be used for cleaning the superfluous columns at the end. For enough sparse  $x^0$ , this elimination step is not necessary.

---

#### Algorithm 2 Greedy algorithm to compute a pre-certificate with elimination

---

- 1: Calculate  $\eta_0 = \mathbf{A}_{\mathbf{I}}^{+*} \text{sign}(x^0)$
  - 2:  $\text{sign}_k \leftarrow$  vector zeros of the length of  $x^0$
  - 3:  $I_{k\_bad} \leftarrow (1)$
  - 4: **while**  $I_{k\_bad}$  is non empty **do**
  - 5:   **while**  $\mathbf{A}_{\mathbf{I}_k}$  is injective **do**
  - 6:      $j_k \leftarrow \max_{j_k \notin I_k} |\langle a_{j_k}, \eta \rangle|$
  - 7:      $\text{sign}_k \leftarrow \text{sign}_k \cup \{\text{sign}(\langle a_{j_k}, \eta \rangle)\}$
  - 8:      $I_k \leftarrow I_k \cup j_k$
  - 9:      $t_k \leftarrow \min_{i \notin I_k} t_i = \frac{\langle a_i, \eta \rangle - \langle a_{j_k}, \eta \rangle}{(\mathbf{A}^* \mathbf{A}_{\mathbf{I}_k}^{+*} \text{sign}_k)_i - (\mathbf{A}^* \mathbf{A}_{\mathbf{I}_k}^{+*} \text{sign}_k)_{j_k}}$  such that  $t_i > 0$
  - 10:      $\eta \leftarrow \eta - t_k \mathbf{A}_{\mathbf{I}_k}^{+*} \text{sign}_k$
  - 11:   **end while**
  - 12:  $I_{k\_bad} \leftarrow$  component of  $I_k$  such that  $\mathbf{A}_{\mathbf{I}_k}^* \text{sign}_k \neq \text{sign}_k$
  - 13: Elimination of components  $I_k(I_{k\_bad})$
  - 14: Elimination of components  $\text{sign}_k(I_{k\_bad})$
  - 15: **end while**
- 

#### V. NUMERICAL RESULTS

In this section, we create synthetic sets of sparse image  $x^0 \in \mathbb{R}^n$  of size  $x = 64 \times 64$ . The operator  $\mathbb{A}$  of the computed tomography will be a classical Radon transform with parallel lines. There is no real gap with a fan-beam like operator. The Radon transform is performed with the code written by Justin K. Romberg. In the noiseless case, the basis pursuit has been computed using linear programming with the Matlab optimization toolbox and OPTI toolbox [5] ( to speed up the calculations). We obtain the same results with those toolbox. In the noisy case, (3) was computed with SPGL1 toolbox [15] with the best precision.

##### A. Empirical performance of the greedy algorithm for noiseless identifiability

In this subsection we present the numerical results for the greedy algorithm regarding its ability to produce dual certificates, and compare it with other existing criteria.

We used as a reference an empirical phase transition of the  $\ell^1$  minimization is obtained, showing the degree of sparsity one can expect for within the identifiability zone. Different sets of sparse images were built for each degree of sparsity. A first one, a non-negative, where non-zero components are set to 1 and the second where non-zero components can take -1 or 1 as value. At first, a hundred signals are selected for a given sparsity and a certain number of angles. The basis pursuit is performed on each signals of a given sparsity and a probability of success is deduced. The operation is repeated for each sparsity. Results are shown as a function of the sparsity in Figure 1.

For the same set of selected sparse images, the pre-certificates are computed via the greedy algorithm and it is checked whether they are dual certificates. As for the basis pursuit, results are shown as a ratio of the number of dual certificates over the pre-certificates function of sparsity. The label  $\eta_{IC}$  denotes the candidate produced by  $IC$  criterion.  $d$  is for the Fuchs criterion that only checks the particular candidate  $d(x^0)$  - see (10) -,  $\eta_g$  is the greedy algorithm without the elimination step and  $\eta_{g2}$  is the greedy algorithm with elimination.

The projection algorithm is not represented in the Figure 2 because it finds a dual certificate for every identifiable  $x^0$ . It is a very good way to estimate the success of sparse approximations by  $\ell^1$  minimization. There is only one issue, the cost of time computation

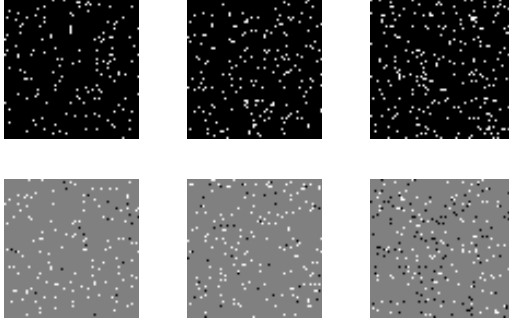


Figure 1: Examples of  $x^0 \in \mathbb{R}^{64 \times 64}$

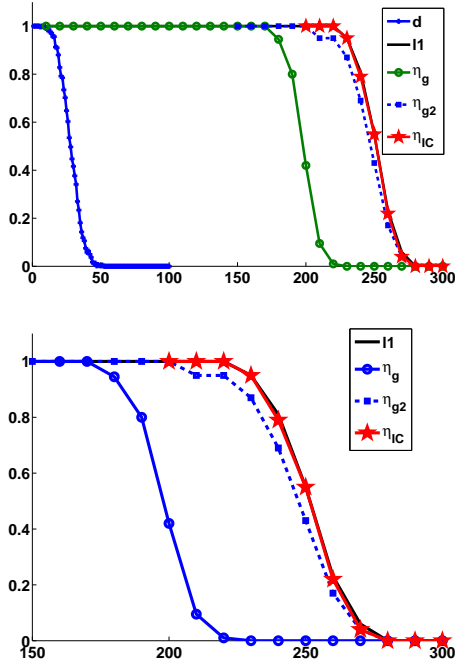


Figure 2: The ratio of vectors recovered by  $\ell^1$  minimization as a function of  $\|x^0\|_0$ . The ratio of pre-certificates  $d$ ,  $\eta_g$ ,  $\eta_{g2}$  and  $\eta_{IC}$  which are dual certificates as a function of  $\|x^0\|_0$

can be near or worst than  $\ell^1$  minimization itself as it is show in the next section. The IC covers almost the whole set of identifiable signals but it gives the worst  $Q$ . The pre-certificate  $\eta_{g2}$  is a dual certificate for nearly all of identifiable signals. One can note that the enhanced greedy algorithm has a huge gain over the simple greedy approach. The Fuchs criterion is verified only on signals with a low degree of sparsity.

### B. Projection vs Greedy

This subsection presents a comparison between the projection algorithm and the greedy algorithm regarding time complexity and quality of the Lipschitz bounds.

The table I and II present the  $Q$  obtained for different degrees of sparsity.

When the degree of sparsity approaches the phase transition,  $Q$  is higher and the time computation explodes for the greedy algorithm,

sparsity	80		100		150	
	Q	time(sec)	Q	time(sec)	Q	time(sec)
mean	11.2262	11.9678	15.29	11.38	37.44	15.13
median	11.1787	11.8658	15.27	11.33	36.71	15.01
min	9.3402	9.9907	12.45	10.47	26.53	13.15
max	13.4216	16.3981	19.35	12.85	56.54	17.37

Table I: Results for  $Q$  obtained by the projection algorithm

sparsity	80		100		150	
	Q	time(sec)	Q	time(sec)	Q	time(sec)
mean	11.5985	3.7014	15.84	4.58	42.95	9.64
median	11.5188	3.5650	15.74	4.44	41.25	9.51
min	9.5915	2.8845	12.83	3.53	29.14	6.71
max	13.7379	5.7184	20.02	6.36	72.16	14.54

Table II: Results for  $Q$  obtained by the greedy algorithm

because of the elimination of bad components. Close to the phase transition, finding a dual certificate is harder and there are many steps of elimination, thus the algorithm struggles in a long loop. Under a certain degree of sparsity, the fast greedy algorithm leads to dual certificates such that the correspondent  $Q$  is neared to those of the projection algorithm.

### C. Theoretical bounds versus real bounds

We end this section by a discussion on the quality of the theoretical Lipschitz bounds. One general feature is that these bounds always emphasize a worst case, so they can be very pessimistic for an arbitrary noise. A classical approach for testing how realistic they are is to compute the recovery of a signal with a large number of gaussian noises. In the context of computed tomography, the choice of the directions of the ray, i.e the angles between the rays and an axis, is important. All the results are obtained with a constant angular pitch, except in the worst case where we first choose the directions which produces the best results for our purpose. We also construct a "worst case" noise.

Locally, one knows that the solution  $x^\lambda$  with  $J_\lambda = \text{supp}(x^\lambda)$  is

$$x_{J_\lambda}^\lambda = x_{J_\lambda}^0 + \mathbf{A}_{J_\lambda}^{+*} b - \lambda (\mathbf{A}_{J_\lambda}^* \mathbf{A}_{J_\lambda})^{-1} \text{sign}(x_{J_\lambda}^\lambda) \quad (30)$$

One can expect to maximize  $\|x_{J_\lambda}^\lambda - x_{J_\lambda}^0\|_2$  by building a noise vector  $b$  based on the last support  $J_\lambda$  during the execution of (3) where the matrix  $(\mathbf{A}_{J_\lambda}^* \mathbf{A}_{J_\lambda})^{-1}$  is not well-conditioned. It could work for a small noise in the sense of  $\ell^2$ . Indeed, a higher noise can impact the sign. Following this idea, the algorithm 3 is proposed.

### Algorithm 3 Building a worst case noise vector

- 1: Find the last support  $J_\lambda$  such that  $(\mathbf{A}_{J_\lambda}^* \mathbf{A}_{J_\lambda})^{-1}$  is not well-conditioned
- 2:  $(U, S, V) \leftarrow \text{SVD}((\mathbf{A}_{J_\lambda}^* \mathbf{A}_{J_\lambda})^{-1})$  where  $\text{SVD}$  is the Singular Value Decomposition.
- 3:  $b \leftarrow V(1)$  where  $V(1)$  is the vector with the maximum singular value.
- 4: normalize  $b$ .

The table III presents the best results with the "worst case" noise vector for small noise and it compares to a hundred gaussian noises and the theoretical boundary. As it is shown, the initial gap in the literature has been filled, while there is still a significant difference.

One sees that the theoretical bounds differ from real bounds by some varying factor. In the case of tomography matrices, this can be explained even in the continuous domain. Indeed, if  $Rf$  is the Radon

$\ x_{J_\lambda}^\lambda - x_{J_\lambda}^0\ _2$	Mean	Max	WC
$TB_1/\ x_{J_\lambda}^\lambda - x_{J_\lambda}^0\ _2$ from theorem 1	133.9	24.8	24.02
$TB_2/\ x_{J_\lambda}^\lambda - x_{J_\lambda}^0\ _2$ from [13]	728.7	134.9	130.7

Table III: Ratio between theoretical results and real  $\ell^2$ -error.  $TB$  is the theoretical boundary, mean is the mean  $\ell^2$ -error with gaussian noises, idem with max and WC is the  $\ell^2$ -error with the "worst" case noise.

transform operator,  $Rf^*Rf$  can be seen as deconvolution operator. Duval et al [8] show in the context of sparse spike deconvolution that there can be small Dirac movements. Thus one cannot expect a too small  $\ell^2$ -recovery error. Moreover, in the discrete domain this lead to a further increase of the  $\ell^2$ -error when the resolution of the discrete grid increases. In this context, even for small noises, where one can generally hope a better approximation, the theoretical estimation of the noise robustness still remains at some factor away.

One can note that this bound not only hold for a solution  $x^\delta$  but for every solution  $x^\delta$  such that  $\|x^\delta\|_1 \leq \|x^0\|_1$ . Indeed, this theoretical result covers the worst case for this set of solutions of (3). It also explains very well the gap with the reality.

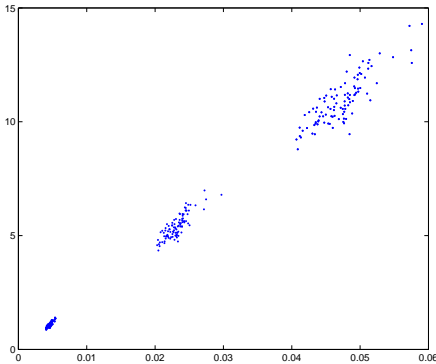


Figure 3: The theoretical boundary in function of the mean of the recovery error of a thousand different gaussian noises.

Even so, the figure 3 shows a correlation between the theoretical results and the reality. There is approximatively a ratio  $K$ .

## VI. CONCLUSION

The presented algorithms allow the numerical optimization of some of the most recent and versatile theoretical results concerning noise-robustness in compressive sensing. Numerical results show that they are rather complementary with respect to different sparsity regimes. Also, one can choose one or the other according to the potential application priorities, speed versus quality. In addition, due to their inherently and natural geometrical content, part of the ideas behind the proposed algorithms is susceptible to be adapted for future refinements of theoretical noise-robustness estimations.

## ACKNOWLEDGEMENT

This work was supported by the TOTAL SA. Authors would like to thank P. Ricoux of TOTAL SA. for his advice and useful discussions.

## REFERENCES

- [1] Martin Burger and Stanley Osher. Convergence rates of convex variational regularization. *Inverse problems*, 20(5):1411, 2004.
- [2] Emmanuel Candes and Carlos Fernandez-Granda. Super-resolution from noisy data. *arXiv preprint arXiv:1211.0290*, 2012.
- [3] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 2013.
- [4] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, December 1998.
- [5] Jonathan Currie and David I. Wilson. OPTI: Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User. In Nick Sahinidis and Jose Pinto, editors, *Foundations of Computer-Aided Process Operations*, Savannah, Georgia, USA, 8–11 January 2012.
- [6] David L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, 2005.
- [7] Charles Dossal and Remi Tesson. Consistency of  $\ell_1$  recovery from noisy deterministic measurements. *Applied and Computational Harmonic Analysis*, 2013.
- [8] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *arXiv preprint arXiv:1306.6909*, 2013.
- [9] J. J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theor.*, 50(6):1341–1344, June 2004.
- [10] Elmer G Gilbert, Daniel W Johnson, and S Sathiya Keerthi. A fast procedure for computing the distance between complex objects in three-dimensional space. *Robotics and Automation, IEEE Journal of*, 4(2):193–203, 1988.
- [11] Markus Grasmair. Linear convergence rates for tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 2011.
- [12] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. Sparse regularization with  $\ell_1$  penalty term. *Inverse Problems*, 24(5):055020, 2008.
- [13] Markus Grasmair, Otmar Scherzer, and Markus Haltmeier. Necessary and sufficient conditions for linear convergence of  $\ell_1$ -regularization. *Commun. Pure and Appl. Math.*, 64(2):161–182, 2011.
- [14] Samuel Vaiter, Gabriel Peyré, Charles Dossal, and Jalal Fadili. Robust sparse analysis regularization. 2011.
- [15] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.