On the Stable Recovery of the Sparsest Overcomplete Representations in Presence of Noise

Massoud Babaie-Zadeh, Christian Jutten

On the Stable Recovery of the Sparsest Overcomplete Representations in Presence of Noise

Massoud Babaie-Zadeh and Christian Jutten, *Fellow, IEEE*

***Abstract*—Let x be a signal to be sparsely decomposed over a redundant dictionary A, i.e., a sparse coefficient vector s has to be found such that $\mathrm{x} = \mathrm{As}$. It is known that this problem is inherently unstable against noise, and to overcome this instability, Donoho, Elad and Temlyakov ["Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006] have proposed to use an "approximate" decomposition, that is, a decomposition satisfying $\|\mathrm{x} - \mathrm{As}\| \leq \delta$ rather than satisfying the exact equality $\mathrm{x} = \mathrm{As}$. Then, they have shown that if there is a decomposition with $\|\mathrm{s}\|_0 < (1 + M^{-1})/2$, where M denotes the *coherence* of the dictionary, this decomposition would be stable against noise. On the other hand, it is known that a sparse decomposition with $\|\mathrm{s}\|_0 < (1/2)\mathrm{spark}(\mathrm{A})$ is unique. In other words, although a decomposition with $\|\mathrm{s}\|_0 < (1/2)\mathrm{spark}(\mathrm{A})$ is unique, its stability against noise has been proved only for highly more restrictive decompositions satisfying $\|\mathrm{s}\|_0 < (1 + M^{-1})/2$, because usually $(1 + M^{-1})/2 \ll (1/2)\mathrm{spark}(\mathrm{A})$. This limitation maybe had not been very important before, because $\|\mathrm{s}\|_0 < (1 + M^{-1})/2$ is also the bound which guaranties that the sparse decomposition can be found via minimizing the ℓ^1 norm, a classic approach for sparse decomposition. However, with the availability of new algorithms for sparse decomposition, namely SL0 and robust-SL0, it would be important to know whether or not unique sparse decompositions with $(1 + M^{-1})/2 \leq \|\mathrm{s}\|_0 < (1/2)\mathrm{spark}(\mathrm{A})$ are stable. In this correspondence, we show that such decompositions are indeed stable. In other words, we extend the stability bound from $\|\mathrm{s}\|_0 < (1 + M^{-1})/2$ to the whole uniqueness range $\|\mathrm{s}\|_0 < (1/2)\mathrm{spark}(\mathrm{A})$. In summary, we show that *all unique sparse decompositions are stably recoverable*. Moreover, we see that sparser decompositions are "more stable."**

***Index Terms*—Compressed sensing, overcomplete dictionaries, sparse component analysis (SCA), sparse recovery, sparse signal decomposition.**

I. Introduction

Let $\mathbf{A}$ be an $n \times m$ matrix with $m > n$, and consider the underdetermined system of linear equations (USLE) $\mathbf{As} = \mathbf{x}$. Such a linear system has typically infinitely many solutions, but let consider its sparsest solution, that is, a solution $\mathbf{s}_0$ which has as much as possible zero components.

This problem has recently attracted a lot of attention from many different viewpoints. It is used, for example, in compressed sensing (CS) [2]–[4], underdetermined sparse component analysis (SCA) and source separation [5]–[8], atomic decomposition on overcomplete dictionaries [9], [1], decoding real field codes [10], image deconvolution [11], [12], image denoising [13], electromagnetic imaging and direction of arrival (DOA) finding [14], etc.

Manuscript received July 12, 2009; accepted May 04, 2010. Date of publication June 10, 2010; date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kostantinos I. Diamantaras. This work has been partially funded by the Iran National Science Foundation (INSF), by the Iran Telecom Research Center (ITRC), and also by the Center for International Research and Collaboration (ISMO) and the French Embassy in Tehran, Iran, in the framework of a GundiShapour collaboration program.

M. Babaie-Zadeh is with the Electrical Engineering Department, Sharif University of Technology, Tehran, Iran (e-mail: mbzadeh@yahoo.com).

C. Jutten is with the GIPSA-Lab, University of Grenoble, and the Institut Universitaire de France, France (e-mail: Christian.Jutten@gipsa-lab.grenoble-inp.fr)*[AU: please provide missing postal code and cities--ed.]*

Digital Object Identifier 10.1109/TSP.2010.2052357

In atomic decomposition viewpoint [15], the columns of $\mathbf{A}$ are called 'atoms' and the matrix $\mathbf{A}$ is called the "dictionary" over which the "signal," the representation is not unique, but by the sparsest solution, we are looking for the representation which uses as small as possible number of atoms to represent the signal.

Sparse solutions of underdetermined linear systems would not be useful unless positive answers can be provided for the following three questions:

1) **Uniqueness:** Is such a solution unique?
2) **Practical algorithm:** Is it practically possible to find the sparsest solution of an USLE?
3) **Stability against noise:** Doesn't a small amount of noise result in a completely different sparse solution?

In this correspondence we study the third question, and we generalize previously available results. To better explain the problem and our contribution, we firstly do a brief review in Section II on the available results about the above questions, and then explain in Section II-D what our contribution is. We state then the main theorem in Section III. Finally, a generalized result will be stated in Section IV.

II. Problem Statement

A. Uniqueness?

The uniqueness problem has been addressed in [14], [16], and [17], and it has been shown that if an underdetermined linear system has a sparse enough solution, it would be its unique sparsest solution. More precisely:

Theorem 1 (Uniqueness [16], [17]): Let $\mathrm{spark}(\mathbf{A})$ denote the minimum number of columns of $\mathbf{A}$ that are linearly dependent, and $\|\cdot\|_0$ denotes the ℓ^0 norm of a vector (i.e., the number of its non-zero components). Then if the USLE $\mathbf{As} = \mathbf{x}$ has a solution $\mathbf{s}_0$ for which $\|\mathbf{s}_0\|_0 < (1/2)\mathrm{spark}(\mathbf{A})$, it is its unique sparsest solution.

A special case of this uniqueness theorem has been stated in [14]: if $\mathbf{A}$ has the unique representation property (URP), that is, if all $n \times n$ submatrices of $\mathbf{A}$ are non-singular, then $\mathrm{spark}(\mathbf{A}) = n + 1$ ad hence $\|\mathbf{s}_0\|_0 \leq (n/2)$ implies that $\mathbf{s}_0$ is the unique sparsest solution.

B. Practical Algorithm?

Finding the sparsest solution of an USLE can be expressed as

$$(P_0): \quad \text{Minimize } \|\mathbf{s}\|_0 \quad \text{subject to} \quad \mathbf{As} = \mathbf{x} \tag{1}$$

where $\|\cdot\|_0$ stands for the ℓ^0 norm of a vector. Solving the above problem requires a combinatorial search and is generally NP-hard. Then, many algorithms have been proposed to indirectly solve the problem. One of the first and most successful ideas is the idea of basis pursuit (BP) [9], which is to replace the above problem by

$$(P_1): \quad \text{Minimize } \|\mathbf{s}\|_1 \quad \text{subject to} \quad \mathbf{As} = \mathbf{x} \tag{2}$$

where $\|\mathbf{s}\|_1 \triangleq \sum_i |s_i|$ is the ℓ^1 norm of $\mathbf{s}$. Note that the problem P_1 is convex and can be easily solved by using linear programming (LP) techniques. Moreover, it has been shown that if the sparsest solution $\mathbf{s}_0$ is highly sparse, then the solution of P_1 is also the sparsest solution, i.e., it is also the solution of P_0.

To express this property more precisely, let the columns of $\mathbf{A}$ be normalized to have unit ℓ^2 (Euclidean) norm. Let also define the "*coherence*" M of the dictionary $\mathbf{A}$ as the maximum correlation between its atoms, that is

$$M \triangleq \max_{i \neq j} \left| \mathbf{a}_i^T \mathbf{a}_j \right| \tag{3}$$

where $\mathbf{a}_i$, $i = 1, \ldots, m$ denote the columns of $\mathbf{A}$. Then:

Theorem 2 (Equivalence of P_0 and P_1 [16], [17]): If the USLE $\mathbf{As} = \mathbf{x}$ has a solution $\mathbf{s}_0$ for which $\|\mathbf{s}_0\|_0 < (1 + M^{-1}/2)$, then it is the unique solution of both problems P_0 and P_1.

In other words, if the sparsest solution satisfies $\|\mathbf{s}_0\|_0 < (1 + M^{-1}/2)$, it can be found by solving the convex program P_1.

Remark 1: Note that the bound on sparsity that guaranties the equivalence of P_0 and P_1 is highly more restrictive than the bound which guaranties the uniqueness of the sparsest solution. For example, suppose that the dictionary $\mathbf{A}$ is constructed by concatenating two orthonormal bases, $\mathbf{A} = [\mathbf{\Phi}, \mathbf{\Psi}]$, and hence $m = 2n$. It can be easily shown [16] that in this case the maximum possible value for M is $1/\sqrt{n}$ (this maximum value for M is obtained for example for concatenation of a Dirac and a Fourier dictionary). Consider for example such a dictionary $\mathbf{A}$ with $m = 1000$ and $n = 500$, which satisfies the URP and has the maximum possible coherence $M = (1/\sqrt{n}) \approx 1/(22.36)$. Then, by Theorem 1 a solution $\mathbf{s}_0$ with $\|\mathbf{s}_0\|_0 \leq 250$ is necessarily the unique sparsest solution. However, from Theorem 2, it is guaranteed that the sparsest solution can be found by P_1 only where $\|\mathbf{s}_0\|_0 < (1 + 22.36)/2$, that is $\|\mathbf{s}_0\|_0 \leq 11$. In other words, if there is a solution $\mathbf{s}_0$ such that among its 1000 entries there are at most 250 non-zero entries, it would be the unique sparsest solution, but we cannot necessarily find it by solving P_1, unless among these 1000 entries, there are at most 11 non-zero entries. Consequently, equivalence of P_1 and P_0 holds only for the case there exists a *"very very" sparse* solution.

Remark 2: Note also that if the unique sparsest solution satisfies $(1 + M^{-1}/2) \leq \|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$, the above theorem does not state that it "cannot" be found by solving P_1; it simply does not "guarantee" that P_1 can recover it. In fact, from the uniqueness Theorem 1, we know that if we find a solution $\hat{\mathbf{s}}_0$ by using any method (e.g., P_1, or even simply by a magic guess), and we see that it happens that $\|\hat{\mathbf{s}}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$, we will know that we have found the unique sparsest solution.

In addition to the methods based on ℓ^1 norm minimization, there are other ideas for finding the sparsest solution, for example, matching pursuit (MP) [15] and smoothed ℓ^0 (SL0) [18]. The latter method (SL0), which has been designed in our group, tries to directly solve the P_0 problem by replacing the ℓ^0 norm by a smooth approximation of it (and hence the name "smoothed" ℓ^0). One of the motivations behind SL0 is the fact stated above: Since the equivalence of P_0 and P_1 holds only where there exist very very sparse solutions, it would probably be better trying to solve P_0 directly. Another motivation is the speed: it has been shown [18] that SL0 is highly faster than solving P_1.

C. Stability Against Noise?

Suppose that $\mathbf{x}_0$ is a linear combination of a few atoms of the dictionary, that is, $\mathbf{x}_0 = \mathbf{As}_0$, where $\mathbf{s}_0$ is sparse. Now consider a noisy measurement of $\mathbf{x}_0$, that is, $\mathbf{x} = \mathbf{x}_0 + \mathbf{n}$, where $\mathbf{n}$ denotes the noise, and $\|\mathbf{n}\|_2 \leq \varepsilon$. The question of 'stability' [1] is then: Even for a very small ε, is it guaranteed that the sparse decomposition of $\mathbf{x}$ over the dictionary (problem P_0) is not too different from the sparse decomposition of $\mathbf{x}_0$? The answer is unfortunately no, that is, the problem P_0 can be too sensitive to noise [19].

To overcome this problem, it has been proposed in [1] that instead of solving P_0 or P_1 one considers solving their noise aware variants:

$$(P_{0,\delta}): \quad \text{Minimize} \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{As}\|_2 \leq \delta \tag{4}$$

$$(P_{1,\delta}): \quad \text{Minimize} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{As}\|_2 \leq \delta. \tag{5}$$

In other words, it has been proposed to do an "approximate" decomposition, that is, a decomposition with $\|\mathbf{x} - \mathbf{As}\| \leq \delta$ instead of the exact decomposition $\mathbf{x} = \mathbf{As}$. These noise aware variants have to be solved for a sufficiently large δ, that is, for $\delta \geq \varepsilon$ to guarantee that the true solution $\mathbf{s}_0$ satisfies the constraints of the above optimization problems. Then, in [1], the authors prove that both problems $P_{0,\delta}$ and $P_{1,\delta}$ are stable against noise, that is, the estimation error is at worst proportional to the noise level. More precisely, the stability of $P_{0,\delta}$ is given by the following theorem:

Theorem 3 (Stability of $P_{0,\delta}$; Theorem 2.1 of [1]): Let M denote the coherence of the dictionary $\mathbf{A}$. Suppose that for the sparse representation of the noiseless signal $\mathbf{x}_0 = \mathbf{As}_0$ we have

$$k \triangleq \|\mathbf{s}_0\|_0 < \frac{1 + M^{-1}}{2}. \tag{6}$$

If $\hat{\mathbf{s}}_{0,\delta}$ denotes the result of applying $P_{0,\delta}$ on the noisy data $\mathbf{x}$ with $\delta \geq \varepsilon$, then

$$\|\hat{\mathbf{s}}_{0,\delta} - \mathbf{s}_0\|_2 \leq \frac{\varepsilon + \delta}{\sqrt{1 - M(2k - 1)}}. \tag{7}$$

Note that (6) implies also that the term under the square root in (7) is positive.

The authors of [1] also prove the stability of $P_{1,\delta}$ for the case $\|\mathbf{s}_0\|_0 < (1 + M^{-1})/4$.

A noise aware variant of SL0 (called robust-SL0), has already been developed [20], which tries to solve directly $P_{0,\delta}$ without passing through $P_{1,\delta}$.

D. Our Contribution

As it was said in Section II-C, the stability of the problem $P_{0,\delta}$ has only been shown for the case $\|\mathbf{s}_0\|_0 < (1 + M^{-1})/2$. This sparsity limit for stability is the same as the sparsity limit for the equivalence of P_0 and P_1 as stated in Theorem 2. However, as was stated in Remark 1 after Theorem 2, this sparsity limit is highly more restrictive than the sparsity limit for the uniqueness of the sparse solution. In other words, current results state that although a sparse representation with $(1 + M^{-1}/2) \leq \|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$ is unique, it is not guaranteed that $P_{0,\delta}$ can stably recover this representation in presence of noise.

Maybe the lack of this guarantee had not been important before, because, the classic idea for solving P_0 was solving P_1, and the sparsity limit for the equivalence of these two solutions is the same as the sparsity limit for the stability of $P_{0,\delta}$. However, with new algorithms like SL0 or robust-SL0, one can now try to solve $P_{0,\delta}$ directly and without relying on $P_{1,\delta}$. Hence, it is now important to know whether or not sparse representations with $(1 + M^{-1}/2) \leq \|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$ are stable.

In the next section, we will show that $P_{0,\delta}$ is stable for the *whole sparsity range* that guarantees the uniqueness, that is, $P_{0,\delta}$ is stable whenever $\|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$. Moreover, we will show that for smaller $\|\mathbf{s}_0\|_0$ the problem is "more stable," that is, *the more sparsity, the more stability*. Finally, we will show in Section IV that *this stability not only holds for $P_{0,\delta}$, but also holds for any estimation $\hat{\mathbf{s}}_0$ such that $\|\hat{\mathbf{s}}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$ and $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_0\|_2 \leq \delta$*.

III. The Main Theorem

To state the main theorems, we need first to define two notations:

- Let $q = q(\mathbf{A}) = \text{spark}(\mathbf{A}) - 1$. Then, by definition, every q columns of $\mathbf{A}$ are linearly independent, and there is at least a set of $q + 1$ columns which are linearly dependent (in the literature, the quantity q is usually called "Kruskal rank" or "k-rank" of the matrix $\mathbf{A}$). It is also obvious that $q \leq n$, in which, $q = n$ corresponds to the case $\mathbf{A}$ has the URP.
- Let $\sigma_{\min}^{(j)}$, $1 \leq j \leq q(\mathbf{A})$, denote the smallest singular value among all of the submatrices of $\mathbf{A}$ formed by taking j columns of

$\mathbf{A}$. Note that since every q columns of $\mathbf{A}$ are linearly independent, we have $\sigma_{\min}^{(j)} > 0, \forall 1 \leq j \leq q(\mathbf{A})$.

Moreover, it is known [21, p. 419], [22, Lemma 3] that if we add a new column to a full-rank tall matrix, its smallest singular value decreases or remains the same (refer to [22] for a simple direct proof). Therefore, $\sigma_{\min}^{(j)}$ is a decreasing sequence in j, that is

$$\sigma_{\min}^{(j)} \geq \sigma_{\min}^{(j+1)} > 0, \quad \forall 1 \leq j \leq q-1. \tag{8}$$

We are now ready to state the following theorem.

Theorem 4 (Stability of $P_{0,\delta}$): Suppose that the noiseless signal $\mathbf{x}_0$ has a sparse representation $\mathbf{x}_0 = \mathbf{A}\mathbf{s}_0$ satisfying $\|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$. Let also $\mathbf{x} = \mathbf{x}_0 + \mathbf{n}$ be a noisy measurement of $\mathbf{x}_0$ and $\|\mathbf{n}\|_2 \leq \varepsilon$. If $\hat{\mathbf{s}}_{0,\delta}$ denotes the result of applying $P_{0,\delta}$ on the noisy signal $\mathbf{x}$ with $\delta \geq \varepsilon$, then

$$\|\hat{\mathbf{s}}_{0,\delta} - \mathbf{s}_0\|_2 \leq \frac{\delta + \varepsilon}{\sigma_{\min}^{(\ell)}} \tag{9}$$

where $\ell = 2\|\mathbf{s}_0\|_0$.

Remark 1: Theorem 4 shows that $P_{0,\delta}$ is stable not only for $\|\mathbf{s}_0\|_0 < (1 + M^{-1}/2)$, but also for the whole uniqueness range $\|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$. The stability is in the sense that the estimation error increases at worst proportionally to the noise level. Moreover, from (8), the upper bound on estimation error decreases or remains the same as the sparsity increases (this is because sparser $\mathbf{s}_0$ means smaller $\|\mathbf{s}_0\|_0$, which implies smaller ℓ and hence larger or the same $\sigma_{\min}^{(\ell)}$). In other words, *sparser solutions are "more stable."*

Remark 2: The main reason for stating Theorem 4 is to provide a stability result for the case $1 + M^{-1} \leq \ell = 2\|\mathbf{s}_0\|_0 < \text{spark}(\mathbf{A})$, because in this case, Theorem 3 provides no stability result. Moreover, note that for the case $\ell < 1 + M^{-1}$, in which both bounds (7) and (9) are applicable, (9) provides also a tighter bound than (7). This is implied from Lemma 2.2 of [1] which states that in this case $\sigma_{\min}^{(\ell)} > \sqrt{1 - M(\ell - 1)}$.

Proof of Theorem 4: Let define $\hat{\mathbf{x}}_{0,\delta} \triangleq \mathbf{A}\hat{\mathbf{s}}_{0,\delta}$. We write

$$\begin{aligned} \|\mathbf{x}_0 - \hat{\mathbf{x}}_{0,\delta}\|_2 &= \|\mathbf{x} - \mathbf{n} - \hat{\mathbf{x}}_{0,\delta}\|_2 \\ &= \|(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_{0,\delta}) - \mathbf{n}\|_2 \\ &\leq \underbrace{\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_{0,\delta}\|_2}_{\leq \delta} + \underbrace{\|\mathbf{n}\|_2}_{\leq \varepsilon} \\ &\leq \delta + \varepsilon. \end{aligned} \tag{10}$$

On the other hand

$$\mathbf{x}_0 - \hat{\mathbf{x}}_{0,\delta} = \mathbf{A}(\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}) = \mathbf{B}\mathbf{v} \tag{11}$$

where $\mathbf{v}$ is a vector composed of non-zero entries of $\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}$, and $\mathbf{B}$ is a submatrix of $\mathbf{A}$ composed of the columns of $\mathbf{A}$ corresponding to the non-zero entries of $\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}$. Since $\delta \geq \varepsilon$, $\mathbf{s}_0$ satisfies the constraint of the optimization problem $P_{0,\delta}$, and hence $\|\hat{\mathbf{s}}_{0,\delta}\|_0 \leq \|\mathbf{s}_0\|_0$. Therefore, $\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}$ has at most $\ell \triangleq 2\|\mathbf{s}_0\|_0 < \text{spark}(\mathbf{A})$ non-zero entries (note that $\ell < \text{spark}(\mathbf{A})$ means $\ell \leq q(\mathbf{A})$). In other words, $\mathbf{B}$ has at most $\ell \leq q$ columns, and hence (by having also in mind (8))

$$\|\mathbf{B}\mathbf{v}\|_2 \geq \sigma_{\min}^{(\ell)}\|\mathbf{v}\|_2. \tag{12}$$

Noting that $\|\mathbf{v}\|_2 = \|\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}\|_2$, and combining the above inequality with (11), we obtain

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}_{0,\delta}\|_2 \geq \sigma_{\min}^{(\ell)}\|\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}\|_2. \tag{13}$$

Combining (10) and (13) gives

$$\sigma_{\min}^{(\ell)}\|\mathbf{s}_0 - \hat{\mathbf{s}}_{0,\delta}\|_2 \leq \delta + \varepsilon \tag{14}$$

which completes the proof. ■

Remark 3: From (8) and $\ell = 2\|\mathbf{s}_0\|_0 \leq q(\mathbf{A})$, we may replace $\sigma_{\min}^{(j)}$ by its worst case to obtain the following looser bound, which does not need knowing the value of $\|\mathbf{s}_0\|_0$:

$$\|\mathbf{s}_{0,\delta} - \mathbf{s}_0\|_2 \leq \frac{\delta + \varepsilon}{\sigma_{\min}^{(q)}}. \tag{15}$$

IV. A Generalized Stability Theorem

If we carefully re-examine the proof of Theorem 4, we notice that the fact that $\|\hat{\mathbf{s}}_{0,\delta}\|_0 \leq \|\mathbf{s}_0\|_0$ is not essential for obtaining the looser bound (15). Hence, the bound (15) holds not only for the sparse recovery methods based on solving $P_{0,\delta}$, but also for *any other estimation* $\hat{\mathbf{s}}_{0,\delta}$ (obtained from any sparse recovery algorithm or even simply from a magic guess), provided that it satisfies $\|\hat{\mathbf{s}}_{0,\delta}\|_0 < (1/2)\text{spark}(\mathbf{A})$ and $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_{0,\delta}\|_2 \leq \delta$. In other words, *not only $P_{0,\delta}$ is stable, but also any other method for "approximate" sparse representation is stable* provided that it provides a sparse enough estimation. More precisely:

Theorem 5 (Stability of Approximate Sparse Representation): Suppose that the noiseless signal $\mathbf{x}_0$ has a sparse representation $\mathbf{x}_0 = \mathbf{A}\mathbf{s}_0$ satisfying $\|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$. Let also $\mathbf{x} = \mathbf{x}_0 + \mathbf{n}$ be a noisy measurement of $\mathbf{x}_0$ and $\|\mathbf{n}\|_2 \leq \varepsilon$. If we have at hand an estimation $\hat{\mathbf{s}}_{0,\delta}$ of the sparse representation coefficients which satisfies $\|\hat{\mathbf{s}}_{0,\delta}\|_0 < (1/2)\text{spark}(\mathbf{A})$ and $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_{0,\delta}\|_2 \leq \delta$, then

$$\|\hat{\mathbf{s}}_{0,\delta} - \mathbf{s}_0\|_2 \leq \frac{\delta + \varepsilon}{\sigma_{\min}^{(q)}}. \tag{16}$$

Proof: It is easily obtained by following the same steps as the proof of Theorem 4: (10) and (11) still hold. We then note that

$$\|\hat{\mathbf{s}}_{0,\delta} - \mathbf{s}_0\|_2 \leq \|\hat{\mathbf{s}}_{0,\delta}\|_0 + \|\mathbf{s}_0\|_2 < \text{spark}(\mathbf{A}) \tag{17}$$

and hence $\|\hat{\mathbf{s}}_{0,\delta} - \mathbf{s}_0\|_0 \leq q(\mathbf{A})$. Consequently, instead of (12) we write

$$\|\mathbf{B}\|_2 \geq \sigma_{\min}^{(q)}\|\mathbf{v}\|_2 \tag{18}$$

which in combination by (10) and (11) proves (16). ■

Remark: Note that the condition $\delta \geq \varepsilon$ does not explicitly appeared in Theorem 5, and is no more essential (while it was essential in Theorem 4, because it was necessary to insure that $P_{0,\delta}$ gives an estimation satisfying $\|\hat{\mathbf{s}}_{0,\delta}\|_0 \leq \|\mathbf{s}_0\|_0$, which was essential in the proof). However, implicitly, the δ in Theorem 5 cannot be too small, because for a very small δ, it is possible that there exists no $\hat{\mathbf{s}}_{0,\delta}$ satisfying $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}_{0,\delta}\|_2 \leq \delta$.

V. Conclusion

Since minimizing ℓ^1 norm has been one of the first and most successful ideas for finding the sparsest solution of an USLE, some theoretical aspects of the sparsest solution are currently too much influenced by the ℓ^1 minimization idea. Currently, with the availability of the algorithms that try to find the sparse solution by means of other approaches, e.g., SL0 and robust-SL0, some of the properties of the sparsest solution need to be revisited. In this correspondence, we studied the stability of the sparsest solution, and we showed that it is stable not only where $\|\mathbf{s}_0\|_0 < (1+M^{-1})/2$, but also for the whole uniqueness range $\|\mathbf{s}_0\|_0 < (1/2)\text{spark}(\mathbf{A})$. These results prove the practical interest of designing ℓ^0-norm minimization algorithms, since they can provide a

good estimation from noisy data, with the weakest condition of sparsity.

REFERENCES

[1] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–124, Jul. 2007.

[5] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: Principles, perspectives, and new challenges," in *Proc. Eur. Symp. Artificial Neural Networks (ESANN)*, Apr. 2006, pp. 323–330.

[6] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process.*, vol. 81, pp. 2353–2362, 2001.

[7] P. G. Georgiev, F. J. Theis, and A. Cichocki, "Blind source separation and sparse component analysis for over-complete mixtures," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, pp. 493–496.

[8] Y. Li, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *Proc. Int. Communication Assoc. Conf. (ICA)*, 2003, pp. 89–94.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[10] E. J. Candè and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 15, no. 12, pp. 4203–4215, 2005.

[11] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, 2003.

[12] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Aug. 2005, pp. II-782–II-785.

[13] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Trans. Image Process.*, vol. 52, no. 12, pp. 5559–5569, 2006.

[14] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS, a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[15] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[16] R. Gribonval and M. Nielsen, "Sparse decompositions in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[17] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[18] H. Mohimani, M. Babaie-Zadeh, and Ch. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.

[19] B. Wohlberg, "Noise sensitivity of sparse signal representations: Reconstruction error bounds for the inverse problem," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3053–3060, Dec. 2003.

[20] A. Eftekhari, M. Babaie-Zadeh, Ch. Jutten, and H. Abrishami-Moghaddam, "Robust-SL0 for stable sparse representation in noisy settings," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 3433–3436.

[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[22] M. Babaie-Zadeh, H. Mohimani, and Ch. Jutten, "An upper bound on the estimation error of the sparsest solution of underdetermined linear systems," presented at the Workshop Signal Processing With Adaptive Sparse Structured Representations (SPARS), Saint-Malo, France, Apr. 6–9, 2009.