

## ROBUST-SL0 FOR STABLE SPARSE REPRESENTATION IN NOISY SETTINGS

Armin Eftekhari<sup>1</sup>, Massoud Babaie-Zadeh<sup>2</sup>, Christian Jutten<sup>3</sup>, Hamid Abrishami Moghaddam<sup>1</sup>

<sup>1</sup>K.N. Toosi University of Technology, Department of Electrical Engineering, Tehran, Iran

<sup>2</sup>Sharif University of Technology, Department of Electrical Engineering, Tehran, Iran

<sup>3</sup>GIPSA-lab, Grenoble, France

### ABSTRACT

In the last few years, we have witnessed an explosion in applications of sparse representation, the majority of which share the need for finding sparse solutions of underdetermined systems of linear equations (USLE's). Based on recently proposed smoothed  $\ell^0$ -norm (SL0), we develop a noise-tolerant algorithm for sparse representation, namely Robust-SL0, enjoying the same computational advantages of SL0, while demonstrating remarkable robustness against noise. The proposed algorithm is developed by adopting the corresponding optimization problem for noisy settings, followed by theoretically-justified approximation to reduce the complexity. Stability properties of Robust-SL0 are rigorously analyzed, both analytically and experimentally, revealing a remarkable improvement in performance over SL0 and other competing algorithms, in the presence of noise.

**Index Terms**— Sparse representation, overcomplete signal representation, compressed sensing, basis pursuit

### 1. INTRODUCTION

In recent years, identifying relevant sparse solutions of under-determined systems of linear equations (USLE) has become a subject of major interest in signal processing and statistics, with applications spanning vast areas, such as overcomplete signal representation, source separation, and compressed sensing [1, 2]. In these applications, a sparse solution corresponds to an efficient representation of data as a linear combination of some collection of predetermined elements. In the rest of this paper, without loss of generality, we will pursue this problem in the context of sparse overcomplete signal representation. Let  $\mathbf{x} \in \mathbb{R}^n$  be a signal we wish to decompose over a given dictionary  $\mathbf{A} \in \mathbb{R}^{n \times m}$  with columns (atoms)  $\mathbf{a}_i \in \mathbb{R}^n$ . We then may write  $\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^m s_i \mathbf{a}_i$ , where  $\mathbf{s} = [s_1, \dots, s_m]^T \in \mathbb{R}^m$  is the vector of coefficients. This problem is a USLE and (if consistent) has

infinite solutions, hence additional constraints, such as sparsity, should be imposed to arrive at a unique solution. In fact, we are interested to use as small as possible number of atoms to represent the signal. To this end, one may directly seek for solutions to  $\mathcal{P}_0: \min_{\mathbf{s}} \|\mathbf{s}\|_0$  s.t.  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , in which  $\|\mathbf{s}\|_0$  denotes the  $\ell^0$ -norm, i.e. number of nonzero components, of  $\mathbf{s}$ . This is, however, intractable as it requires combinatorial search. Moreover, since any small amount of noise completely changes the  $\ell^0$ -norm of a vector, this method is prone to errors in noisy settings. As a result, alternative approaches have been considered to pursue sparse solutions. Based on the observation that, under some mild constraints, solving  $\mathcal{P}_1: \min_{\mathbf{s}} \|\mathbf{s}\|_1$  s.t.  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , recovers the sparsest solution, a family of algorithms represented by basis pursuit (BP) have been developed [1]. Another family, with FOCUSS as an important member, is iterative re-weighted least squares, with the performance that is generally worse than BP [3]. These algorithms are typically associated with prohibitive computational complexity.

Recently, Mohimani et al. [4] proposed SL0 algorithm for sparse representation. SL0 iteratively minimizes a smoothed version of  $\ell^0$ -norm and was shown to run much faster than the competing algorithms, while producing solutions with the same or better accuracy. In this paper, we improve the performance of SL0 in noisy settings by developing a noise-aware variant of SL0, namely Robust-SL0, which we then analytically prove it to be globally robust against noise. Robust-SL0 enjoys the same computational advantages of SL0, while having higher degree of immunity to noise. Our simulations in presence of noise, verified the improved performance of Robust-SL0 in terms of signal to noise ratio. The rest of this paper is organized as follows. Necessary background on SL0 algorithm is provided in Section 2. The proposed Robust-SL0 is systematically developed in Section 3. In Section 4, stability properties of Robust-SL0 are considered analytically. Experimental results are presented in Section 5 and the paper concludes in Section 6.

### 2. SMOOTHED $\ell^0$

Direct minimization of  $\ell^0$ -norm is mainly impeded by the fact that the  $\ell^0$ -norm of a vector is a discontinuous function

---

This work has been partially supported by Iran NSF (INSF) under contract number 86/994, by Iran Telecommunications Research Center (ITRC), and also by ISMO and French embassy in Iran in the framework of a GundiShapour collaboration program.

of that vector. Interestingly, SL0 approximates  $\ell^0$ -norm of  $\mathbf{s} \in \mathbb{R}^m$  with the continuous function  $m - F_\sigma(\mathbf{s})$ , where  $F_\sigma(\mathbf{s}) = \sum_{i=1}^m f_\sigma(s_i)$  and  $f_\sigma$  belongs to a family of continuous (one variable) functions that, generally speaking, approximate Kronecker delta as  $\sigma \rightarrow 0$ . Therefore, instead of minimizing  $\|\mathbf{s}\|_0$  subject to data, SL0 attempts to solve the problem  $Q: \lim_{\sigma \rightarrow 0} \max_{\mathbf{s}} F_\sigma(\mathbf{s})$  s.t.  $\mathbf{A}\mathbf{s} = \mathbf{x}$ . However, it is observed that  $F_\sigma$ , for small  $\sigma$ , contains numerous local maxima, while having no local maxima for sufficiently large values of  $\sigma$ . Therefore, to avoid getting trapped into local maxima, SL0 solves a sequence of problems of the form  $Q_\sigma: \max_{\mathbf{s}} F_\sigma(\mathbf{s})$  s.t.  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , decreasing  $\sigma$  at each step, and initializing the next step at the maximizer of the previous (larger) value of  $\sigma$ . Each  $Q_\sigma$  is solved using few iterations of gradient ascent. Thus, SL0 algorithm consists of two loops: the external loop that is responsible for gradually decreasing of  $\sigma$  and the internal loop, being a simple steepest ascent algorithm for finding the maximizer of  $Q_\sigma$  for given  $\sigma$ . The convergence analysis of SL0 has been thoroughly considered in [4] and it was shown that, under mild conditions, the sequence of maximizers of  $Q_\sigma$  indeed converges to the unique minimizer of  $\mathcal{P}_0$ , whenever such answer exists. Moreover, SL0 runs significantly faster than the competing algorithms, while producing answers with the same or better accuracy [4].

### 3. ROBUST-SL0

As a result of the presence of noise in practical situations,  $\mathbf{A}\mathbf{s} = \mathbf{x}$  not exactly, but approximately holds and it would be more appropriate to seek for sparse approximate representations, instead. This scenario considers the situation where a noiseless underlying signal  $\mathbf{x}_0$  indeed has a sparse representation  $\mathbf{x}_0 = \mathbf{A}\mathbf{s}_0$ , but we can observe only a noisy version  $\mathbf{x}$  which is corrupted by additive noise  $\mathbf{n}$ :  $\mathbf{x} = \mathbf{x}_0 + \mathbf{n}$ ,  $\|\mathbf{n}\|_2 \leq \epsilon$ . For stable recovery in this noisy situation, the noise-aware variant of  $\mathcal{P}_0$  can be adapted as:  $\mathcal{P}_{0,\epsilon}: \min_{\mathbf{s}} \|\mathbf{s}\|_0$  s.t.  $\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 \leq \epsilon$ . In fact, it has been shown that, when  $\mathbf{s}_0$  is sufficiently sparse and the noise level is given, solving  $\mathcal{P}_{0,\epsilon}$  enables stable recovery of  $\mathbf{s}_0$  with an error which is at worst proportional to the noise level [1]. Just as with  $\mathcal{P}_0$ , the noise-aware variant of SL0, to which we will refer as Robust-SL0, can be adapted for noisy settings:  $Q_\epsilon: \lim_{\sigma \rightarrow 0} \max_{\mathbf{s}} F_\sigma(\mathbf{s})$  s.t.  $\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 \leq \epsilon$ , where  $\sigma \geq 0$ . As we will formally prove in Section 4, solving  $Q_\epsilon$  indeed enables stable recovery of  $\mathbf{s}_0$ , provided some mild conditions are met. For now, however, we consider solving  $Q_\epsilon$ .

As was the case for SL0, to avoid getting trapped into local maxima, we should solve a sequence of problems of the form  $Q_{\epsilon,\sigma}: \max_{\mathbf{s}} F_\sigma(\mathbf{s})$  s.t.  $\frac{1}{2} \|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2^2 < \frac{1}{2} \epsilon^2$ , decreasing  $\sigma$  at each step, and initializing the next step at the maximizer of the previous (larger) value of  $\sigma$ . Therefore, for each  $\sigma$ , we are left with the optimization problem  $Q_{\epsilon,\sigma}$ . To make the inequality constraint implicit in the functional, we rewrite  $Q_{\epsilon,\sigma}$  as:  $\max_{\mathbf{s}} F_\sigma(\mathbf{s}) - I\left(\frac{-1}{2} (\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2^2 - \epsilon^2)\right)$ , where  $I(u)$

takes zero when  $u \leq 0$  and is infinite when  $u > 0$  (i.e. when the constraint is violated). It is easily observed that the differentiable function  $\log(-u)/\tau$  approximates  $I(u)$  and that the approximation becomes more accurate as  $\tau$  increases [5]. Therefore, for large  $\tau$ ,  $Q_{\epsilon,\sigma}$  can be approximated by  $Q_{\epsilon,\sigma,\tau}: \max_{\mathbf{s}} F_\sigma(\mathbf{s}) + \frac{1}{\tau} \tilde{I}(\mathbf{s})$ , where we have  $\tilde{I}(\mathbf{s}) = \log\left(\frac{-1}{2} (\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2^2 - \epsilon^2)\right)$ . Nonetheless, maximizing  $Q_{\epsilon,\sigma,\tau}$  for large  $\tau$  is difficult as the gradient varies rapidly near the boundary of the feasible set  $\mathcal{S} = \{\mathbf{s} \mid \|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 \leq \epsilon\}$  [5]. Instead, we may solve a sequence of problems of the form  $Q_{\epsilon,\sigma,\tau}$ , increasing  $\tau$  at each step, and initializing the next step at the maximizer of the previous (smaller) value of  $\tau$  [5]. Using iterations of gradient ascent to solve each  $Q_{\epsilon,\sigma,\tau}$ , the final algorithm would consist of three loops instead of two loops of SL0 and the computational efficiency of SL0 is potentially removed. Therefore, we have to resort to approximations, the detailed justification of which is not reported here and only an insightful explanation is given in the following. Notice that even for relatively large, but meaningful noise levels, the feasible set  $\mathcal{S}$  is still a thin convex region enclosing  $\mathbf{A}\mathbf{s} = \mathbf{x}$ . In fact, it can be shown that for small  $\epsilon$ , the gradient of  $\tilde{I}(\mathbf{s})$ ,  $\nabla \tilde{I}(\mathbf{s})$ , is approximately perpendicular to  $\mathbf{A}\mathbf{s} = \mathbf{x}$  and hence, informally speaking,  $\nabla \tilde{I}(\mathbf{s})$  only moves  $\mathbf{s}$  along the so-called minor axis of  $\mathcal{S}$ . In other words, the second term in the functional of  $Q_{\epsilon,\sigma,\tau}$  has negligible effect and can be omitted. Therefore, to find the maximizer of  $Q_{\epsilon,\sigma}$ , instead of solving  $Q_{\epsilon,\sigma,\tau}$  for an increasing sequence of  $\tau$ , we can simply maximize  $F_\sigma$  while forcing  $\mathbf{s}$  to stay in the feasible set and yet be sure that the answer approximates the maximizer of  $Q_{\epsilon,\sigma}$ , provided  $\epsilon$  is small. For instance, one can use gradient ascent to update  $\mathbf{s}$  and project the updated  $\mathbf{s}$  onto  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , whenever it falls outside the feasible set<sup>1</sup>. To be more specific, let  $f_\sigma(s) = e^{-s^2/2\sigma^2}$ , which indeed meets the requirements of SL0 (see Section 2). Then, the gradient of  $F_\sigma(\mathbf{s})$  would be  $\nabla F_\sigma(\mathbf{s}) = \frac{-1}{\sigma^2} \cdot [s_1 e^{-s_1^2/2\sigma^2}, \dots, s_m e^{-s_m^2/2\sigma^2}]^T$  and the update rule of gradient ascent becomes  $\mathbf{s} \leftarrow \mathbf{s} + \mu_\sigma \nabla F_\sigma$ . Following the same reasoning as in [4], we choose  $\mu_\sigma = \sigma^2 \mu_0$  and the update rule reduces to  $\mathbf{s} \leftarrow \mathbf{s} - \mu_0 [s_1 e^{-s_1^2/2\sigma^2}, \dots, s_m e^{-s_m^2/2\sigma^2}]^T$ , followed by projecting onto  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , whenever  $\mathbf{s}$  falls outside the feasible set. The remaining issues of initialization and selection of the sequence of  $\sigma$  are addressed as in [4]. Robust-SL0 is initialized by the minimum  $\ell^2$ -norm solution of  $\mathbf{A}\mathbf{s} = \mathbf{x}$ , i.e.  $\hat{\mathbf{s}}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{x}$ . In addition, the sequence of  $\sigma$  is characterized by  $\sigma_1 = 2 \max_i |s_i|$  and  $\sigma_{k+1} = c\sigma_k$  where  $c \leq 1$  and  $1 \leq k < K$ . Fig. 1 summarizes the Robust-SL0 algorithm. Further study of the proposed algorithm reveals that Robust-SL0 has the same computational complexity as SL0, while being more robust against noise, as we shall verify in Section 5.

<sup>1</sup> $\mathbf{A}\mathbf{s} = \mathbf{x}$  is naturally selected as it is the maximum likelihood loci for  $\mathbf{s}$  with no other prior in hand.

- Initialization:
  1. Let  $\hat{\mathbf{s}}_0$  equal to the minimum  $\ell^2$  norm solution of  $\mathbf{s} = \mathbf{x}$
  2. Choose a suitable decreasing sequence for  $\sigma$ ,  $[\sigma_1, \dots, \sigma_K]$  (see Section 3).
- For  $k = 1, \dots, K$ 
  1. Let  $\sigma = \sigma_k$
  2. Maximize (approximately) the function  $F_\sigma$  on the feasible set  $\mathcal{S} = \{\mathbf{s} \mid \|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 \leq \epsilon\}$ :
    - Initialization:  $\mathbf{s} = \hat{\mathbf{s}}_{k-1}$
    - For  $j = 1, \dots, L$ 
      - a) Let  $\Delta\mathbf{s} = [s_1 e^{-s_1^2/2\sigma^2}, \dots, s_m e^{-s_m^2/2\sigma^2}]^T$ .
      - b) Let  $\mathbf{s} \leftarrow \mathbf{s} - \mu_0 \Delta\mathbf{s}$ .
      - c) If  $\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 > \epsilon$  project  $\mathbf{s}$  onto  $\mathbf{A}\mathbf{s} = \mathbf{x}$  with:
 
$$\mathbf{s} \leftarrow \mathbf{s} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{s} - \mathbf{x}).$$
    - Set  $\hat{\mathbf{s}}_k = \mathbf{s}$ .
- Final answer is  $\hat{\mathbf{s}} = \hat{\mathbf{s}}_K$

Fig. 1. Robust-SL0 algorithm.

#### 4. STABILITY PROPERTIES OF ROBUST-SL0

In this section, we formally investigate the stability properties of the proposed Robust-SL0. We shall prove that, whenever sufficient sparsity is present, solving  $Q_\epsilon$  will indeed recover the unique minimizer of  $\ell^0$ -norm with error which is at worst proportional to the noise level. The proof offered here, is based upon the assumption that for a given  $\sigma$ , exact maximizer of  $F_\sigma$  is found without getting trapped into local maxima, which was achieved by gradually decreasing of  $\sigma$ . However, the question of how much gradually, is not answered in this paper, but was experimentally investigated in [4]. Also, without loss of generality, it is assumed that the columns of  $\mathbf{A}$  are normalized to unity. We now start with a few definitions. Given  $\mathbf{A} \in \mathbb{R}^{n \times m}$  and  $m > n$ , mutual coherence of  $\mathbf{A}$ ,  $M(\mathbf{A})$ , is defined as the maximum off-diagonal element of  $\mathbf{G} = \mathbf{A}^T \mathbf{A}$  [1]. Also, given  $\alpha \in \mathbb{R}^+$ , we will denote by  $I_\alpha(\mathbf{s})$  the set of indices  $i$  of  $\mathbf{s}$  for which  $|s_i| > \alpha$ . The cardinality of  $I_\alpha(\mathbf{s})$ , i.e. the number of elements of  $\mathbf{s}$  with amplitude  $> \alpha$ , will be denoted by  $|\mathbf{s}|_\alpha$ . Before stating the stability properties of Robust-SL0 (Theorem 1), a few lemmas are given, that are either proved here or the proof of which is offered in corresponding references.

**Lemma 1** [1]: Given an  $h \times h$  symmetric matrix  $\mathbf{H}$  with diagonal entries set to one and off-diagonal entries  $\leq M$  in amplitude, all eigenvalues of  $\mathbf{H}$  are  $\geq 1 - M(h - 1)$ .

**Lemma 2:** Let  $\mathbf{v} \in \mathbb{R}^m$  satisfy  $|\mathbf{v}|_\alpha \leq 2k$  and  $\|\mathbf{A}\mathbf{v}\|_2 \leq 2\epsilon$ . Then  $\|\mathbf{v}\|_2 \leq \psi(k, \epsilon, \alpha) = \frac{2\epsilon}{\sqrt{1-M(2k-1)}} + \left(1 + \frac{1}{\sqrt{1-M(2k-1)}}\right) am$ .

*Proof:* Suppose  $\mathbf{v}_1$  and  $\mathbf{v}_2$  comprise the elements of  $\mathbf{v}$  being  $> \alpha$  and  $\leq \alpha$  in amplitude, respectively. Note that,  $\|\mathbf{A}_1 \mathbf{v}_1\|_2 - \|\mathbf{A}_2 \mathbf{v}_2\|_2 \leq 2\epsilon$  or  $\|\mathbf{A}_1 \mathbf{v}_1\|_2 \leq 2\epsilon + \|\mathbf{A}_2 \mathbf{v}_2\|_2$ . At the same time,  $\|\mathbf{A}_2 \mathbf{v}_2\|_2 \leq \sum_{i \in I_\alpha} \|\mathbf{a}_i\|_2 |v_i|$ , where  $\mathbf{a}_i$  and  $v_i$  are the  $i$ th column of  $\mathbf{A}$  and  $i$ th element of  $\mathbf{v}$ , respectively. Hence,  $\|\mathbf{A}_2 \mathbf{v}_2\|_2 \leq \sum_{i \in I_\alpha} |v_i| \leq \alpha(m - |\mathbf{v}|_\alpha) \leq am$ , from which we conclude

$\|\mathbf{A}_1 \mathbf{v}_1\|_2 \leq 2\epsilon + \|\mathbf{A}_2 \mathbf{v}_2\|_2 \leq 2\epsilon + am$ . Now, notice that  $\|\mathbf{A}_1 \mathbf{v}_1\|_2^2 = \mathbf{v}_1^T \mathbf{A}_1^T \mathbf{A}_1 \mathbf{v}_1$ . Observe that, all diagonal elements of  $\mathbf{A}_1^T \mathbf{A}_1$  are equal to one. Also, by the definition of mutual coherence, all off-diagonal elements of  $\mathbf{A}_1^T \mathbf{A}_1$  are  $\leq M$ . Therefore, Lemma 1 requires all eigenvalues of  $\mathbf{A}_1^T \mathbf{A}_1$  and  $\mathbf{A}_1$  to be  $\geq 1 - M(|\mathbf{v}|_\alpha - 1)$  and  $\geq \sqrt{1 - M(|\mathbf{v}|_\alpha - 1)}$ , respectively. Then it follows that  $\sqrt{1 - M(|\mathbf{v}|_\alpha - 1)} \|\mathbf{v}_1\|_2 \leq \|\mathbf{A}_1 \mathbf{v}_1\|_2$  and consequently  $\sqrt{1 - M(|\mathbf{v}|_\alpha - 1)} \|\mathbf{v}_1\|_2 \leq 2\epsilon + am$  or  $\|\mathbf{v}_1\|_2 \leq \frac{2\epsilon + am}{\sqrt{1 - M(|\mathbf{v}|_\alpha - 1)}}$ . Finally, since  $|\mathbf{v}|_\alpha \leq 2k$ , an upper is attained for  $\|\mathbf{v}\|_2$ :  $\|\mathbf{v}\|_2 \leq \|\mathbf{v}_1\|_2 + \|\mathbf{v}_2\|_2 \leq \frac{2\epsilon + am}{\sqrt{1 - M(|\mathbf{v}|_\alpha - 1)}} + am \leq \frac{2\epsilon}{\sqrt{1 - M(2k - 1)}} + \left(1 + \frac{1}{\sqrt{1 - M(2k - 1)}}\right) am$  ■

**Lemma 3:** Let  $\hat{\mathbf{s}}_{\epsilon, \sigma}$  be the maximizer of  $Q_{\epsilon, \sigma}$  and  $f_\sigma$  belongs to the family of functions described in [4] (see Section 2). Also, suppose that  $\mathbf{s}_0$  is the unique sparsest answer of noise-free problem  $\mathbf{A}\mathbf{s} = \mathbf{x}_0$  and has  $k$  nonzero elements. Then, for any given  $\alpha \in \mathbb{R}^+$ , there exists  $\sigma_0$  such that for all  $\sigma \leq \sigma_0$ , we have  $|\hat{\mathbf{s}}_{\epsilon, \sigma}|_\alpha \leq k$ .

*Flavor of the Proof:* Detailed proof, which uses similar ideas to [4], is not reported here and only the proof outline is given. Utilizing the properties of  $f_\sigma$ , a proper upper bound on  $F_\sigma(\hat{\mathbf{s}}_{\epsilon, \sigma})$  and a lower bound on  $F_\sigma(\mathbf{s}_0)$  are attained. On the other hand, since  $\mathbf{s}_0$  is merely feasible for  $Q_{\epsilon, \sigma}$ , while  $\hat{\mathbf{s}}_{\epsilon, \sigma}$  is optimal, we have  $F_\sigma(\mathbf{s}_0) \leq F_\sigma(\hat{\mathbf{s}}_{\epsilon, \sigma})$ . Combining these two observations eventually yields  $|\mathbf{s}_{\epsilon, \sigma}|_\alpha \leq k$ .

**Lemma 4** [1]: Suppose the noiseless case  $\mathbf{A}\mathbf{s}_0 = \mathbf{x}_0$ , where  $\mathbf{s}_0$  satisfies  $\|\mathbf{s}_0\|_0 = k \leq \frac{1}{2}(1 + M^{-1})$  and  $M$  stands for the mutual coherence of  $\mathbf{A}$ . Then,  $\mathbf{s}_0$  is the unique sparsest such representation of  $\mathbf{x}_0$ . Also, if  $\hat{\mathbf{s}}_{0, \epsilon}$  denotes the minimizer of  $\mathcal{P}_{0, \epsilon}$ , then  $\|\hat{\mathbf{s}}_{0, \epsilon} - \mathbf{s}_0\|_2 \leq \frac{2\epsilon}{\sqrt{1 - M(2k - 1)}}$ , for any  $\epsilon \geq 0$ .

**Theorem 1:** Suppose the noiseless case  $\mathbf{A}\mathbf{s}_0 = \mathbf{x}_0$ , where  $\mathbf{s}_0$  satisfies  $\|\mathbf{s}_0\|_0 = k \leq \frac{1}{2}(1 + M^{-1})$  and  $M$  stands for the mutual coherence of  $\mathbf{A}$ . Then, denoting the maximizer of  $Q_{\epsilon, \sigma}$  by  $\hat{\mathbf{s}}_{\epsilon, \sigma}$ , we have  $\lim_{\sigma \rightarrow 0} \|\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{s}_0\|_2 \leq \frac{2\epsilon}{\sqrt{1 - M(2k - 1)}}$ .

*Proof:* As implied by Lemma 4,  $\mathbf{s}_0$  is indeed the unique sparsest answer. Now, it suffices to show that, for any desired accuracy  $\beta \geq 2\epsilon/\sqrt{1 - M(2k - 1)}$ , there exists  $\sigma_0$  such that for all  $\sigma \leq \sigma_0$ , we have  $\|\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{s}_0\|_2 \leq \beta$ . First, note that  $\|\mathbf{A}\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{x}\|_2 \leq \epsilon$  and  $\|\mathbf{A}\mathbf{s}_0 - \mathbf{x}\|_2 \leq \epsilon$  require that  $\|\mathbf{A}(\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{s}_0)\|_2 \leq 2\epsilon$ . Also, Lemma 3 implies that, for any given  $\alpha$ , there exists  $\sigma_0$  such that for all  $\sigma \leq \sigma_0$ ,  $|\hat{\mathbf{s}}_{\epsilon, \sigma}|_\alpha \leq k$ . Before continuing with the rest of the proof, we need the following observation.

**Lemma 5:**  $|\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{s}_0|_\alpha \leq |\hat{\mathbf{s}}_{\epsilon, \sigma}|_\alpha + |\mathbf{s}_0|_\alpha$ .

*Proof:* Let us denote indices of vector by last subscript  $i$  and the cardinality of a set  $X$  with  $|X|$ . Now, observe that number of indices for which either  $s_{0,i}$  or  $\hat{s}_{\epsilon, \sigma, i}$  are  $\geq \alpha$  is less than  $|\hat{\mathbf{s}}_{\epsilon, \sigma}|_\alpha + |\mathbf{s}_0|_\alpha$ , since clearly  $|I_\alpha(\mathbf{s}_0) \cup I_\alpha(\hat{\mathbf{s}}_{\epsilon, \sigma})| \leq |I_\alpha(\mathbf{s}_0)| + |I_\alpha(\hat{\mathbf{s}}_{\epsilon, \sigma})| = |\mathbf{s}_0|_\alpha + |\hat{\mathbf{s}}_{\epsilon, \sigma}|_\alpha$ . Also, it is observed that other indices of  $\hat{\mathbf{s}}_{\epsilon, \sigma} - \mathbf{s}_0$  cannot exceed  $\alpha$ . This is because

$\forall i \notin I_\alpha(\mathbf{s}_0)$ ,  $s_{0,i} = 0$  and  $\forall i \in I_\alpha(\hat{\mathbf{s}}_{\epsilon,\sigma})$ ,  $|s_{\epsilon,\sigma,i}| \leq \alpha$ , from which we get  $\forall i \notin I_\alpha(\mathbf{s}_0) \cup I_\alpha(\hat{\mathbf{s}}_{\epsilon,\sigma})$ ,  $|\hat{s}_{\epsilon,\sigma,i} - s_{0,i}| \leq \alpha$ . These two observations together yield  $|\hat{\mathbf{s}}_{\epsilon,\sigma} - \mathbf{s}_0|_\alpha \leq |\mathbf{s}_0|_\alpha + |\hat{\mathbf{s}}_{\epsilon,\sigma}|_\alpha$ . ■

With  $|\mathbf{s}_0|_\alpha$  and  $|\hat{\mathbf{s}}_{\epsilon,\sigma}|_\alpha$  being  $\leq k$ , Lemma 5 implies  $|\hat{\mathbf{s}}_{\epsilon,\sigma} - \mathbf{s}_0|_\alpha \leq |\hat{\mathbf{s}}_{\epsilon,\sigma}|_\alpha + |\mathbf{s}_0|_\alpha \leq 2k$ . Assuming  $\mathbf{v} = \hat{\mathbf{s}}_{\epsilon,\sigma} - \mathbf{s}_0$  and applying Lemma 2, we get  $\|\hat{\mathbf{s}}_{\epsilon,\sigma} - \mathbf{s}_0\|_2 \leq \psi(k, \epsilon, \alpha)$ . Hence, for any  $\beta \geq 2\epsilon/\sqrt{1-M(2k-1)}$ , it suffices to select  $\alpha$  to satisfy  $\beta = \psi(k, \epsilon, \alpha)$  and then Lemma 3 insures that for this value of  $\alpha$  there exists  $\sigma_0$  such that for all  $\sigma \leq \sigma_0$ ,  $|\hat{\mathbf{s}}_{\epsilon,\sigma}|_\alpha \leq k$  or  $|\hat{\mathbf{s}}_{\epsilon,\sigma} - \mathbf{s}_0|_\alpha \leq 2k$ . This completes the proof. ■

**Remark.** For a given  $\mathbf{A}$ , Theorem 1 states that solving  $\mathcal{Q}_\epsilon$  enables recovery of the sparsest solution with an error less than  $2\epsilon/\sqrt{1-M(2k-1)}$ , whenever such answer exists and  $k = \|\mathbf{s}_0\|_0 \leq \frac{1}{2}(1+M^{-1})$ . On the other hand, for a given  $\mathbf{A}$ , best existing result implies that solving the noise-aware variant of  $\mathcal{P}_1$ , i.e.  $\mathcal{P}_{1,\epsilon} : \min_{\mathbf{s}} \|\mathbf{s}\|_1$  s.t.  $\|\mathbf{A}\mathbf{s} - \mathbf{x}\|_2 \leq \epsilon$ , enables stable recovery with an error less than  $2\epsilon/\sqrt{1-M(4k-1)}$ , provided such answer exists and  $k = \|\mathbf{s}_0\|_0 \leq \frac{1}{4}(1+M^{-1})$  [1]. During our simulations, for a large range of noise levels, Robust-SL0 demonstrated improved performance compared to the noise-aware variant of BP.

## 5. SIMULATION RESULTS

In this section, the performance of the proposed algorithm in presence of noise is quantitatively compared to FOCUSS [7] and  $\ell_1$ -magic [6], as a powerful implementation of the noise-tolerant variant of BP. The simulations are performed using synthetically generated signals. Each sparse signal  $\mathbf{s}_0$  is independently obtained using a Bernoulli-Gaussian model. To be specific, each source is derived from a  $\mathcal{N}(0,1)$  density with probability  $p$  and from a  $\mathcal{N}(0, \sigma_{\text{off}}^2)$  with probability  $1-p$ , where  $\sigma_{\text{off}} \ll 1$ . Here,  $\sigma_{\text{off}}$  is set to zero. In addition, instead of directly working with  $p$ , sparsity of  $\mathbf{s}_0$  is controlled using  $k = mp$ , representing the average number of nonzero elements of  $\mathbf{s}_0$ . To construct each column of the dictionary  $\mathbf{A}$ , samples are drawn from a  $\mathcal{N}(0,1)$  density, followed by normalization to unity. Finally, the observations  $\mathbf{x}$  are generated using the noisy model  $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$ , where  $\mathbf{n}$  obeys  $\mathcal{N}(0, \sigma_n^2 \mathbf{I}_{n \times n})$ , and  $\mathbf{I}_{n \times n}$  stands for the  $n \times n$  unity matrix. The signal to noise ratio (SNR), defined as  $20 \log(\|\mathbf{s}_0\|_2 / \|\mathbf{s}_0 - \hat{\mathbf{s}}\|_2)$ , with  $\hat{\mathbf{s}}$  denoting our estimation, will be used as our measure of performance. In this experiment, the following set of parameters are used:  $m = 1000$ ,  $n = 400$ ,  $p = 0.1$ . The sequence of  $\sigma$  is chosen as described in Section 3 with  $c = .5$ . The internal loop of the algorithm is characterized by  $L = 10, \mu_0 = 1$ . Finally, we have set  $\epsilon = \sigma_n$ . The noise power  $\sigma_n^2$  is varied between 0 and .5 and the average performance of algorithms over 100 runs is reported in Fig. 2. The elapsed time for Robust-SL0, SL0, BP, and FOCUSS was .09, .09, 1.24, and .82 seconds, respectively, with a regular Intel Core 2 Duo processor. Obtained results reveal that, for a large range of noise levels, Robust-SL0 is

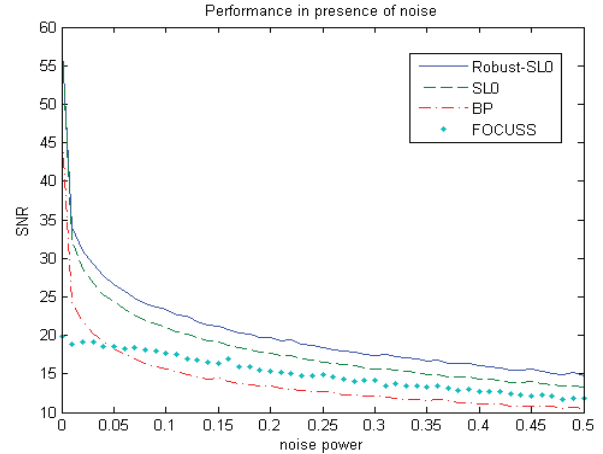


Fig. 2. Averaged SNR (in dB) as a function of noise power  $\sigma_n^2$ .

more immune to noise compared to other competing algorithms. Finally, it is worth noting that the dependence of Robust-SL0 on its parameters is expectedly similar to SL0, and has been thoroughly scrutinized in [4].

## 6. CONCLUSIONS

In this article, we presented Robust-SL0 for stable recovery of sparse solutions of USLE's in presence of noise. Theoretical and experimental justifications concerning Robust-SL0 have also been closely considered. Strong theoretical justification in companion with remarkable experimental results and low computational complexity, suggest Robust-SL0 as a powerful choice for real-world applications involving sparse representation.

## 7. REFERENCES

- [1] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Info. Theory*, vol. 52, no. 1, pp. 6–18, Jan 2006.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [3] I. F. Gorodnitsky, and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS, a re-weighted minimum norm algorithm," *IEEE Trans. on Signal Proc.*, vol. 45, pp. 600–616, March 1997.
- [4] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed L0 norm," *IEEE Trans. on Signal Proc.*, vol. 57, no. 1, pp. 289–301, January 2009.
- [5] S. Boyd, and L. Vandenberghe, "Convex Optimization", *Cambridge University Press*, 2004.
- [6] E. Candes and J. Romberg, " $\ell_1$ -magic: Recovery of sparse signals via convex programming," 2005, URL: [www.acm.caltech.edu/l1magic/downloads/l1magic.pdf](http://www.acm.caltech.edu/l1magic/downloads/l1magic.pdf).
- [7] Available online at: <http://dsp.ucsd.edu/~jfmurray/software.htm>