

# Damping Noise-Folding and Enhanced Support Recovery in Compressed Sensing

Extended Technical Report

Marco Artina,<sup>\*</sup> Massimo Fornasier,<sup>†</sup> and Steffen Peter<sup>‡</sup>

Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany

November 24, 2014

The practice of compressed sensing suffers importantly in terms of the efficiency/accuracy trade-off when acquiring noisy signals prior to measurement. It is rather common to find results treating the noise affecting the measurements, avoiding in this way to face the so-called *noise-folding* phenomenon, related to the noise in the signal, eventually amplified by the measurement procedure. In this paper, we present two new decoding procedures, combining  $\ell_1$ -minimization followed by either a regularized selective least  $p$ -powers or an iterative hard thresholding, which not only are able to reduce this component of the original noise, but also have enhanced properties in terms of support identification with respect to the sole  $\ell_1$ -minimization or iteratively re-weighted  $\ell_1$ -minimization. We prove such features, providing relatively simple and precise theoretical guarantees. We additionally confirm and support the theoretical results by extensive numerical simulations, which give a statistics of the robustness of the new decoding procedures with respect to more classical  $\ell_1$ -minimization and iteratively re-weighted  $\ell_1$ -minimization. Noise folding in compressed sensing, support identification,  $\ell_1$ -minimization, selective least  $p$ -powers, iterative hard thresholding, phase transitions.

2000 Math Subject Classification: 94A12, 94A20, 65F22, 90C05, 90C30

## 1 Introduction

Compressive sensing focuses on the robust recovery of nearly sparse vectors from the minimal amount of measurements obtained by a randomized linear process. So far, a vast literature appeared considering problems where deterministic or random noise is added after the measurement process, while it is not strictly related to the signal. One typically considers model problems of the type

$$y = Ax + w \tag{1}$$

---

<sup>\*</sup>Marco.Artina@ma.tum.de

<sup>†</sup>Massimo.Fornasier@ma.tum.de

<sup>‡</sup>Steffen.Peter@ma.tum.de

where  $x \in \mathbb{R}^N$  is a nearly sparse vector,  $A \in \mathbb{R}^{m \times N}$  is the linear measurement matrix,  $y \in \mathbb{R}^m$  is the result of the measurement, and  $w$  is a white noise vector affecting the measurements. However, in practice it is very uncommon to have a signal detected by a certain device, totally free from some external noise. Therefore, it is reasonable to consider the more realistic model

$$y = A(x + n) + w,$$

instead of (1) where  $n \in \mathbb{R}^N$  is the noise on the original signal.

The recent work [2, 32] shows how the measurement process actually causes the *noise-folding phenomenon*, which implies that the variance of the noise on the original signal is amplified by a factor of  $\frac{N}{m}$ , additionally contributing to the measurement noise, playing to our disadvantage in the recovery phase. We refer also to the papers [34, 35], which seem to have independently reconsidered this problem more recently. More formally, if we add to the signal  $x$  a noise vector  $n$ , whose entries have normal distribution  $\mathcal{N}(0, \sigma_n)$ , the measurement  $y$  given by

$$y = A(x + n), \tag{2}$$

can be considered equivalently obtained by a measurement procedure of the form (1) (possibly with another measurement matrix  $A$  of equal statistics) where now the vector  $w$  is composed by i.i.d. Gaussian entries with distribution  $\mathcal{N}(0, \sigma_w)$  and  $\sigma_w^2 = \frac{N}{m} \sigma_n^2$ . In this stochastic context, the so-called Dantzig selector has been analyzed in [8] showing that the recovered signal  $x^*$  from the measurement  $y$  fulfills the following nearly-optimal distortion guarantees, under the assumption that  $A$  satisfies the so-called restricted isometry property (compare Definition 4):

$$\|x - x^*\|_{\ell_2}^2 \leq C^2 \cdot 2 \log N \cdot \left( \sigma_w^2 + \sum_{i=1}^N \min\{x_i^2, \sigma_w^2\} \right), \tag{3}$$

which, for a sparse vector  $x$  with at most  $k$ -nonzero entries, reduces to the following estimate

$$\begin{aligned} \|x - x^*\|_{\ell_2}^2 &\leq C^2 \cdot 2 \log N \cdot ((1 + k) \sigma_w^2) \\ &= C^2 \cdot 2 \log N \cdot \left( (1 + k) \frac{N}{m} \sigma_n^2 \right). \end{aligned}$$

Therefore, the noise-folding phenomenon may significantly reduce in practice the potential advantages of compressed sensing in terms of the trade-off between robustness and efficient compression (here represented by the factor  $\frac{N}{m}$ ), with respect to other more traditional subsampling encoding methods [15].

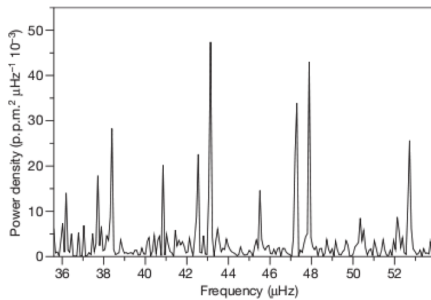
In the following, we wish to focus on two fundamental consequences of noise folding: the loss of accuracy in the recovery of the large entries of the original vector  $x$ , and the correct detection of their index support.

To this end, let us introduce for  $r > \eta > 0$  and  $1 \leq k < m$  the class of *sparse vectors affected by bounded noise*,

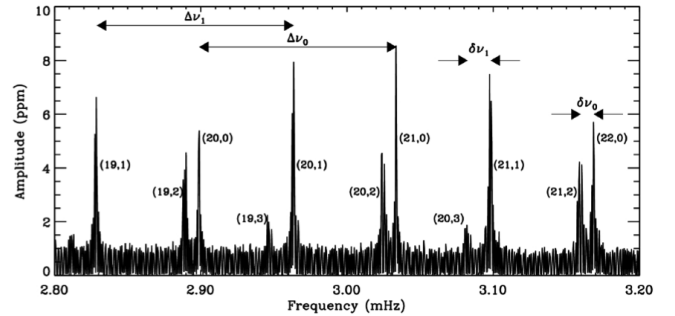
$$\mathcal{S}_{\eta, k, r}^p := \left\{ x \in \mathbb{R}^N \mid \#S_r(x) \leq k \text{ and } \sum_{i \in (S_r(x))^c} |x_i|^p \leq \eta^p \right\}, \quad 1 \leq p \leq 2, \tag{4}$$

where  $S_r(x) := \{i \in \{1, \dots, N\} \mid |x_i| > r\}$  is the index support of the large entries exceeding in absolute

value the threshold  $r$ . This class contains all vectors for which at most  $1 \leq k < m$  large entries exceed the threshold  $r$  in absolute value, while the  $p$ -norm of the other entries stays below a certain noise level. To highlight the concrete modeling potential of such a class, in Figure 1 we show typical examples of real-life signals from Asteroseismology falling into it, for some choice of the parameters  $\eta, k, r$ .



(a) Power density spectrum of the red giant candidate CoRoT-101034881 showing a frequency pattern with a regular spacing [30].



(b) Amplitude spectra of solar oscillations measured by the VIRGO instrument on SOHO. The figure shows only a portion of the solar oscillation spectrum [6].

Figure 1: Examples of signals out of the class defined in (4).

Let us now informally explain how the class  $\mathcal{S}_{\eta,k,r}^p$  can be crucially used to analyze the effects of noise folding in terms of support recovery depending on the parameters  $\eta, r, k$ . Therefore, assume now that  $x$  is a sparse vector with at most  $k$  nonzero entries exceeding the threshold  $r$  in absolute value and  $x + n \in \mathcal{S}_{\eta,k,r}^2$  in expectation (notice that we specified here  $p = 2$ ). By the statistical equivalence mentioned above of the model (2) and (1), we infer that the recovered vector  $x^*$  by means of the Dantzig selector will fulfill the following error estimate:

$$\begin{aligned} \|x - x^*\|_{\ell_2}^2 &\leq C^2 \cdot 2 \log N \cdot \left( (1+k) \frac{N}{m} \sigma_n^2 \right) \\ &\leq C^2 \cdot 2 \log N \cdot \left( (1+k) \frac{N}{m} \frac{\eta^2}{N-k} \right), \end{aligned} \quad (5)$$

where the last inequality follows by the requirement

$$(N-k)\sigma_n^2 = \mathbb{E} \left( \sum_{i \in (S_r(x))^c} |n_i|^2 \right) \leq \eta^2,$$

considering here  $S_r(x)$  as previously defined in (4). Since we assume that  $k \ll N$  and  $\frac{k+1}{m} \leq 1$ , the right-hand-side of (5) can be further bounded from above by

$$C^2 \cdot 2 \log N \cdot \left( (1+k) \frac{N}{m} \frac{\eta^2}{N-k} \right) \leq C_1^2 \cdot 2 \log N \cdot \eta^2.$$

It easily follows (and we will use similar arguments below for different decoding methods) that a sufficient condition for the identification of the support of  $x$ , i.e.,  $\text{supp}(x) \subset \text{supp}(x^*)$ , is

$$C_1^2 \cdot 2 \log N \eta^2 < r^2,$$

or, equivalently,

$$\eta < \frac{r}{C_1 \sqrt{2 \log N}}.$$

Notice that such a sufficient condition on  $\eta$  actually implies a rather large gap between the significant entries of  $x$  and the noise components in  $n$ . Hence, it is of utmost practical interest to understand how small this gap is actually allowed to be, i.e., how large  $\eta$  can be relatively to  $r$ , for the most used recovery algorithms in compressed sensing (not only the Dantzig selector) to be able to have both support identification and a good approximation of the significant entries of  $x$ .

An approach to control the noise folding, is proposed in [2]. In this case, one may tune the linear measurement process in order to a priori filter the noise. However, this strategy requires to have a precise knowledge of the noise statistics and to design proper filters. Other related work [21, 22, 23] addresses the problem of designing adaptive measurements, called *distilled sensing*, in order to detect and locate the signal within white noise. In this paper, we shall follow a *blind-to-statistics* approach (although, as recalled below, we do not consider the scenario of impulsive noise), which does not modify the non-adaptive measurements, and, differently from the Dantzig selector analysis in [8], we restrict ourself to a purely deterministic setting.

First of all, we show that, unfortunately, the classical  $\ell_1$ -minimization, but also the iteratively re-weighted  $\ell_1$ -minimization [10, 27], considered one of the most robust in the field, easily fails in both the tasks mentioned above as soon as  $\eta \approx r$ . The deep reason of this failure is the lack of selectivity of these algorithms, which are designed to promote not only the sparsity of the signal  $x$  but also of the recovered noise. This principle drawback affects necessarily also recent modifications of re-weighted  $\ell_1$ -minimization as appearing in [25]. This has the consequence, as a sort of balancing principle, that the miscomputed components of the noise  $n$ , if it is not originally of impulsive nature, blow up the inaccuracy in the detection and approximation of  $x$ . (Notice that the problem of separating sparse signals within impulsive noise is much harder and it will not be considered within the scope of the present paper.)

To overcome these difficulties of these popular methods, we propose a new decoding procedure, combining  $\ell_1$ -minimization and regularized *selective least  $p$ -powers*, which is able to reduce the noise component affecting the signal and also to enhance the support identification. In fact, for certain applications, such as radar [24], the support recovery can be even more relevant than an accurate estimate of the signal values. Although the analysis and the numerical results of this new procedure greatly outperform  $\ell_1$ -minimization and its re-weighted iterative version, its computational complexity becomes prohibitive in very high-dimension. For we conclude by proposing and analyzing an eventual method, based on similar principles, which combines again a warm-up step based on  $\ell_1$ -minimization with a nonconvex optimization realized by the well-known iterative hard thresholding [7], and a final correction step realized by a convex optimization. We show that this latter method performs as robustly as the previous one, but with a drastic reduction of the complexity.

The paper is organized as follows. In the next section, we concisely recall the pertinent features of the theory of compressive sensing. In Section 3, we shall describe the limitations of  $\ell_1$ -minimization when noise on the signal is present, and we mention that very similarly an analogue analysis can be performed for the iteratively re-weighted  $\ell_1$ -minimization based on the results in [27]. Afterwards, as an alternative, we propose the linearly constrained minimization of the regularized selective  $p$ -potential functional, and show that certain sufficient conditions for recovery indicate significantly better performance than the one provided by  $\ell_1$ -minimization and iteratively re-weighted  $\ell_1$ -minimization. Within this paper, we measure the *performance* of a method by its ability to identify and approximate the relevant entries of the original signal, where the *relevant entries* of a signal are the ones which exceed

a predefined threshold  $r$ . For this purpose, we introduced the class (4). With respect to the notation in the introduction, this class is changing the meaning of the notation  $x$  throughout the rest of the paper and makes redundant the notation  $x + n$ , which emphasizes the fact that the noise is directly on the signal. Section 4 recalls the main properties of a very robust and efficient algorithm to perform the linearly constrained minimization of the regularized selective least  $p$ -powers. In Section 5, we address the issue of the high computational cost of the regularized selective  $p$ -potential optimization and propose, exploiting a similar *selectivity principle*, an alternative method based on iterative hard thresholding. Finally, in Section 6, we report the results of extensive numerical experiments, which we made to illustrate and support our theoretical guarantees, and the comparisons between all the mentioned decoding methods.

## 2 Compressive sensing

It is possible to uniquely and robustly identify the solution  $x \in \mathbb{R}^N$  of the linear system  $Ax = y$  for an arbitrary given measurement vector  $y \in \mathbb{R}^m$ , if  $A \in \mathbb{R}^{m \times N}$  has rank  $m$ , and  $m = N$ . However, in many applications, we may be either not able to take enough measurements, or interested in taking much fewer measurements to save costs or time, i.e.,  $m \ll N$ . The theory of compressive sensing studies this scenario under some restrictions, and assumes that the original signal  $x$  is nearly sparse. In this section, we recall concisely terms and principles of this theory, and we refer to some of the known tutorials for more details [4, 9, 19, 17].

In compressive sensing, we call the matrix  $A$  the *encoder* which transforms the  $N$ -dimensional signal  $x$  into to the *measurement vector*  $y \in \mathbb{R}^m$  of dimension  $m \ll N$ . Further, we assume  $A$  to have rank  $m$  from now on in this article. In practice, we do not know  $x$  and wonder if it is possible to recover it somehow robustly by an efficient nonlinear *decoder*  $\Delta: \mathbb{R}^m \rightarrow \mathbb{R}^N$ . As already mentioned, the theory only works if we assume the signal  $x$  to be sparse or at least compressible.

**Definition 1** (*k*-sparse vector). *Let  $k \in \mathbb{N}^+$ ,  $k \leq N$ . We call the vector  $x \in \mathbb{R}^N$   $k$ -sparse if*

$$x \in \Sigma_k := \{z \in \mathbb{R}^N \mid \#\text{supp}(z) \leq k\}$$

where  $\text{supp}(z) := \{i \in \{1, \dots, N\} \mid z_i \neq 0\}$  denotes the support of  $z$ .

In applications, signals are often not exactly sparse but at least compressible. We refer to [26] for more details. We define compressibility in terms of the *best  $k$ -term approximation error* with respect to the  $\ell_p$ -norm, given by

$$\|x\|_{\ell_p} = \left( \sum_{i=1}^N |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty.$$

**Definition 2** (Best  $k$ -term approximation). *Let  $x$  be an arbitrary vector in  $\mathbb{R}^N$ . We denote the best  $k$ -term approximation of  $x$  by*

$$x_{[k]} := \arg \min_{z \in \Sigma_k} \|x - z\|_{\ell_p}, \quad 1 \leq p < \infty,$$

and the respective best  $k$ -term approximation error of  $x$  by

$$\sigma_k(x)_{\ell_p} := \min_{z \in \Sigma_k} \|x - z\|_{\ell_p} = \|x - x_{[k]}\|_{\ell_p}.$$

**Remark 1.** *The best  $k$ -term approximation error is the minimal distance of  $x$  to a  $k$ -sparse vector. Informally, vectors having a relatively small best  $k$ -term approximation error are considered to be compressible.*

**Remark 2.** *If we define the nonincreasing rearrangement of  $x$  by*

$$r(x) = (|x_{i_1}|, \dots, |x_{i_N}|)^T, \text{ and } |x_{i_j}| \geq |x_{i_{j+1}}|, j = 1, \dots, N-1,$$

then

$$\sigma_k(x)_{\ell_p} = \left( \sum_{j=k+1}^N r_j(x)^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty.$$

Alternatively, we can describe the best- $k$ -term approximation error by

$$\sigma_k(x)_{\ell_p} = \left( \sum_{j \in \Lambda^c} |x_j|^p \right)^{\frac{1}{p}},$$

where  $\Lambda := \text{supp}(x_{[k]})$ , and  $\Lambda^c$  is its complement in  $\{1, \dots, N\}$ .

A desirable property of an encoder/decoder pair  $(A, \Delta)$  is given by the following stability estimate, called *instance optimality*

$$\|x - \Delta(Ax)\|_{\ell_p} \leq C \sigma_k(x)_{\ell_p}, \quad (6)$$

for all  $x \in \mathbb{R}^N$ , with a positive constant  $C$  independent of  $x$ , and  $k$  the closest possible to  $m$  [12]. This would in particular imply that by means of  $\Delta$  we are able to recover a  $k$ -sparse signal  $x$  exactly, since in this case  $\sigma_k(x)_{\ell_p} = 0$ . It turns out that the existence of such a pair restricts the range of  $k$  to be maximally of the order of  $\frac{m}{\log \frac{m}{N} + 1}$ . We refer to [5, 12, 16] for more details.

Actually, the above mentioned condition (6) can be realized in practice, at least for  $p = 1$ , by pairing the  $\ell_1$ -minimization as the decoder with the choice of an encoder which has the so-called *Null Space Property* of optimal order  $k$ . (For realizations of the instance optimality in other  $\ell_p$ -norms, for instance for  $p = 2$ , one needs more restrictive requirements, see [33]. For the analysis within this paper, we shall use (6) just for  $p = 1$  for the sake of simplicity.)

**Definition 3** (Null Space Property). *A matrix  $A \in \mathbb{R}^{m \times N}$  has the Null Space Property of order  $k$  and for positive constant  $\gamma_k > 0$  if*

$$\|z|_{\Lambda}\|_{\ell_1} \leq \gamma_k \|z|_{\Lambda^c}\|_{\ell_1},$$

for all  $z \in \ker A$  and all  $\Lambda \subset \{1, \dots, N\}$  such that  $\#\Lambda \leq k$ . We abbreviate this property with the writing  $(k, \gamma_k)$ -NSP.

The Null Space Property states that the kernel of the encoding matrix  $A$  contains no vectors where some entries have a significantly larger magnitude with respect to the others. In particular, no compressible vector is contained in the kernel. This is a natural requirement since otherwise no decoder would be able to distinguish a sparse vector from zero.

**Lemma 1.** *Let  $A \in \mathbb{R}^{m \times N}$  have the  $(k, \gamma_k)$ -NSP, with  $\gamma_k < 1$ , and define*

$$\mathcal{F}(y) := \{z \in \mathbb{R}^N | Az = y\},$$

the set of feasible vectors for the measurement vector  $y \in \mathbb{R}^m$ . Then the decoder

$$\Delta_1(y) := \arg \min_{z \in \mathcal{F}(y)} \|z\|_{\ell_1}, \quad (7)$$

which we call  $\ell_1$ -minimization, performs

$$\|x - \Delta_1(y)\|_{\ell_1} \leq C \sigma_k(x)_{\ell_1}, \quad (8)$$

for all  $x \in \mathcal{F}(y)$  and the constant  $C := \frac{2(1+\gamma_k)}{1-\gamma_k}$ .

Although the result we stated here is by now well-known, see also [31], we report its short proof for the sake of completeness, and for comparison with the enhanced guarantees given in Theorem 4 below.

*Proof.* Let us denote  $x^* = \Delta_1(y)$  and  $z = x^* - x$ . Then  $z \in \ker A$  and

$$\|x^*\|_{\ell_1} \leq \|x\|_{\ell_1},$$

because  $x^*$  is a solution of the  $\ell_1$ -minimization problem (7). Let  $\Lambda$  be the set of the  $k$ -largest entries of  $x$  in absolute value. One has

$$\|x^*|_{\Lambda}\|_{\ell_1} + \|x^*|_{\Lambda^c}\|_{\ell_1} \leq \|x|_{\Lambda}\|_{\ell_1} + \|x|_{\Lambda^c}\|_{\ell_1}.$$

It follows immediately from the triangle inequality that

$$\|x|_{\Lambda}\|_{\ell_1} - \|z|_{\Lambda}\|_{\ell_1} + \|z|_{\Lambda^c}\|_{\ell_1} - \|x|_{\Lambda^c}\|_{\ell_1} \leq \|x|_{\Lambda}\|_{\ell_1} + \|x|_{\Lambda^c}\|_{\ell_1}.$$

Hence, by the  $(k, \gamma_k)$ -NSP

$$\|z|_{\Lambda^c}\|_{\ell_1} \leq \|z|_{\Lambda}\|_{\ell_1} + 2\|x|_{\Lambda^c}\|_{\ell_1} \leq \gamma_k \|z|_{\Lambda^c}\|_{\ell_1} + 2\sigma_k(x)_{\ell_1},$$

or, equivalently,

$$\|z|_{\Lambda^c}\|_{\ell_1} \leq \frac{2}{1-\gamma_k} \sigma_k(x)_{\ell_1}. \quad (9)$$

Finally, again by the  $(k, \gamma_k)$ -NSP

$$\|x - x^*\|_{\ell_1} = \|z|_{\Lambda}\|_{\ell_1} + \|z|_{\Lambda^c}\|_{\ell_1} \leq (\gamma_k + 1) \|z|_{\Lambda^c}\|_{\ell_1} \leq \frac{2(1+\gamma_k)}{1-\gamma_k} \sigma_k(x)_{\ell_1},$$

and the proof is completed.  $\square$

Unfortunately, the NSP is hard to verify in practice. Therefore one can introduce another property which is called the *Restricted Isometry Property* which implies the NSP. Being a spectral concentration property, the Restricted Isometry Property is particularly suited to be verified with high probability by certain random matrices; we mention some instances of such classes of matrices below.

**Definition 4** (Restricted Isometry Property). *A matrix  $A \in \mathbb{R}^{m \times N}$  has the Restricted Isometry Property (RIP) of order  $K$  with constant  $0 < \delta_K < 1$  if*

$$(1 - \delta_K) \|z\|_{\ell_2} \leq \|Az\|_{\ell_2} \leq (1 + \delta_K) \|z\|_{\ell_2},$$

for all  $z \in \Sigma_K$ . We refer to this property by  $(K, \delta_K)$ -RIP.

**Lemma 2.** *Let  $k, h \in \mathbb{N}^+$  and  $K = k + h$ . Assume that  $A \in \mathbb{R}^{m \times N}$  has  $(K, \delta_K)$ -RIP. Then  $A$  has  $(k, \gamma_k)$ -NSP, where*

$$\gamma_k := \sqrt{\frac{k}{h} \frac{1 + \delta_K}{1 - \delta_K}}.$$

The proof of this latter result can be found, for instance, in [17], and not being of specific relevance for this paper we do not include it here, although it will be important at some points below to connect the RIP to a corresponding NSP. Encoders which have the RIP with optimal constants, i.e., with  $k$  in the order of  $\frac{m}{\log \frac{m}{N} + 1}$  exist, but, so far, as mentioned above, they can be realized exclusively by randomization. By now, classical examples of stochastic encoders are i.i.d. Gaussian matrices [16] or discrete Fourier matrices with randomly chosen rows [11]. Further details and generalizations are provided in [19, 29]. In the rest of the paper we will use as prototypical cases mainly such stochastic encoders.

A recent ansatz to increase the reconstruction accuracy of sparse vectors is iteratively re-weighted  $\ell_1$ -minimization [10, 27]. The idea of this approach is to iteratively solve the weighted  $\ell_1$ -minimization

$$z^{n+1} = \arg \min_{z \in \mathcal{F}(y)} \sum_{i=1}^N w_i^n |z_i|,$$

while updating the weights according to  $w_i^n = (|z_i^n| + a)^{-1}$  for all  $i = 1, \dots, N$ , for a suitably chosen stability parameter  $a > 0$ . We denote this decoder by  $\Delta_{\text{Irew}}$ , and recall the following respective instance optimality result.

**Lemma 3** ([27, Theorem 3.2]). *Let  $A \in \mathbb{R}^{m \times N}$  have the  $(2k, \delta_{2k})$ -RIP, with  $\delta_{2k} < \sqrt{2} - 1$ , and assume the smallest nonzero coordinate of  $x_{[k]}$  in absolute value larger than the threshold*

$$\bar{r} := 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left( \sigma_k(x)_{\ell_2} + \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \right). \quad (10)$$

Then the decoder  $\Delta_{\text{Irew}}$  performs

$$\|x - \Delta_{\text{Irew}}(y)\|_{\ell_2} \leq 4.8 \frac{\sqrt{1 + \delta_{2k}}}{1 + (\sqrt{2} - 1)\delta_{2k}} \left( \sigma_k(x)_{\ell_2} + \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \right). \quad (11)$$

What we recalled up to this point of the theory of compressive sensing tells us that we are able to recover by (iteratively re-weighted)  $\ell_1$ -minimization compressible vectors within a certain accuracy, given by (8) or (11) respectively. If we re-interpret compressible vectors as sparse vectors which are corrupted by noise, we immediately see that the accuracy of the recovered solution is basically driven by the noise level affecting the vector. Nevertheless, neither inequalities (8) and (11) tell us immediately if the recovered support of the  $k$  largest entries of the decoded vector in absolute value is the same as the one of the original signal nor are we able to identify the large entries exceeding a given threshold in absolute value. Section 3 addresses these issues in detail and investigates the limitations of  $\ell_1$ -minimization and iteratively re-weighted  $\ell_1$ -minimization.

Furthermore, we propose a new decoder, which is able to outperform  $\ell_1$ -minimization and iteratively re-weighted  $\ell_1$ -minimization in terms of simultaneously identifying the exact position of the significant entries and reducing the noise level on the signal.



### 3 Support identification

We have seen in Lemma 1 that sparse vectors can be recovered exactly by the  $\ell_1$ -minimization decoder  $\Delta_1$  if the matrix has the NSP. Moreover, a sparse signal which is disturbed by noise is recovered within a certain accuracy depending on the best  $k$ -term approximation error. In this section, we investigate in detail the noise level we can tolerate without losing the ability of  $\ell_1$ -minimization to recover the support of the undisturbed sparse signal.

For later use, let us denote, for  $1 \leq p \leq 2$  and  $q$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ ,

$$\kappa_p := \kappa_p(N, k) := \begin{cases} 1, & p = 1, \\ \sqrt[q]{N - k}, & 1 < p \leq 2. \end{cases} \quad (12)$$

#### 3.1 The $\ell_1$ -minimization result

The following simple proposition shows how we can recover the support of the original signal if we know the  $\ell_1$ -minimizer. It turns out that the large entries of the original signal in absolute value need to exceed a certain threshold, which depends on the noise level.

**Theorem 1.** *Let  $x \in \mathbb{R}^N$  be a noisy signal with  $k$  relevant entries and the noise level  $\eta \in \mathbb{R}$ ,  $\eta \geq 0$ , i.e., for  $\Lambda = \text{supp}(x_{[k]})$ ,*

$$\sum_{j \in \Lambda^c} |x_j|^p \leq \eta^p, \quad (13)$$

for a fixed  $1 \leq p \leq 2$ . Consider further an encoder  $A \in \mathbb{R}^{m \times N}$  which has the  $(k, \gamma_k)$ -NSP, with  $\gamma_k < 1$ , the respective measurement vector  $y = Ax \in \mathbb{R}^m$ , and the  $\ell_1$ -minimizer

$$x^* := \arg \min_{z \in \mathcal{F}(y)} \|z\|_{\ell_1}.$$

If the  $i$ -th component of the original signal  $x$  is such that

$$|x_i| > \frac{2(1 + \gamma_k)}{1 - \gamma_k} \kappa_p \eta, \quad (14)$$

then  $i \in \text{supp}(x^*)$ .

*Proof.* We know by (8) that

$$\|x^* - x\|_{\ell_1} \leq \frac{2(1 + \gamma_k)}{1 - \gamma_k} \sigma_k(x)_{\ell_1}. \quad (15)$$

Thus, by Hölder's inequality and the assumption (13), we obtain the estimate

$$\|x^* - x\|_{\ell_1} \leq \frac{2(1 + \gamma_k)}{1 - \gamma_k} \sigma_k(x)_{\ell_1} \leq \frac{2(1 + \gamma_k)}{1 - \gamma_k} \kappa_p \eta. \quad (16)$$

We now choose a component  $i \in \{1, \dots, N\}$  such that

$$|x_i| > \frac{2(1 + \gamma_k)}{1 - \gamma_k} \kappa_p \eta,$$

and assume  $i \notin \text{supp}(x^*)$ . This leads to the contradiction:

$$|x_i| = |x_i - x_i^*| \leq \|x - x^*\|_{\ell_1} \leq \frac{2(1 + \gamma_k)}{1 - \gamma_k} \kappa_p \eta < |x_i|. \quad (17)$$

Hence, necessarily  $i \in \text{supp}(x^*)$ .  $\square$

The noise level substantially influences the ability of support identification. Here, the noisy signal should have (as a sufficient condition) the  $k$  largest entries in absolute value above

$$r_1 := \frac{2(1 + \gamma_k)}{1 - \gamma_k} \kappa_p \eta,$$

in order to guarantee support identification.

### 3.2 The result for iteratively re-weighted $\ell_1$ -minimization

We are able also in the case of the iteratively re-weighted  $\ell_1$ -minimization to show a similar support identification result, which follows similarly from Lemma 3.

**Theorem 2.** *Let  $x \in \mathbb{R}^N$  be a noisy signal with  $k$  relevant entries and the noise level  $\eta \in \mathbb{R}$ ,  $\eta \geq 0$ , i.e., for  $\Lambda = \text{supp}(x_{[k]})$ ,*

$$\sum_{j \in \Lambda^c} |x_j|^p \leq \eta^p, \quad (18)$$

for a fixed  $1 \leq p \leq 2$ . Consider further an encoder  $A \in \mathbb{R}^{m \times N}$  which has the  $(2k, \delta_{2k})$ -RIP, with  $\delta_{2k} < \sqrt{2} - 1$ , the respective measurement vector  $y = Ax \in \mathbb{R}^m$ , and the iteratively re-weighted  $\ell_1$ -minimizer  $x^* := \Delta_{1rew}(y)$ . If for all  $i \in \text{supp}(x_{[k]})$

$$|x_i| > 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left(1 + \frac{\kappa_p}{\sqrt{k}}\right) \eta, \quad (19)$$

then  $\text{supp}(x_{[k]}) \subset \text{supp}(x^*)$ .

*Proof.* First, notice that  $\eta \geq \sqrt[p]{\sum_{j \in \Lambda^c} |x_j|^p} = \sigma_k(x)_{\ell_p} \geq \sigma_k(x)_{\ell_2}$  and  $\kappa_p \sigma_k(x)_{\ell_p} \geq \sigma_k(x)_{\ell_1}$  by Hölder's inequality. Thus, we have for all  $i \in \text{supp}(x_{[k]})$  that

$$\begin{aligned} |x_i| &> 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left(1 + \frac{\kappa_p}{\sqrt{k}}\right) \eta \geq 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left(\sigma_k(x)_{\ell_p} + \frac{\kappa_p}{\sqrt{k}} \sigma_k(x)_{\ell_p}\right) \\ &\geq 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left(\sigma_k(x)_{\ell_2} + \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}}\right). \end{aligned}$$

Consequently, we fulfill the conditions of Lemma 3 for which, for all  $i \in \text{supp}(x_{[k]})$ ,  $|x_i| > \bar{r}$ , as defined in (10).

Assume now that there is  $i \in \text{supp}(x_{[k]})$  and  $i \notin \text{supp}(x^*)$ . By means of Lemma 3, we obtain the

contradiction

$$\begin{aligned}
|x_i| &= |x_i - x_i^*| \leq \|x - x^*\|_{\ell_2} \leq 4.8 \frac{\sqrt{1 + \delta_{2k}}}{1 + (\sqrt{2} - 1)\delta_{2k}} \left( \sigma_k(x)_{\ell_2} + \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \right) \\
&\leq 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left( \sigma_k(x)_{\ell_2} + \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}} \right) = \bar{r} < |x_i|.
\end{aligned} \tag{20}$$

Hence,  $i \in \text{supp}(x^*)$ . □

As negative aspects  $\Delta_{1\text{rew}}$ , we notice that, in view of (20) already the conditions of Lemma 3 for the stable convergence of the algorithm do imply support identification. Moreover, we stress that Theorem 2 requires the RIP, which is in general a stronger condition than the NSP. Again, one observes that the noise level influences the ability of support identification. Here, the noisy signal should have the  $k$  largest entries in absolute value above

$$r_{1\text{rew}} := 9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left( 1 + \frac{\kappa_p}{\sqrt{k}} \right) \eta,$$

in order to guarantee support identification.

In the following, we shall show that for the class  $\mathcal{S}_{\eta,k,r}^p$ , as defined in (4), a smaller threshold is required for support identification, under the sole request of the NSP (and not of the RIP).

### 3.3 Support identification stability in the class $\mathcal{S}_{\eta,k,r}^p$

In this section, we present results in terms of support discrepancy once we consider two elements of the class  $\mathcal{S}_{\eta,k,r}^p$  defined in (4), having the same measurements, i.e., they are both in  $\mathcal{F}(y)$  for  $y \in \mathbb{R}^m$ .

**Theorem 3.** *Let  $A \in \mathbb{R}^{m \times N}$  have the  $(2k, \gamma_{2k})$ -NSP, for  $\gamma_{2k} < 1$ ,  $1 \leq p \leq 2$ , and  $x, x' \in \mathcal{S}_{\eta,k,r}^p$  such that  $Ax = Ax'$ , and  $0 \leq \eta < r$ . Then*

$$\#(S_r(x) \Delta S_r(x')) \leq \frac{(2\gamma_{2k}\kappa_p\eta)^p}{(r - \eta)^p}. \tag{21}$$

(Here we denote by “ $\Delta$ ” the set symmetric difference, not to be confused with the symbol of a generic decoder.) If additionally

$$r > \eta(1 + 2\gamma_{2k}\kappa_p) =: r_S, \tag{22}$$

then  $S_r(x) = S_r(x')$ , i.e., we have unique identification of the large entries in absolute value.

*Proof.* As  $Ax = Ax'$ , the difference  $(x - x') \in \ker(A)$ . By the  $(2k, \gamma_{2k})$ -NSP, Hölder’s inequality, and the triangle inequality we have

$$\begin{aligned}
\|(x - x')|_{S_r(x) \cup S_r(x')}\|_{\ell_p} &\leq \|(x - x')|_{S_r(x) \cup S_r(x')}\|_{\ell_1} \\
&\leq \gamma_{2k} \|(x - x')|_{(S_r(x) \cup S_r(x'))^c}\|_{\ell_1} \\
&\leq \gamma_{2k}\kappa_p \|(x - x')|_{(S_r(x) \cup S_r(x'))^c}\|_{\ell_p} \\
&\leq 2\gamma_{2k}\kappa_p\eta.
\end{aligned} \tag{23}$$

Now we estimate the symmetric difference of the supports of the large entries of  $x$  and  $x'$  in absolute value as follows: if  $i \in S_r(x) \Delta S_r(x')$ , then either  $|x_i| > r$  and  $|x'_i| \leq \eta$  or  $|x_i| \leq \eta$  and  $|x'_i| > r$ . This

implies that  $|x'_i - x_i| > (r - \eta)$ . Thus we have

$$\|(x - x')|_{S_r(x)\Delta S_r(x')}\|_{\ell_p}^p \geq (\#(S_r(x)\Delta S_r(x'))) (r - \eta)^p.$$

Together with the inequality (23) and the non-negativity of  $\|(x - x')|_{S_r(x)\cap S_r(x')}\|_{\ell_p}$ , we obtain the chain of inequalities

$$\begin{aligned} (2\gamma_{2k}\kappa_p\eta)^p &\geq \|(x - x')|_{S_r(x)\cup S_r(x')}\|_{\ell_p}^p \\ &\geq \|(x - x')|_{S_r(x)\cap S_r(x')}\|_{\ell_p}^p + \|(x - x')|_{S_r(x)\Delta S_r(x')}\|_{\ell_p}^p \\ &\geq (\#(S_r(x)\Delta S_r(x'))) (r - \eta)^p, \end{aligned}$$

and therefore

$$\#(S_r(x)\Delta S_r(x')) \leq \frac{(2\gamma_{2k}\kappa_p\eta)^p}{(r - \eta)^p}. \quad (24)$$

For the unique support identification, we want the symmetric difference between the sets  $S_r(x)$  and  $S_r(x')$  to be empty. Thus the left-hand side of inequality (24) has to be zero. Since  $\#(S_r(x)\Delta S_r(x')) \in \mathbb{N}$ , it is sufficient to require that the right-hand side be strictly less than one, and this is equivalent to condition (22).  $\square$

**Remark 3.** *One additional implication of this latter theorem is that we can give a bound on the difference of  $x$  and  $x'$  restricted to the relevant entries. Indeed, in case of unique identification of the relevant entries, i.e.,  $\Lambda := S_r(x) = S_r(x')$  we obtain, by the inequality (23), that*

$$\|(x - x')_\Lambda\|_{\ell_1} \leq 2\gamma_k\kappa_p\eta. \quad (25)$$

Notice that we replaced  $\gamma_{2k}$  by  $\gamma_k \leq \gamma_{2k}$ , because now  $\#\Lambda \leq k$ .

**Remark 4.** *Unfortunately, we are not able to provide the necessity of the gap conditions (14), (19), (22) for successful support recovery, simply because we lack optimal deterministic error bounds in general: one way of producing a lower bound would be to construct a counterexample for each algorithm for which a certain gap condition is violated and recovery of support fails. Since most of the algorithms we shall illustrate below are iterative, it is likely extremely difficult to provide such explicit counterexamples. Therefore, we limit ourselves here to discuss the difference of  $r_1$  and  $r_S$  and of  $r_{1rew}$  and  $r_S$ . We shall see in the numerical experiments that the sufficient gap conditions (14), (19), (22) nevertheless provide actual indications of performance of the algorithms.*

- The gap between the two thresholds  $r_1, r_S$  is given by

$$r_1 - r_S = \left(2 \left(\frac{1 + \gamma_k}{1 - \gamma_k} - \gamma_{2k}\right) \kappa_p(N, k) - 1\right) \eta.$$

As  $\gamma_{2k} < 1 < \frac{1 + \gamma_k}{1 - \gamma_k}$  and  $\kappa_p(N, k)$  is very large for  $N \gg k$ , this positive gap is actually very large, for  $N \gg 1$ .

- The gap between the two thresholds  $r_{1rew}, r_S$  is given by

$$r_{1rew} - r_S = \left(9.6 \frac{\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}} \left(1 + \frac{\kappa_p}{\sqrt{k}}\right) - (1 + 2\gamma_{2k}\kappa_p)\right) \eta.$$

Following for example the arguments in [14], we know that the matrix  $A$  having the  $(2k, \delta_{2k})$ -RIP implies having the  $(2k, \gamma_{2k})$ -NSP with  $\gamma_{2k} = \frac{\sqrt{2}\delta_{2k}}{1 - (\sqrt{2}+1)\delta_{2k}}$ , which, substituted into the above equation, yields

$$r_{1rew} - r_S = \left( \frac{\left( \frac{9.6\sqrt{1+\delta_{2k}}}{\sqrt{k}} - 2\sqrt{2}\delta_{2k} \right) \kappa_p + [9.6\sqrt{1+\delta_{2k}} - (1 - (\sqrt{2}+1)\delta_{2k})]}{1 - (\sqrt{2}+1)\delta_{2k}} \right) \eta.$$

Since  $0 < \delta_{2k} < \sqrt{2} - 1$ , we have  $0 < 1 - (\sqrt{2}+1)\delta_{2k} < 1$ , and therefore the denominator and the right summand in the numerator are positive. The left summand of the numerator is positive and very large as soon as  $k < \left( \frac{9.6\sqrt{1+\delta_{2k}}}{2\sqrt{2}\delta_{2k}} \right)^2$ . Thus, even in the limiting scenario where  $\delta_{2k} \approx \sqrt{2} - 1$ , we still have  $k \leq 94$ , which may be considered sufficient for a wide range of applications. If one wants to exceed this threshold a more sophisticated estimate of the above term will reveal even less restrictive bounds on  $k$ . Thus, in general, since  $k$  and  $\delta_{2k}$  are small, also the left summand is positive. We conclude that again the gap is very large.

Unfortunately, this discrepancy cannot be amended because the  $\ell_1$ -minimization decoder  $\Delta_1$  and the iteratively re-weighted  $\ell_1$ -minimization decoder  $\Delta_{1rew}$  have not in general the property

$$x \in \mathcal{S}_{\eta, k, r}^p \Rightarrow \{\Delta_1(Ax), \Delta_{1rew}(Ax)\} \ni x^* \in \mathcal{S}_{\eta, k, r}^p.$$

Hence, it allows us neither to say that also the (iteratively re-weighted)  $\ell_1$ -minimizer has a bounded noise component

$$\sum_{i \in (S_r(x^*))^c} |x_i^*|^p \leq \eta^p,$$

nor to apply Theorem 3 to obtain support stability. We present several examples in Section 6, showing these ineliminable limitations of  $\Delta_1$  and  $\Delta_{1rew}$ .

### 3.4 The regularized selective $p$ -potential functional and its properties

To overcome the shortcomings of methods based exclusively on  $\ell_1$ -minimizations in 1. damping the noise-folding and consequently in 2. having a stable support recovery, in this section, we design a new decoding procedure which allows us to have both these very desirable properties.

Let us first introduce the following functional.

**Definition 5** (Regularized selective  $p$ -potential). *We define the regularized truncated  $p$ -power function  $W_r^{p, \epsilon}: \mathbb{R} \rightarrow \mathbb{R}_0^+$  by*

$$W_r^{p, \epsilon}(t) = \begin{cases} t^p & 0 \leq t < r - \epsilon, \\ \pi_p(t) & r - \epsilon \leq t \leq r + \epsilon, \\ r^p & t > r + \epsilon, \end{cases} \quad t \geq 0, \quad (26)$$

where  $0 < \epsilon < r$ , and  $\pi_p(t)$  is the third degree interpolating polynomial

$$\pi_p(t) := A(t - s_2)^3 + B(t - s_2)^2 + C,$$

with

$$\begin{cases} C = \mu_3, \\ B = \frac{\mu_1}{s_2 - s_1} - \frac{3(\mu_3 - \mu_2)}{(s_2 - s_1)^2}, \\ A = \frac{\mu_1}{3(s_2 - s_1)^2} + \frac{2B}{3(s_2 - s_1)}. \end{cases} \quad (27)$$

The parameters which appeared in the definition of the interpolating function are defined as:  $s_1 = (r - \varepsilon)$ ,  $s_2 = (r + \varepsilon)$ ,  $\mu_1 = p(r - \varepsilon)^{p-1}$ ,  $\mu_2 = (r - \varepsilon)^p$ , and  $\mu_3 = r^p$ . Moreover, we define it for  $t < 0$  as  $W_r^{p,\varepsilon}(t) = W_r^{p,\varepsilon}(-t)$ . We call the functional  $\mathcal{SP}_r^{p,\varepsilon}: \mathbb{R}^N \rightarrow \mathbb{R}_0^+$ ,

$$\mathcal{SP}_r^{p,\varepsilon}(x) = \sum_{j=1}^N W_r^{p,\varepsilon}(x_j), \quad r > 0, \quad 1 \leq p \leq 2, \quad (28)$$

the regularized selective  $p$ -potential (SP) functional.

To easily capture the function  $W_r^{p,\varepsilon}$ , we express it explicitly for  $p = 2$ . In this case, the polynomial has the analytic form

$$\pi_2(t) := \frac{[t + (r - \varepsilon)][\varepsilon(r + t) - (r - t)^2]}{4\varepsilon},$$

and the graph of  $W_r^{p,\varepsilon}$  is shown in Figure 2 for  $p = 2$ ,  $r = 1$ , and  $\varepsilon = 0.4$ . Notice that  $W_r^{p,0}$  is the truncated  $p$ -power function, which is widely used both in statistics and signal processing [3, 18], and also shown in Figure 2 for  $p = 2$ .

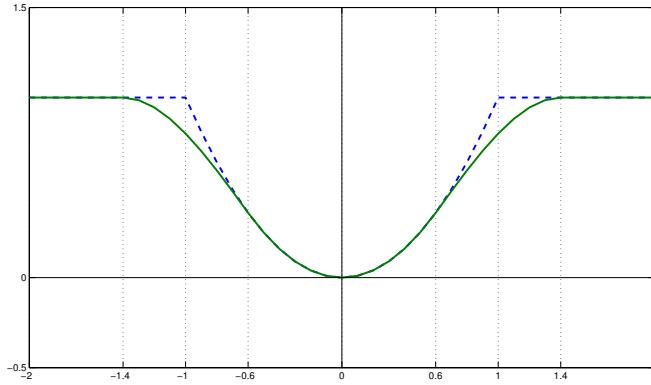


Figure 2: Truncated quadratic (2-power) potential  $W_1^{2,0}$  (dashed) and its regularization  $W_1^{2,0.4}$ .

**Remark 5.** Let us notice that the functional  $\mathcal{SP}_r^{p,\varepsilon}$  is semi-convex for  $\varepsilon > 0$ , that means that there exists a constant  $\omega > 0$  such that  $\mathcal{SP}_r^{p,\varepsilon}(\cdot) + \omega \|\cdot\|^2$  is convex.

**Theorem 4.** Let  $A \in \mathbb{R}^{m \times N}$  have the  $(2k, \gamma_{2k})$ -NSP, with  $\gamma_{2k} < 1$ , and  $1 \leq p \leq 2$ . Furthermore, we assume  $x \in \mathcal{S}_{\eta,k,r+\varepsilon}^p$ , for  $\varepsilon > 0$ ,  $0 < \eta < r + \varepsilon$ , with the property of having the minimal  $\#S_{r+\varepsilon}(x)$  within  $\mathcal{F}(y)$ , where  $y = Ax$  is its associated measurement vector, i.e.,

$$\#S_{r+\varepsilon}(x) \leq \#S_{r+\varepsilon}(z) \text{ for all } z \in \mathcal{F}(y). \quad (29)$$

If  $x^*$  is such that

$$\mathcal{SP}_r^{p,\varepsilon}(x^*) \leq \mathcal{SP}_r^{p,\varepsilon}(x), \quad (30)$$

and

$$|x_i^*| < r - \varepsilon, \quad (31)$$

for all  $i \in (S_{r+\varepsilon}(x^*))^c$ , then also  $x^* \in \mathcal{S}_{\eta,k,r+\varepsilon}^p$ , implying noise-folding damping. Moreover, we have

the support stability property

$$\#(S_{r+\epsilon}(x)\Delta S_{r+\epsilon}(x^*)) \leq \frac{(2\gamma 2k\kappa_p\eta)^p}{(r+\epsilon-\eta)^p}. \quad (32)$$

*Proof.* Notice that we can equally rewrite the  $\mathcal{SP}_r^{p,\epsilon}$  functional as

$$\mathcal{SP}_r^{p,\epsilon}(z) = r^p \#S_{r+\epsilon}(z) + \sum_{i \in (S_{r+\epsilon}(z))^c} |z_i|_\epsilon^p,$$

where  $|t|_\epsilon^p := W_r^{p,\epsilon}(t)$  for  $|t| \leq r+\epsilon$ . Here, by construction, we have  $|t|_\epsilon^p \leq |t|^p$ . By the assumptions (30) and  $x \in \mathcal{S}_{\eta,k,r+\epsilon}^p$ , we have

$$\begin{aligned} r^p \#S_{r+\epsilon}(x^*) &\leq \mathcal{SP}_r^{p,\epsilon}(x^*) \leq \mathcal{SP}_r^{p,\epsilon}(x) = r^p \#S_{r+\epsilon}(x) + \sum_{i \in (S_{r+\epsilon}(x))^c} |x_i|_\epsilon^p \\ &\leq r^p \#S_{r+\epsilon}(x) + \sum_{i \in (S_{r+\epsilon}(x))^c} |x_i|^p \leq r^p \#S_{r+\epsilon}(x) + \eta^p, \end{aligned} \quad (33)$$

and thus,

$$\#S_{r+\epsilon}(x^*) \leq \left(\frac{\eta}{r}\right)^p + \#S_{r+\epsilon}(x).$$

As  $\frac{\eta}{r} < 1$  by assumption, the minimality property (29) yields immediately

$$\#S_{r+\epsilon}(x^*) = \#S_{r+\epsilon}(x) \leq k. \quad (34)$$

Assumption (31) and again (30) yield

$$\begin{aligned} r^p \#S_{r+\epsilon}(x^*) + \sum_{i \in (S_{r+\epsilon}(x^*))^c} |x_i^*|^p &= r^p \#S_{r+\epsilon}(x^*) + \sum_{i \in (S_{r+\epsilon}(x^*))^c} |x_i^*|_\epsilon^p \\ &\leq r^p \#S_{r+\epsilon}(x) + \sum_{i \in (S_{r+\epsilon}(x))^c} |x_i|_\epsilon^p \\ &\leq r^p \#S_{r+\epsilon}(x) + \sum_{i \in (S_{r+\epsilon}(x))^c} |x_i|^p. \end{aligned}$$

By this latter inequality and (34) we obtain

$$\sum_{i \in (S_{r+\epsilon}(x^*))^c} |x_i^*|^p \leq \sum_{i \in (S_{r+\epsilon}(x))^c} |x_i|^p \leq \eta^p,$$

which implies  $x^* \in \mathcal{S}_{\eta,k,r+\epsilon}^p$ . By an application of Theorem 3, we obtain (32).  $\square$

**Remark 6.** *Let us comment on the assumptions of the latter result.*

- (i) *The assumption that  $x$  is actually the vector with minimal essential support  $S_r(x)$  among the feasible vectors in  $\mathcal{F}(y)$  corresponds to the request of being the "simplest" explanation to the data, in a certain sense we acknowledge the Occam's razor principle;*

(ii) The best candidate  $x^*$  to fulfill condition (30) would be actually

$$x^* := \arg \min_{z \in \mathcal{F}(y)} \mathcal{SP}_r^{p,\epsilon}(z) \quad (35)$$

because this will make (30) true, whichever  $x$  is. However (35) is a highly nonconvex problem whose solution is in general NP-hard [1]. The way we will circumvent this drawback is to employ an algorithm, which we describe in details in the following section, to compute  $x^*$  by performing a local minimization of  $\mathcal{SP}_r^{p,\epsilon}$  in  $\mathcal{F}(y)$  around a given starting vector  $x_0$ . Ideally, the best choice for  $x_0$  would be  $x$  itself, so that (30) may be fulfilled. But, obviously, we do not dispose yet of the original vector  $x$ . Therefore, a heuristic rule, which we will show to be very robust in our numerical simulations, is to choose  $x_0 = \Delta_1(y) \approx x$ , i.e., we use  $x_0$  as the result of the  $\ell_1$ -minimization as a warm-up for the iterative algorithm described below. The reasonable hope is that actually

$$\mathcal{SP}_r^{p,\epsilon}(x^*) \leq \mathcal{SP}_r^{p,\epsilon}(\Delta_1(y)) \approx \mathcal{SP}_r^{p,\epsilon}(x);$$

(iii) The assumption that the outcome  $x^*$  of the algorithm has additionally the property  $|x_i^*| < r - \epsilon$ , for all  $i \in (S_{r+\epsilon}(x^*))^c$  is justified by observing that in our implementation  $x^*$  will be the result of a thresholding operation, i.e.,  $x_i^* = \mathbb{S}_p^\mu(\xi_i)$ , for  $i \in (S_{r+\epsilon}(x^*))^c$ , see (38). The particularly steep shape of the thresholding function in the interval  $[r - \epsilon, r + \epsilon]$ , especially for  $p = 2$ , see Figure 3(c), makes it highly unlikely for  $\epsilon$  sufficiently small that  $r - \epsilon \leq |x_i^*|$  for  $i \in (S_{r+\epsilon}(x^*))^c$ . Actually, our numerical experiments confirm that typically this algorithm promotes solutions in  $\mathcal{F}(y)$  selectively characterized by a relevant gap between large components exceeding  $r$  and small components, significantly below  $r$ .

## 4 Minimization of the regularized selective $p$ -potential functional

In the latter section, we introduced the functional  $\mathcal{SP}_r^{p,\epsilon}$ , which is nonconvex. Unluckily, this makes its linearly constrained minimization (35), which we want to call selective least  $p$ -powers (SLP), also nontrivial. Here, we recall a novel and very robust algorithm for linearly constrained nonconvex and nonsmooth minimization, introduced and analyzed first in [3]. The algorithm is particularly suited for our purpose, since it only requires a  $C^1$ -regular functional. This distinguishes it from other well-known methods such as SQP and Newton methods, which require a more restrictive  $C^2$ -regularity. All notions and results written in this section are collected more in general in [3]. Nevertheless we report them directly adapted to our specific case in order to have a simplified and more immediate application.

### 4.1 The algorithm

In this section, we present the algorithm to perform the local minimization of (28) in  $\mathcal{F}(y)$ . Before describing it, it is necessary to introduce the concept of  $\nu$ -convexity, which plays a key-role in the minimization process. In fact, to achieve the minimization of the functional  $\mathcal{SP}_r^{p,\epsilon}$ , we use a Bregman-like inner loop, which requires this property to converge with an a priori rate.

**Definition 6** ( $\nu$ -convexity). A function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  is  $\nu$ -convex if there exists a constant  $\nu > 0$  such that for all  $x, x' \in \mathbb{R}^N$  and  $\phi \in \partial f(x), \psi \in \partial f(x')$

$$\langle \phi - \psi, x - x' \rangle \geq \nu \|x - x'\|_{\ell_2}^2, \quad (36)$$

where  $\partial f$  is the subdifferential of the function  $f$ .



The starting values  $x_0 = x_{(0,0)} \in \mathbb{R}^N$  and  $q_{(0,0)} \in \mathbb{R}^m$  are taken arbitrarily. For a fixed scaling parameter  $\lambda > 0$  and an adaptively chosen sequence of integers  $(L_\ell)_{\ell \in \mathbb{N}}$ , we formulate

---

**Algorithm 1** SLP

---

**while**  $\|x_{\ell-1} - x_\ell\|_{\ell_2} \leq TOL$  **do**  
 $x_{(\ell,0)} = x_{\ell-1} := x_{(\ell-1, L_{\ell-1})}$   
 $q_{(\ell,0)} = q_{\ell-1} := q_{(\ell-1, L_{\ell-1})}$   
**for**  $k = 1, \dots, L_\ell$  **do**  
 $x_{(\ell,k)} = \arg \min_{x \in \mathbb{R}^N} (\mathcal{SP}_{\omega, x_{\ell-1}}^{p,\epsilon}(x) - \langle q_{(\ell,k-1)}, Ax \rangle + \lambda \|Ax - y\|_{\ell_2}^2)$   
 $q_{(\ell,k)} = q_{(\ell,k-1)} + 2\lambda(y - Ax_{(\ell,k)})$   
**end for**  
**end while**

---

The reader can notice that the functional  $\mathcal{SP}_{\omega, x_{\ell-1}}^{p,\epsilon}$ , which appears in the algorithm, has not been yet introduced. Indeed, a modification to the functional  $\mathcal{SP}_r^{p,\epsilon}$  must be introduced in order to have  $\nu$ -convexity, which is necessary for the convergence of the algorithm. It is defined as

$$\mathcal{SP}_{\omega, x'}^{p,\epsilon}(x) := \mathcal{SP}_r^{p,\epsilon}(x) + \omega \|x - x'\|_{\ell_2}^2,$$

where  $\omega$  is chosen such that the new functional is  $\nu$ -convex. The finite adaptively chosen number of inner loop iterates  $L_\ell$  is defined by the condition

$$(1 + \|q_{\ell-1}\|_{\ell_2}) \|Ax_{(\ell, L_\ell)} - y\|_{\ell_2} \leq \frac{1}{\ell^\alpha},$$

for a given parameter  $\alpha > 1$ , which in our numerical experiments will be set to  $\alpha = 1.1$ . We refer to [3, Section 2.2] for details on the finiteness of  $L_\ell$  and for the proof of convergence of Algorithm 1 to critical points of  $\mathcal{SP}_r^{p,\epsilon}$  in  $\mathcal{F}(y)$ . According to Remark 6 (ii), and as we will empirically verify in our numerical experiments reported in Section 6, this algorithm finds critical points (hopefully close to a global minimizer) with the desired properties illustrated in Theorem 4, as soon as we select the starting point  $x_0$  by an appropriate warm-up procedure.

Algorithm 1 does not yet specify how to minimize the convex functional

$$(\mathcal{SP}_{\omega, x_{\ell-1}}^{p,\epsilon}(x) - \langle q_{(\ell,k-1)}, Ax \rangle + \lambda \|Ax - y\|_{\ell_2}^2),$$

in the inner loop. For that we can use an iterative thresholding algorithm introduced in [3, Section 3.7], inspired by the previous work [18] for the corresponding *unconstrained* optimization of regularized selective  $p$ -potentials. This method ensures the convergence to a minimizer and is extremely agile to be implemented, as it is based on matrix-vector multiplications and very simple componentwise nonlinear thresholdings.

By the iterative thresholding algorithm, we actually equivalently minimize the functional

$$\mathcal{SP}_{\omega, x'}^{p,\epsilon}(x, q) = \mathcal{SP}_{\omega, x'}^{p,\epsilon}(x) + \lambda \|Ax - (y + q)\|_{\ell_2}^2,$$

where we set  $\lambda = \frac{1}{2}$  only for simplicity. The thresholding functions  $\mathbb{S}_p^\mu$  we use are defined in [3, Lemma

3.15] and, for the relevant case  $p = 2$ , it has the analytic form

$$\mathbb{S}_2^\mu(\xi) := \begin{cases} \frac{\xi}{1 + \mu} & |\xi| < (\bar{r} - \epsilon)(1 + \mu), \\ \frac{4\epsilon}{3\mu} \left( 1 + \frac{\mu}{4\epsilon}(2\epsilon + \bar{r}) - \sqrt{\frac{\Gamma(\xi)}{4}} \right) & (\bar{r} - \epsilon)(1 + \mu) \leq |\xi| \leq r + \epsilon, \\ \xi & |\xi| > \bar{r} + \epsilon, \end{cases} \quad (37)$$

where

$$\Gamma(\xi) := 4 \left( 1 + \left( \frac{\mu}{4\epsilon} \right)^2 (2\bar{r} + \epsilon)^2 + \frac{\mu}{2\epsilon}(2\epsilon + \bar{r}) - \frac{3\mu}{2\epsilon}\xi \right).$$

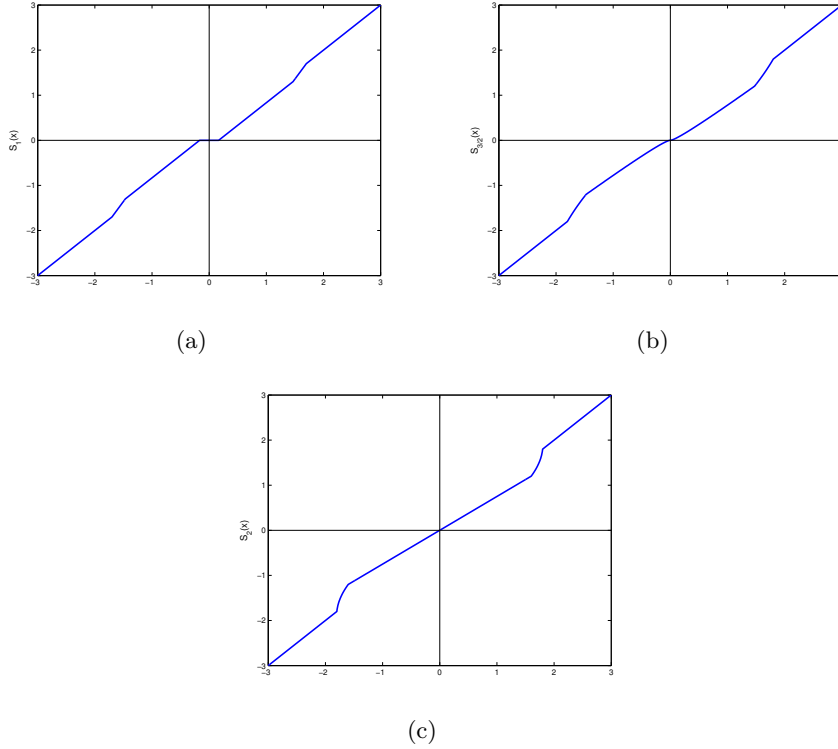


Figure 3: The Lipschitz continuous thresholding functions  $\mathbb{S}_1^\mu$ ,  $\mathbb{S}_{3/2}^\mu$ , and  $\mathbb{S}_2^\mu$ , with parameters  $p = 1, 3/2$ , and  $2$ , respectively, and  $r = 1.5$ ,  $\mu = 5$ ,  $\epsilon = 0.3$ .

Note that in case  $p \neq 2$  the only part that varies is the one for  $|\xi| < (\bar{r} - \epsilon)(1 + \mu)$  because the remaining ones do not depend on  $p$ . We show in Figure 3 the typical shapes of these thresholding functions for different choices of  $p \in \{1, 3/2, 2\}$ . By means of these thresholding functions, the minimizing algorithm in the inner loop is given by the componentwise fixed point iteration, for  $n \geq 0$ ,

$$x_i^{n+1} = \mathbb{S}_p^\mu \left( \left\{ \frac{1}{2} \left[ \left( I - \frac{1}{2} A^* A \right) + (1 - \omega) I \right] x^n + \frac{1}{2} A^* (y + q) + \omega x' \right\}_i \right), \quad i = 1, \dots, N. \quad (38)$$

We refer to [3, Theorem 3.17] for the convergence properties of this algorithm.

To summarize, Algorithm 1 can be realized in practice by nesting three loops. One external loop makes slowly vanishing the quadratic convexification, the second external loop updates the Lagrange multipliers  $q_{(\ell,k)}$  for a fixed quadratic convexification, and the final inner loop implements (38).

## 5 Selectivity principle

In Section 3, we showed a certain superiority of the regularized selective least  $p$ -powers (SLP) in terms of support identification and damping noise-folding. These theoretical results are supported further below in Section 6 by extensive numerical results. However, the realization of the algorithm SLP described above turns out to be computationally demanding as soon as the dimension  $N$  gets large. Since the computational time is a crucial point when it comes to practical applications, we shall introduce in this section another and more efficient approach based on the same principles to obtain equally good support identification results. The investigations in Section 3 reveal some weakness of the sparsity concept:

The decoders based on  $\ell_1$ -minimization and iteratively re-weighted  $\ell_1$ -minimization prefer sparse solutions and have the undesirable effect of sparsifying also the noise. Thus all the noise may concentrate on fewer entries which, by a balancing principle, may eventually exceed in absolute value the smallest entries of the actual original signal. This makes it impossible to separate the relevant entries of the signal from the noise by only knowing the threshold  $r$  which bounds the relevant entries from below. On the contrary, SLP follows a *selectivity* principle, where the recovery process focuses on the extraction of the relevant entries, while uniformly distributing the noise elsewhere. We want now to export this latter mechanism to formulate the method described below.

In the following, we will show that the very well known iterative hard thresholding [7] shows similar support identification properties as SLP while being very efficient in terms of computational time. This method iteratively computes

$$x^{n+1} := \mathbb{H}_{\sqrt{\tau}}(x^n + A^T(y - Ax^n)), \quad (39)$$

where

$$(\mathbb{H}_{\sqrt{\tau}}(z))_i := \begin{cases} z & \text{if } |z| > \sqrt{\tau}, \\ 0 & \text{else,} \end{cases}$$

for  $\tau \geq 0$ . It is converging to a fixed point  $x^{\text{IHT}}$  fulfilling

$$x^{\text{IHT}} := \mathbb{H}_{\sqrt{\tau}}(x^{\text{IHT}} + A^T(y - Ax^{\text{IHT}})), \quad (40)$$

which is a local minimizer of the functional

$$\mathcal{J}_0(x) := \|Ax - y\|_{\ell_2}^2 + \tau \# \text{supp}(x),$$

see [7, Theorem 3] for details. One immediately observes that iterative hard thresholding has a very selective nature: The computed fixed point  $x^{\text{IHT}}$  is a vector with an unknown number, although expected to be small, of non-zero entries whose absolute value is above the threshold  $\sqrt{\tau}$ . In the following, we show under which conditions this method is able to exactly identify the support of the relevant entries of the original vector  $x$ .

**Theorem 5.** Assume  $A \in \mathbb{R}^{m \times N}$  to have the  $(2k, \delta_{2k})$ -RIP, with  $\delta_{2k} < 1$ ,  $\|A\| \leq 1$ , and define  $\beta(A) > 0$  such that

$$\sup_{i \in \{1, \dots, N\}} |A_i^T z| \geq \beta(A) \|z\|_{\ell_2}, \quad \text{for all } z \in \mathbb{R}^m,$$

where  $A_i$  are the columns of the matrix  $A$ . Let  $x \in \mathcal{S}_{\eta, k, r}^p$ , for a fixed  $1 \leq p \leq 2$ , and  $y = Ax$  the respective measurements. Assume further

$$r > \eta \left( 1 + \frac{1}{1 - \delta_{2k}} \left( 1 + \frac{1}{\beta(A)} \right) \right), \quad (41)$$

and define  $\tau$  such that

$$\eta < \sqrt{\tau} < \frac{r - \frac{\eta}{1 - \delta_{2k}}}{1 + \frac{1}{(1 - \delta_{2k})\beta(A)}}. \quad (42)$$

Let  $x^{\text{IHT}}$  be the limit of the iterative hard thresholding algorithm (39), and thus fulfilling the fixed point equation (40). Further, we assume

$$\mathcal{J}_0(x^{\text{IHT}}) \leq \mathcal{J}_0(x|_{S_r(x)}), \quad (43)$$

then we have that  $\Lambda := S_r(x) = \text{supp}(x^{\text{IHT}})$ , and

$$|x_i - x_i^{\text{IHT}}| < r - \sqrt{\tau}, \quad \text{for all } i \in \Lambda. \quad (44)$$

*Proof.* Assume  $\#\text{supp}(x^{\text{IHT}}) > \#S_r(x) = k$ . By (43), we have that

$$\begin{aligned} 0 &< \#\text{supp}(x^{\text{IHT}}) - \#\text{supp}(x|_{S_r(x)}) \\ &= \#\text{supp}(x^{\text{IHT}}) - \#S_r(x) \\ &\leq \frac{1}{\tau} \left( \|A(x|_{S_r(x)}) - y\|_{\ell_2}^2 - \|Ax^{\text{IHT}} - y\|_{\ell_2}^2 \right) \leq \frac{1}{\tau} \|A(x|_{S_r(x)}) - y\|_{\ell_2}^2 \\ &= \frac{1}{\tau} \|A(x|_{(S_r(x))^c})\|_{\ell_2}^2 \leq \frac{1}{\tau} \|A\|^2 \|x|_{(S_r(x))^c}\|_{\ell_2}^2 \leq \frac{\eta^2}{\tau} \\ &< 1, \end{aligned}$$

where the last inequality follows by (42). Since  $(\#\text{supp}(x|_{S_r(x)}) - \#\text{supp}(x^{\text{IHT}})) \in \mathbb{N}$ , the upper inequality yields to a contradiction.

Thus  $\#\text{supp}(x^{\text{IHT}}) \leq \#S_r(x) = k$  and therefore  $x^{\text{IHT}}$  and  $x|_{S_r(x)}$  are both  $k$ -sparse, and  $(x^{\text{IHT}} - x|_{S_r(x)})$  is  $2k$ -sparse. Under our assumptions we can apply [7, Lemma 4] to obtain

$$\|Ax^{\text{IHT}} - y\|_{\ell_2} \leq \frac{\sqrt{\tau}}{\beta(A)}. \quad (45)$$

In addition to this latter estimate, we use the RIP, the sparsity of  $x^{\text{IHT}} - x|_{S_r(x)}$ , and (42) to obtain

for all  $i \in \{1, \dots, N\}$  that

$$\begin{aligned}
|(x|_{S_r(x)})_i - x_i^{\text{IHT}}| &\leq \|x|_{S_r(x)} - x^{\text{IHT}}\|_{\ell_2} \leq \frac{\|A(x|_{S_r(x)} - x^{\text{IHT}})\|_{\ell_2}}{1 - \delta_{2k}} \\
&\leq \frac{\|A(x - x^{\text{IHT}})\|_{\ell_2} + \|A(x|_{(S_r(x))^c})\|_{\ell_2}}{1 - \delta_{2k}} \\
&\leq \frac{\|y - Ax^{\text{IHT}}\|_{\ell_2} + \|A(x|_{(S_r(x))^c})\|_{\ell_2}}{1 - \delta_{2k}} \\
&\leq \frac{\sqrt{\tau}}{\beta(A)(1 - \delta_{2k})} + \frac{\eta}{1 - \delta_{2k}} < r - \sqrt{\tau}.
\end{aligned}$$

Assume now that there is  $\tilde{i} \in \mathbb{N}$  such that  $\tilde{i} \in S_r(x)$  and  $\tilde{i} \notin \text{supp}(x^{\text{IHT}})$ . But then we would also have  $|x_{\tilde{i}} - x_{\tilde{i}}^{\text{IHT}}| = |x_{\tilde{i}}| > r$ , which leads to a contradiction. Thus, we have equivalently  $S_r(x) \subset \text{supp}(x^{\text{IHT}})$ . In addition to  $\#\text{supp}(x^{\text{IHT}}) \leq \#S_r(x)$ , this concludes the proof.  $\square$

**Remark 7.** *Let us discuss some of the assumptions and implications of this latter result.*

- (i) *Since iterative hard thresholding is only computing a local minimizer of  $\mathcal{J}_0$ , condition (43) may not be always fulfilled for any given initial iteration  $x^0$ . Similarly to the argument in Remark 6 (ii), using the  $\ell_1$ -minimizer as the starting point  $x^0$ , or equivalently choosing the vector  $x^0$  as composed of the entries of  $\Delta_1(Ax)$  exceeding  $\sqrt{\tau}$  in absolute value, we may allow us to approach a local minimizer which fulfills (43).*
- (ii) *Condition (41) is comparable to the one derived in (22). If  $A$  is “well-conditioned”, i.e., we have that  $(1 - \delta_{2k}) \sim 1$ , and  $\beta(A) \sim 1$ , then*

$$1 + \frac{1}{1 - \delta_{2k}} \left(1 + \frac{1}{\beta(A)}\right) \sim 3.$$

Although we are able to exactly identify the support  $\Lambda$  by means of Theorem 5, we only have the very poor error estimate (44) of the relevant part of the signal. The reason why we cannot obtain an estimate as good as the one in Remark 3 is that the conditions  $x^{\text{IHT}} \in \mathcal{S}_{\eta, k, r}^p$ , and  $Ax = Ax^{\text{IHT}}$  to apply Theorem 3 are in general not fulfilled. Hence an additional correction to  $x^{\text{IHT}}$  is necessary. As we now dispose of the support  $\Lambda = S_r(x)$ , a natural approach is to seek for an additional vector  $x'$  which is the solution of

$$\begin{aligned}
\min_{z \in \mathbb{R}^n} \quad &\|Az - y\|_{\ell_2}^2 \\
\text{s.t.} \quad &\|z_{\Lambda^c}\|_{\ell_p} \leq \eta, \\
&|z_i| \geq r, \text{ for all } i \in \Lambda.
\end{aligned} \tag{46}$$

Since the original signal  $x$  fulfills  $Ax - y = 0$ , and  $x \in \mathcal{S}_{\eta, k, r}^p$ , it is actually a solution of problem (46). Thus, we conclude that for any minimizer  $x'$  of problem (46) the objective function equals zero, thus  $\|Ax' - y\|_{\ell_2} = 0$ , and we conclude  $Ax = Ax'$  and  $x' \in \mathcal{S}_{\eta, k, r}^p$  as well. The optimization (46) is in general nonconvex, but, luckily, we can easily recast it in an equivalent convex one: Since  $|x_i - x_i^{\text{IHT}}| < r - \sqrt{\lambda}$ , and  $|x_i| > r$ , we know that the relevant entries of  $x$  and  $x^{\text{IHT}}$  have the same sign. Since we are searching for solutions which are close to  $x$ , the second inequality constraint becomes  $\text{sign}(x_i^{\text{IHT}})z_i \geq r$ , for all

$i \in \Lambda$ . Together with the equivalence of  $\ell_2$ - and  $\ell_p$ -norm, we rewrite problem (46) as

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & \frac{1}{2} z^T (A^T A) z - y^T A z \\ \text{s.t.} \quad & z^T P_0 z - (N - k)^{1 - \frac{2}{p}} \eta^2 \leq 0, \\ & z^T P_j z - (\text{sign}(x_{i_j}^{\text{IHT}}) e_{i_j})^T z + r \leq 0, \text{ for all } i_j \in \Lambda, j = 1, \dots, \#\Lambda, \end{aligned} \quad (47)$$

where  $P_0 \in \mathbb{R}^{N \times N}$  is defined componentwise by

$$(P_0)_{r,s} := \begin{cases} 1 & \text{if } r = s \in \Lambda \\ 0 & \text{else} \end{cases},$$

and  $P_j = 0$ ,  $j = 1, \dots, \#\Lambda$ . Since  $A^T A$ ,  $P_0$ , and  $P_j$ ,  $j = 1, \dots, \#\Lambda$ , are semi-definite, problem (47) is a *convex quadratically constrained quadratic program* (QCQP) which can be efficiently solved by standard methods, e.g., interior point methods [28]. Since we combine here three very efficient methods ( $\ell_1$ -minimization, iterative hard thresholding, and a QCQP), the resulting procedure is much faster than the computation of SLP while, as we will show in the numerics, keeping similar support identification properties.

## 6 Numerical results

The following numerical simulations provide empirical confirmation of the theoretical observations in Section 3 and Section 5. In particular, we want to show that SLP and IHT, initialized by the  $\ell_1$ -minimizer as a starting value, are very robust and provide a significantly enhanced rate of recovery of the support of the unknown sparse vector as well as a better accuracy in approximating its large entries, whenever limiting noise, i.e.,  $\eta \approx r$ , is present on the signal.

In the previous sections we provided a very detailed description of the parameter choice for the concrete implementation of SLP and IHT, which depends in a rather explicit way on the threshold parameter  $r > 0$  and it is independent of the dimensionalities  $N, m, k$  of the problem. Instead, it turns out to be extremely hard to tune the parameter  $\delta$ , appearing in [10, 27] as the  $\ell_2$ -norm residual discrepancy, in order to obtain the best performances for the iteratively re-weighted  $\ell_1$ -minimization (IRW $\ell_1$ ) in terms of support identification and accuracy in approximating the large entries of the original vector. In the papers [10, 27] the authors indicated  $\delta^2 = \sigma^2(m + 2\sqrt{2m})$ , depending on  $m$ , as the best parameter choice for ameliorating the discrepancy in  $\ell_2$ -norm between original and decoded vector with respect to the sole  $\ell_1$ -minimization. However, in our experiments we found out that for the two purposes mentioned above a much smaller  $\delta$  has to be chosen. The stability parameter  $a$ , which avoids the denominator to be zero in the weight updating rule seems to have instead no strong influence, and it is set to 0.1 in our experiments. We executed 8 iterations of IRW $\ell_1$  as a reasonable compromise between computational effort and accuracy.

We also consider as one of the test methods  $\ell_1$ -minimization, where we substituted the equality constraint  $Ax = y$  with an inequality constraint which takes into account the noise level, folded from the noise on the signal though; thus

$$\|Ax - y\|_{\ell_2} \leq \|An\|_{\ell_2} \leq \delta.$$

In this constraint, we used the same parameter  $\delta$  as for IRW $\ell_1$ .

As we shall argue in detail below, the following numerical tests indicate that  $\ell_1$ +IHT is much

faster and usually more robust than  $\ell_1$ +SLP, and that both of them perform much better than  $\ell_1$ -minimization and IRW $\ell_1$  in terms of support recovery and accuracy in approximating the large entries in absolute value of the original signal.

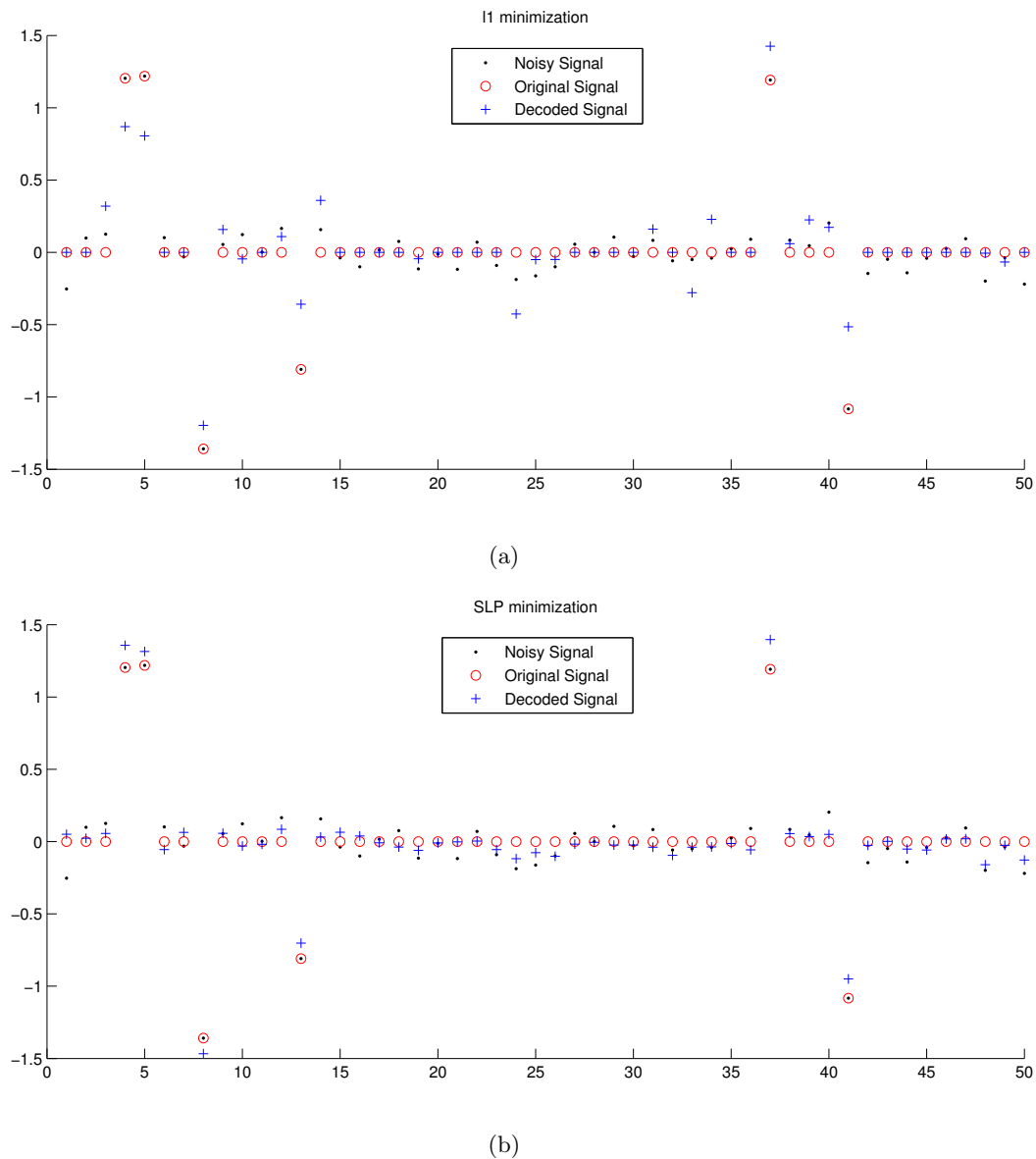


Figure 4: The results of the  $\ell_1$ -minimization and SLP are shown in Subfigure 4(a) and 4(b) respectively. The two decoders are intended to recover the original signal ( $\circ$ ), starting from the measurement of the noisy signal ( $\cdot$ ). The output of the two processes is represented by ( $+$ ). In this case the starting value for both minimizations is  $x_0 = 0$ .

**Advantages of SLP with respect to  $\ell_1$ -minimization** We shall start the discussion on numerical experiments with a comparison between the  $\ell_1$ -minimization and SLP for one typical example reported in Figure 4. Although the setting of the two methods is the same, the results are very different: SLP-

minimization can recover the signal with a very good approximation of its large entries in absolute value and a significant reduction of the noise level, while  $\ell_1$ -minimization may approximate the signal in a bad way, due to the amplification of the noise. For instance, it is evident in Subfigure 4(a) that  $|x_{\ell_1}(13)| < |x_{\ell_1}(24)|$  gives a wrong information about the location of the relevant entries, mismatching them with the noise. However, in this particular example, we were lucky to choose the right starting value for SLP. Due to its nonconvex character, in general SLP is computing a local minimizer, which might be far away from the original signal.

**Choosing  $\ell_1$ -minimization as a warm up** In Section 4, we mentioned that the algorithm, which minimizes the functional  $\mathcal{SP}_r^{p,\epsilon}$ , finds only a critical point, so the condition  $\mathcal{SP}_r^{2,\epsilon}(x^*) \leq \mathcal{SP}_r^{2,\epsilon}(x)$  (30) used in the proof of Theorem 4 may not be always valid. In order to enhance the chance of validity of this condition, the choice of an appropriate starting point is crucial. As we know that the  $\ell_1$ -minimization converges to its global minimizer with at least some guarantees given by Theorem 1, we use the result of this minimization process as a warm up to select the starting point of Algorithm 1. In the following, we distinguish between SLP which starts at  $x_0 = 0$  and  $\ell_1$ +SLP which starts at the  $\ell_1$ -minimizer.

In Figure 5, we illustrate the robustness of  $\ell_1$ +SLP (bottom left subfigure) in comparison to the  $\ell_1$ -minimization based methods and SLP starting at 0. Here, SLP converged to a feasible critical point, but it is quite evident that the decoding process failed since the large entry at position 83 (signal) was badly recovered and even the entry at position 89 (noise) is larger. If we look at the  $\ell_1$ -minimization result (top left subfigure) or the  $\ell_1$ -minimization with inequality constraint (top right subfigure), the minimization process brings us close to the solution, but the results still significantly lack accuracy. By  $\ell_1$ +SLP (center left subfigure) we obtain a good approximation of the relevant entries of the original signal and we get a significant correction and an improved recovery. Also IRW $\ell_1$  improves the result of  $\ell_1$ -minimization significantly, but still approximates the large entries worse than  $\ell_1$ +SLP. Although the difference is minor, we observe another important aspect of IRW $\ell_1$ : the noise part is sparsely recovered, while  $\ell_1$ +SLP distributes the noise in a more uniform way in a much smaller stripe around zero. This drawback of IRW $\ell_1$  can be crucial when it comes to the distinction of the relevant entries from noise.

**Massive computations** The previously presented specific examples in support of our new decoding strategies are actually typical. In order to support this work with even more impressive and convincing evidences, we present some statistical data obtained by solving series of problems. We decided to fix the parameters in order to have the most coherent data to be analyzed; in particular, we set  $N = 100$ ,  $m = 40$ ,  $r = 0.8$ ,  $k = 1, \dots, 7$ , and  $\eta = 0.75$ . The vector  $n$  is composed of random entries with normal distribution and then it is rescaled in order to have  $\|n\|_{\ell_2} = \eta$ . Figures 6, 7, and 8 report the results obtained considering 30 different i.i.d. Gaussian encoding matrices while Figures 9, 10, and 11 used 30 random subsampled cosine transformation encoding matrices. In the following, we use  $x^*$  generically for the decoded vector of any method.

Figure 6 reports the first part of the statistics which we have collected for the Gaussian matrices. We start commenting the subfigures clockwise. The first subfigure, on the upper-left, represents the mean value of the error between the exact signal and the decoded one  $\|x - x^*\|_{\ell_2}$ . For  $\ell_1$ +SLP,  $\ell_1$ +IHT, and IRW $\ell_1$  the absolute  $\ell_2$ -norm discrepancy between original and decoded vector seems to be stable and independent of the number  $k$  of large entries. These methods outperform  $\ell_1$ -minimization; IRW $\ell_1$  performs slightly better. We also observe that the choice of the starting point is crucial for SLP and IHT.



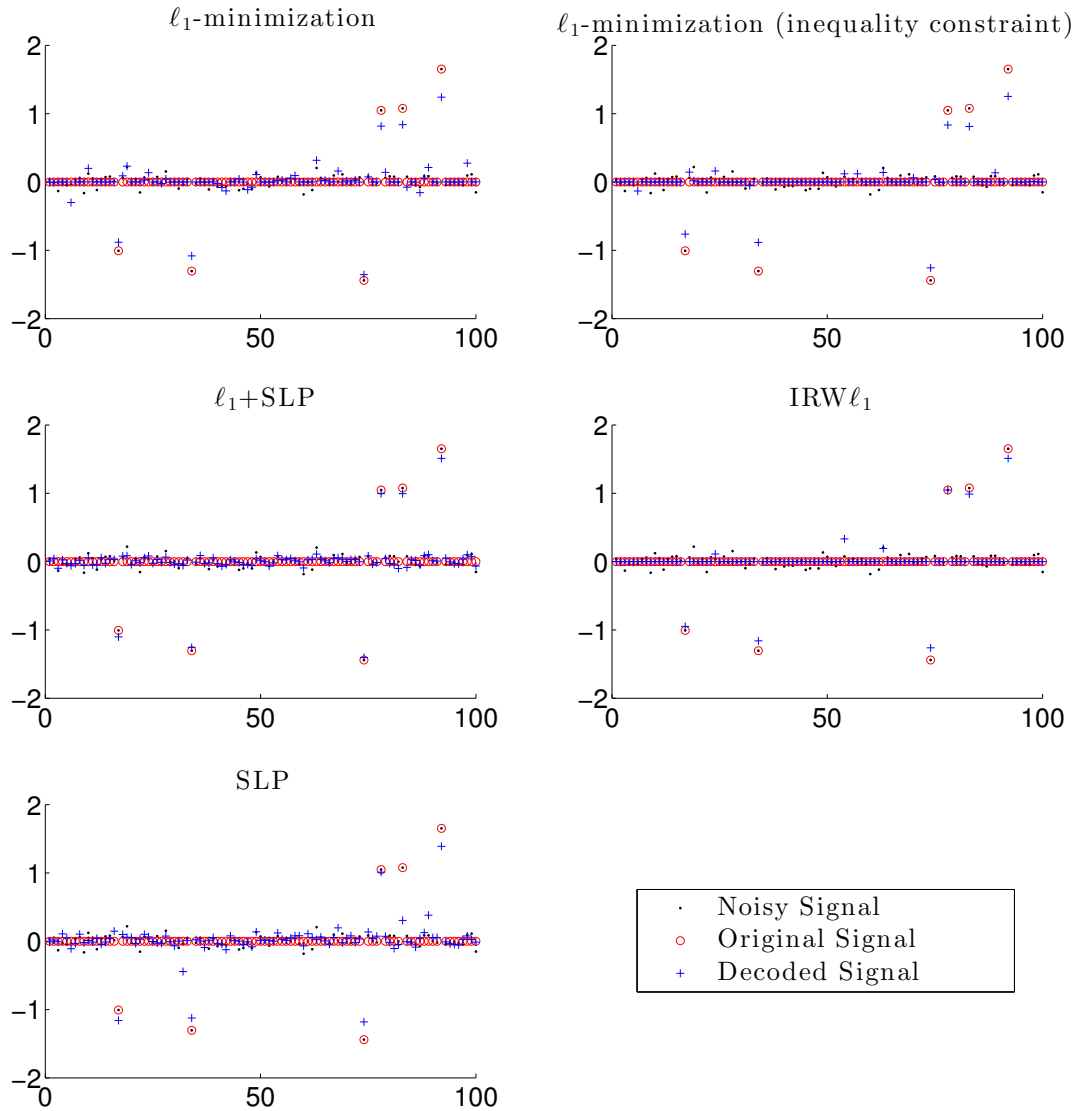


Figure 5: The figure reports the results of five different decoding processes (+) of the same problem where the circles (○) represent the original signal and the points (·) represent the original signal corrupted by the noise.

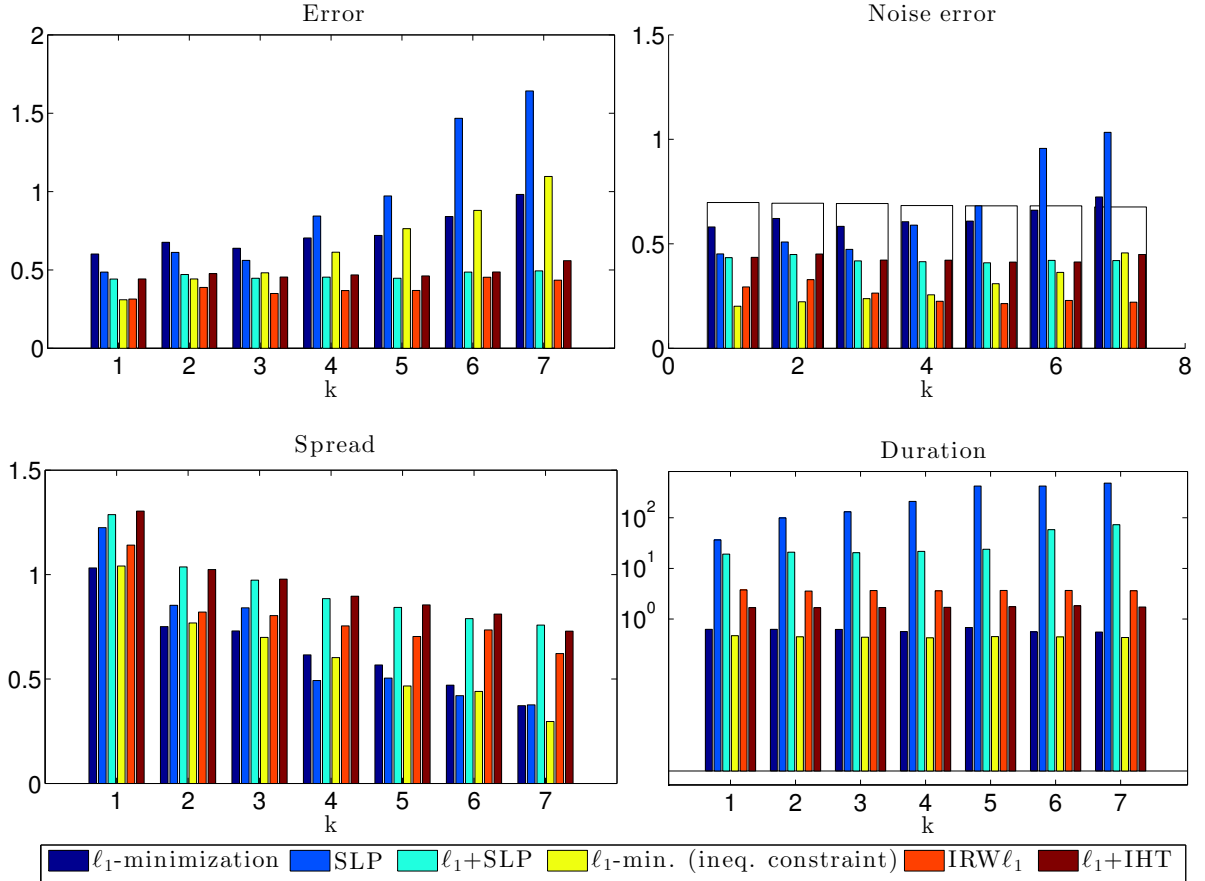


Figure 6: The columns refer to the different results of  $\ell_1$ -minimization (dark blue), SLP (blue),  $\ell_1$ +SLP (cyan),  $\ell_1$ -minimization with inequality constraint (yellow), IRW $\ell_1$  (orange), and  $\ell_1$ +IHT (brown). In the *Noise error* subfigure the white column in the background represents the noise level. On the x-axis the different values of  $k$  are displayed and each column is the mean of the results given by 30 trials. The results were obtained by Gaussian matrices.

The second subfigure is the mean value of the noise level and we can see exactly what we inferred looking at Figure 5:  $\ell_1$ -minimization returns a larger noise level with respect to all the other methods, except SLP; and IRW $\ell_1$  has the best noise reduction property.

The third is the mean computational time, presented in logarithmic scale. All tests were implemented and run in Matlab R2013b in combination with CVX [13, 20], to solve the  $\ell_1$ -minimization with equality and inequality constraint, its iteratively re-weighted version, and the QCQP. We observe that SLP and  $\ell_1$ +SLP are extremely slow. However, in comparison, the good starting point for SLP provides an advantage in terms of computational time. IHT has a computational complexity in between  $\ell_1$ -minimization and IRW $\ell_1$ .

The fourth plot reports the mean value of the discrepancy between noise level and the large entries of the signal, thus  $\min_{i \in S_r(x)} |x_i^*| - \max_{i \in S_r(x)^c} |x_i^*|$ . This plot shows how good the small entries are distinguished from the large ones in absolute value. We see that  $\ell_1$ +SLP and  $\ell_1$ +IHT perform best, which again is a result of their non-sparse noise recovery.

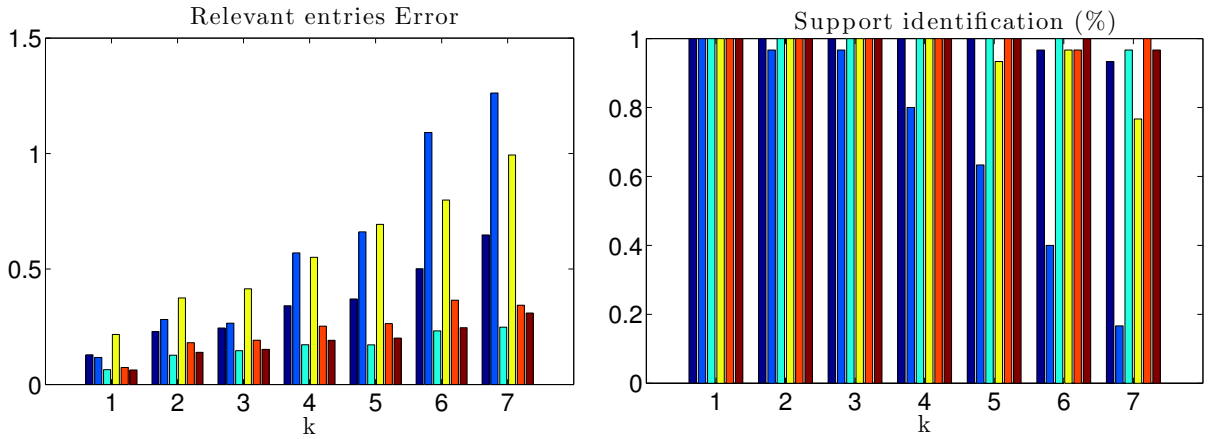


Figure 7: The subfigures represent the error on the relevant entries and the support identification property by knowledge of  $k$ . For more details on the displayed data we refer to the caption of Figure 6. The results were obtained by Gaussian matrices.

In Figure 7, we report the histogram of the mean-value of the errors on the relevant entries: the quantities on the left subfigure are computed as the mean values of  $\|x_{S_r(x)} - x_{S_r(x)}^*\|_{\ell_2}$  where we suppose to know the  $k$  largest entries of the original signal. The right subfigure shows how often the  $k$  largest entries of  $x^*$  coincided with  $S_r(x)$ . Notice that there might be entries below the threshold  $r$  among the  $k$  largest entries of  $x^*$ . We conclude that, knowing the number of large entries, IRW $\ell_1$ ,  $\ell_1$ -minimization,  $\ell_1$ +SLP, and  $\ell_1$ +IHT recover the support with nearly 100% success. In addition,  $\ell_1$ +SLP approximates best the magnitudes of the relevant entries.

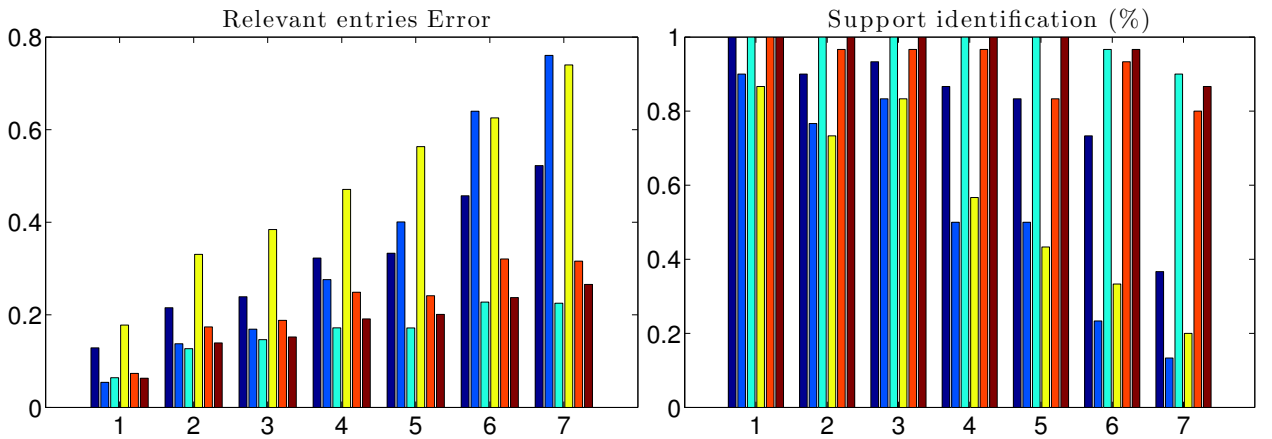


Figure 8: The subfigures represent the error on the relevant entries and the support identification property by knowledge of  $r$ . For more details on the displayed data we refer to the caption of Figure 6. The results were obtained by Gaussian matrices.

In Figure 8 we compute again the mean-value of the relevant entries, but this time without the

knowledge of  $k$  but the knowledge of  $r$  and therefore  $S_r(x^*)$ : the quantities on the left subfigure are the mean values of  $\|x|_{S_r(x^*)} - x^*|_{S_r(x^*)}\|_{\ell_2}$ . In the right subfigure we attribute a positive match in case  $S_r(x^*) = S_r(x)$  so that the relevant entries of  $x^*$  coincide with the ones of the original signal. By our theory, we expect  $\ell_1$ +SLP and  $\ell_1$ +IHT to produce a high rate of success of correctly recovered support. Actually this is confirmed by the experiments: Both methods do a very accurate recovery, as it gives us almost always 100% of the correct result while the other methods perform worse.

Figures 9, 10, and 11 represent the same statistical data reported respectively in the Figures 6, 7, and 8 but using random subsampled cosine transformation encoding matrices. Without describing the result in detail, we state that they are very similar for these problems as well.

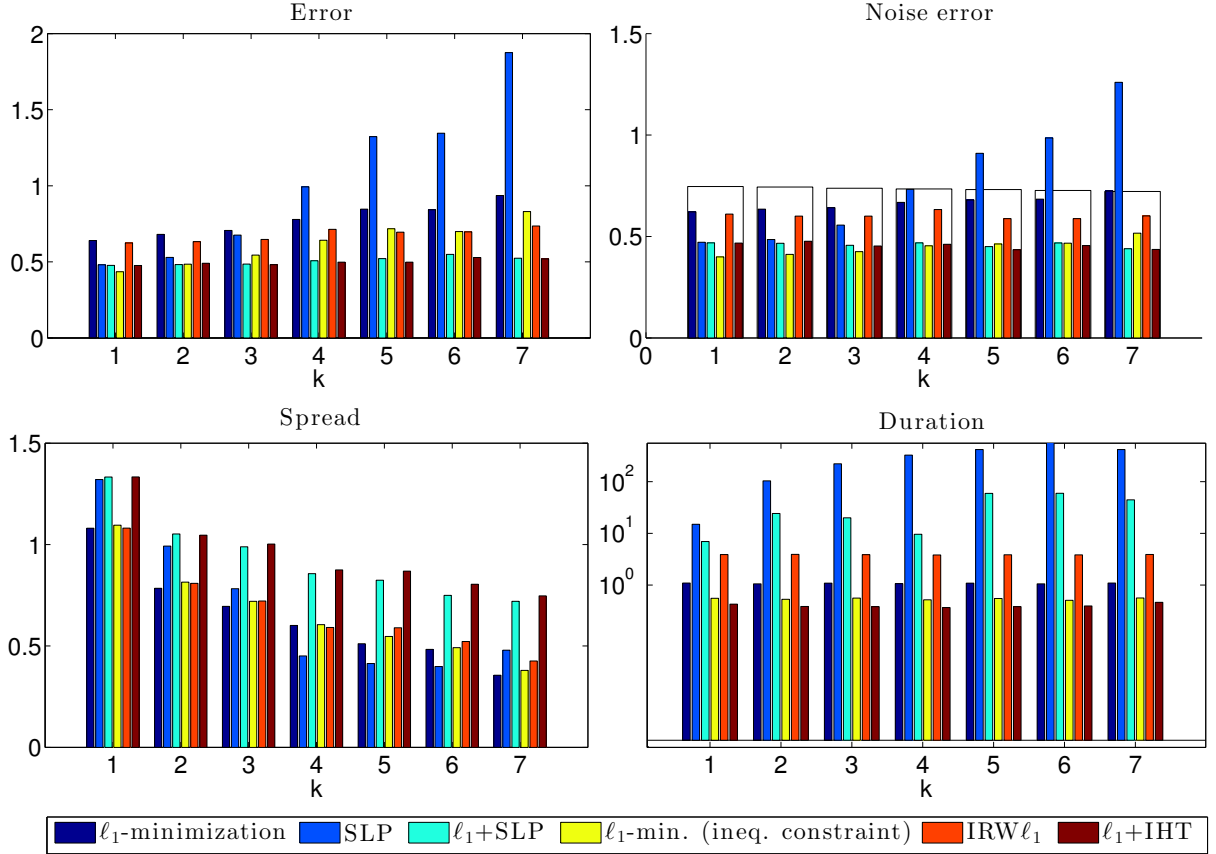


Figure 9: The subfigures represent the errors, the computation time, and the separation of the component. For more details on the displayed data we refer to the caption of Figure 6. The results were obtained by randomly subsampled cosine transformation encoding matrices.

**Phase transition diagrams** To give an even stronger support of the results in the previous paragraph, we extended the results of Figure 8 to a wider range of  $m$  and  $k$ . In Figure 12, we present phase transition diagrams of success rates in support recovery for  $\ell_1$ -minimization, IRW $\ell_1$ ,  $\ell_1$ +SLP, and  $\ell_1$ +IHT in presence of nearly maximally allowed noise, i.e.,  $0.8 = r > \eta = 0.75$ .

To produce phase transition diagrams, we varied the dimension of the measurement vector  $m = 1, \dots, N$  with  $N = 100$ , and solved 20 different problems for all the admissible  $k = \#S_r(x) = 1, \dots, m$ . We colored black all the points  $(m, k)$ , with  $k \leq m$ , which reported 100% of correct support

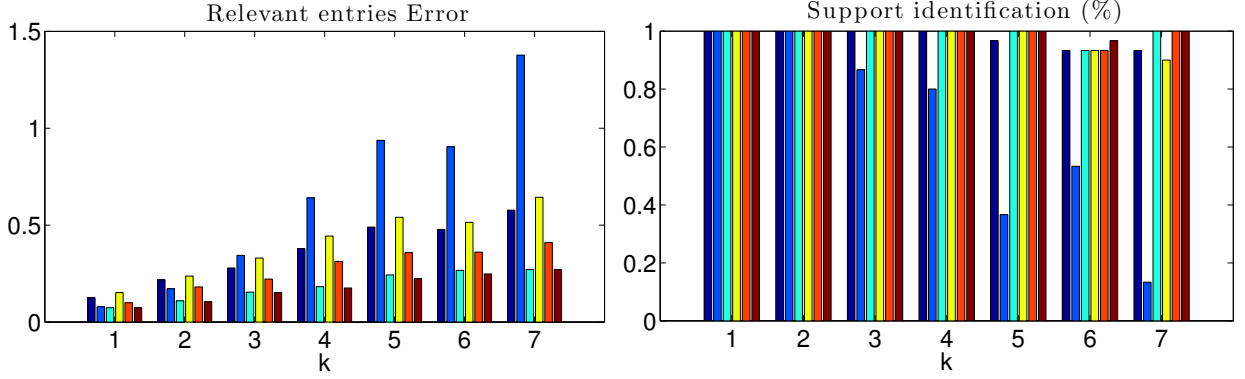


Figure 10: The subfigures represent the error on the relevant entries and the support identification property by knowledge of  $k$ . For more details on the displayed data we refer to the caption of Figure 6. The results were obtained by random subsampled cosine transformation encoding matrices.

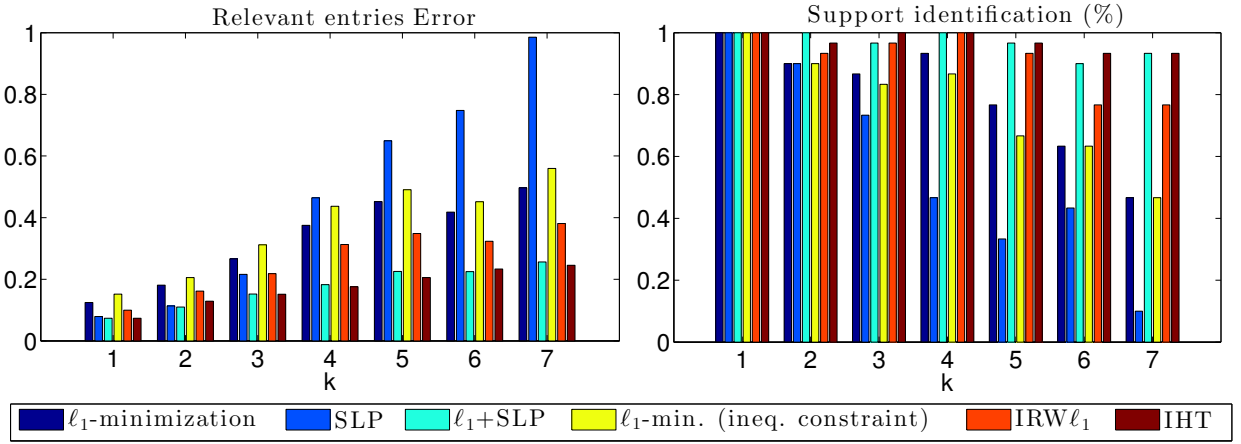


Figure 11: The subfigures represent the error on the relevant entries and the support identification property by knowledge of  $r$ . For more details on the displayed data we refer to the caption of Figure 6. The results were obtained by random subsampled cosine transformation encoding matrices.

identification, and gradually we reduce the tone up to white for the 0% result. The level bound of 50% and 90% is highlighted by a magenta and red line respectively. A visual comparison of the corresponding phase transitions confirms our previous expectations. In particular,  $\ell_1$ +SLP and  $\ell_1$ +IHT very significantly outperform  $\ell_1$ -minimization in terms of correct support recovery. The difference of both methods towards IRW $\ell_1$  is less significant but still important. In Figure 13, we compare the level bounds of 50% and 90% among the four different methods. Observe that the 90% probability bound indicates the largest positive region for  $\ell_1$ +IHT, followed by  $\ell_1$ +SLP, and only eventually by IRW $\ell_1$ , while the bounds are much closer to each other in the case of the 50% bound. Thus, surprisingly,  $\ell_1$ +IHT works in practice even better than  $\ell_1$ +SLP for some range of  $m$ , and offers the most stable support recovery results.

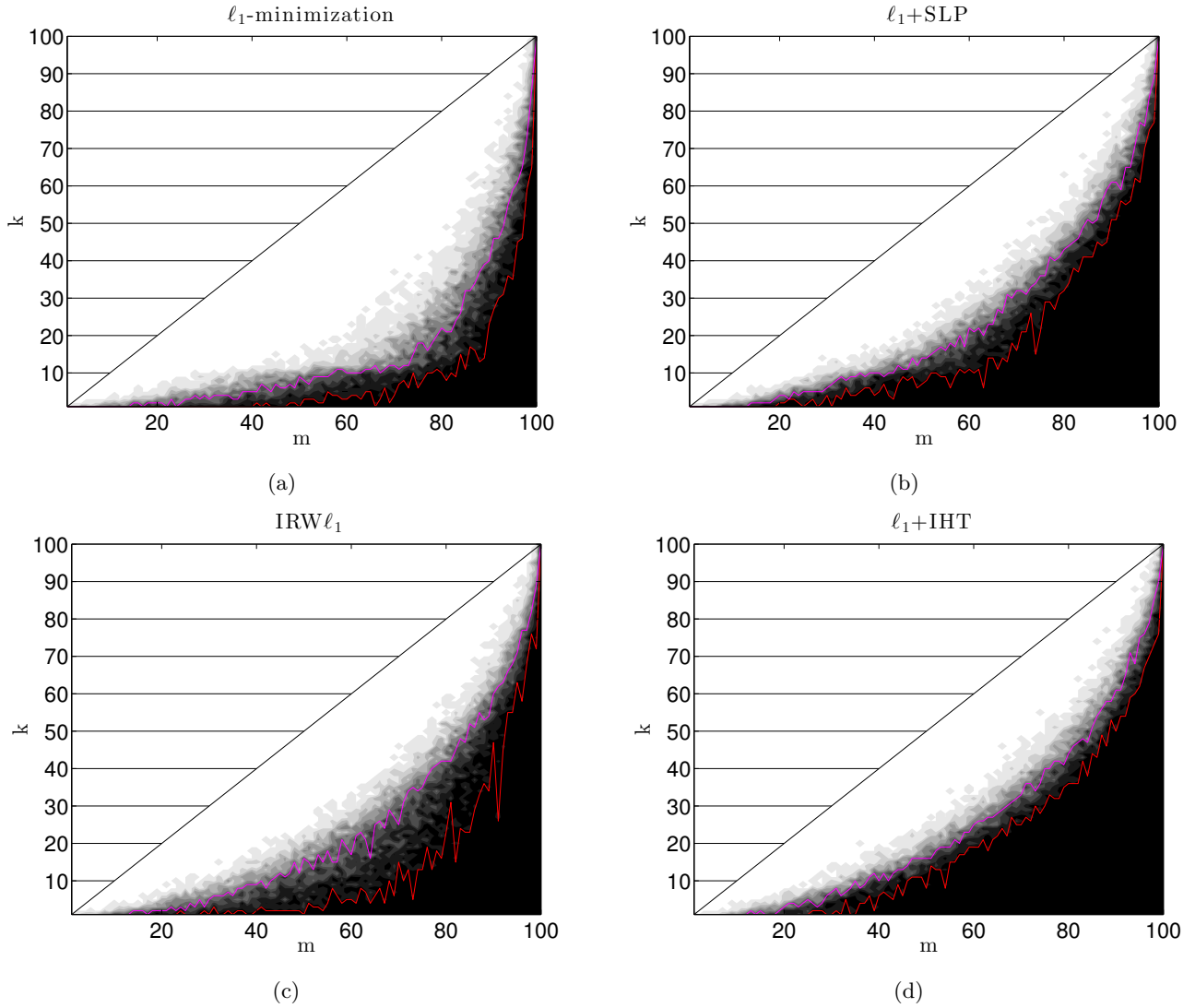


Figure 12: Phase transition diagrams. The black area represents the couple  $(m, k)$  for which we had 100% of support recovery. The results of (a)  $\ell_1$ -minimization, (b)  $\ell_1$ +SLP, (c) IRW $\ell_1$ , and (d)  $\ell_1$ +IHT are reported. Note that the area for  $k > m$  is not admissible. The red line shows the level bound of 90% of support recovery, and the magenta line 50% respectively.

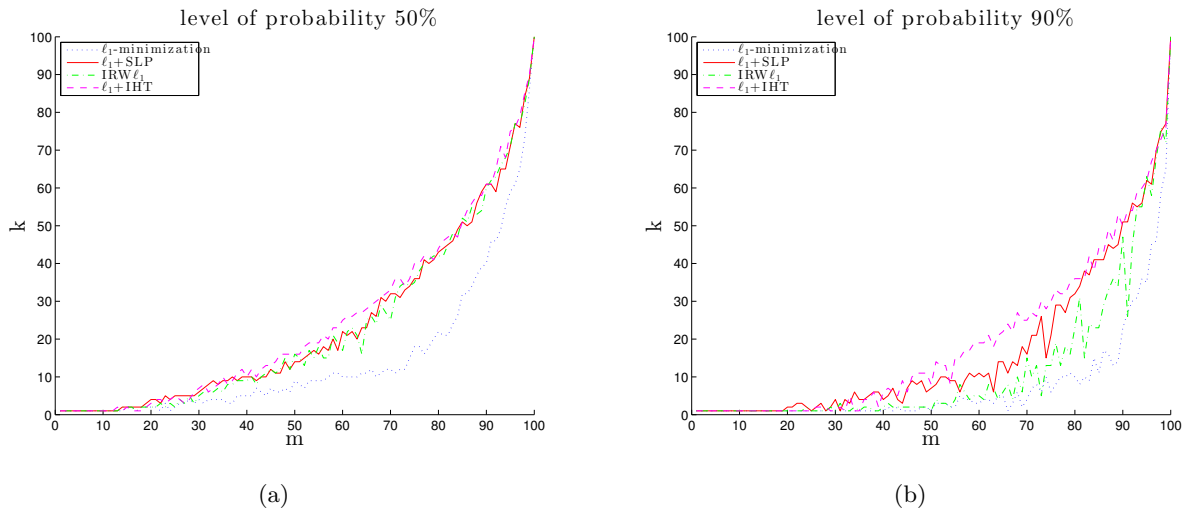


Figure 13: Comparison of phase transition diagrams for  $\ell_1$ -minimization (dark blue, dotted),  $\ell_1$ +SLP (red), IRW  $\ell_1$  (green, dash-dotted), and  $\ell_1$ +IHT (magenta, dashed). The level bound of 50% and 90% as it is displayed in Figure 12 is compared in (a) and (b) respectively.

## Funding

Marco Artina acknowledges the support of the International Research Training Group IGDK 1754 “Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures” of the German Science Foundation. Massimo Fornasier acknowledges the support of the ERC-Starting Grant HDSPCONTR “High-Dimensional Sparse Optimal Control”. Steffen Peter acknowledges the support of the Project “SparsEO: Exploiting the Sparsity in Remote Sensing for Earth Observation” funded by Munich Aerospace.

## References

- [1] B. Alexeev and R. Ward. On the complexity of Mumford-Shah-type regularization, viewed as a relaxed sparsity constraint. *IEEE Trans. Image Process.*, 19(10):2787–2789, 2010.
- [2] E. Arias-Castro and Y. C. Eldar. Noise folding in compressed sensing. *IEEE Signal Process. Lett.*, pages 478–481, 2011.
- [3] M. Artina, M. Fornasier, and F. Solombrino. Linearly constrained nonsmooth and nonconvex minimization. *SIAM J. Opt.*, 23,(3):1904–1937, 2012.
- [4] R. Baraniuk. Compressive sensing. *IEEE Signal Process. Magazine*, 24(4):118–121, 2007.
- [5] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
- [6] T. R. Bedding and H. Kjeldsen. Solar-like oscillations. *Publications of the Astronomical Society of Australia*, 20(2):203–212, 2003.
- [7] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

- [8] E. Candes and T. Tao. The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *The Annals of Statistics*, 35(6):2313–2351, 2007.
- [9] E. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Magazine*, 25(2):21–30, 2008.
- [10] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted  $\ell_1$  minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, 2008.
- [11] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [12] A. Cohen, W. Dahmen, and R. A. DeVore. Compressed sensing and best  $k$ -term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.
- [13] I. CVX Research. CVX: Matlab software for disciplined convex programming, version 2.0. <http://cvxr.com/cvx>, 2012.
- [14] M. Davenport. The RIP and the NSP. @Connexions. <http://cnx.org/content/m37176/1.5/>, Apr. 2011.
- [15] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk. The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range. *IEEE Transactions on Signal Processing*, 60(9):4628–4642, 2012.
- [16] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [17] M. Fornasier and H. Rauhut. Compressive Sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.
- [18] M. Fornasier and R. Ward. Iterative thresholding meets free-discontinuity problems. *Found. Comput. Math.*, 10(5):527–567, 2010.
- [19] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Appl. Numer. Harmon. Anal. Birkhäuser, Boston, 2013.
- [20] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. [http://stanford.edu/~boyd/graph\\_dcp.html](http://stanford.edu/~boyd/graph_dcp.html).
- [21] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak. Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements. In *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, Asilomar’09, Piscataway, NJ, USA, 2009. IEEE Press.
- [22] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak. Sequentially designed compressed sensing. In *Proc. IEEE/SP Workshop on Statistical Signal Processing*, 2012.
- [23] J. Haupt, R. Castro, and R. Nowak. Distilled sensing: Adaptive sampling for sparse detection and estimation. *IEEE Transactions on Information Theory*, 57(9):6222–6235, 2011.
- [24] M. Hügel, H. Rauhut, and T. Strohmer. Remote sensing via  $\ell_1$ -minimization. *Preprint*, 2012.



- [25] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Improved sparse recovery thresholds with two-step reweighted  $\ell_1$  minimization. In *ISIT*, pages 1603–1607, 2010.
- [26] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition, 2008.
- [27] D. Needell. Noisy signal recovery via iterative reweighted  $\ell_1$ -minimization. In *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, Asilomar’09, pages 113–117, Piscataway, NJ, USA, 2009. IEEE Press.
- [28] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [29] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
- [30] J. D. Ridder, C. Barban, F. Baudin, F. Carrier, A. P. Hatzes, S. Hekker, T. Kallinger, W. W. Weiss, A. Baglin, and M. Auvergne. Non-radial oscillation modes with long lifetimes in giant stars, 2009.
- [31] M. Stojnic, W. Xu, A. S. Avestimehr, and B. Hassibi. Compressed sensing of approximately sparse signals. In *ISIT*, pages 2182–2186, 2008.
- [32] J. Treichler, M. A. Davenport, and R. G. Baraniuk. Application of compressive sensing to the design of wideband signal acquisition receivers. In *6th U.S. / Australia Joint Workshop on Defense Applications of Signal Processing (DASP)*, Lihue, Hawaii, Sept. 2009.
- [33] P. Wojtaszczyk. Stability and instance optimality for Gaussian measurements in compressed sensing. *Found. Comput. Math.*, 10(1):1–13, 2010.
- [34] D. Wu, W.-P. Zhu, and M. Swamy. Compressive sensing based speech enhancement in nonsparse noisy environments. *IET Signal Processing*, 7(5):450–457, 2013.
- [35] D. Wu, W.-P. Zhu, and M. Swamy. A note on error bounds of compressed sensing in considering nonsparse noise from the time domain. *preprint*, 2014.