An upper bound on the estimation error of the sparsest solution of underdetermined linear systems

Massoud Babaie-Zadeh, G. Hosein Mohimani, Christian Jutten

An upper bound on the estimation error of the sparsest solution of underdetermined linear systems

Massoud Babaie-Zadeh,
Department of Electrical Engineering
Sharif University of Technology
Tehran, IRAN
mbzadeh@yahoo.com

Hosein Mohimani
Department of Electrical Engineering
University of California, San Diego
USA
hmohiman@ucsd.edu

Christian Jutten
GIPSA-LAB
Grenoble, France
Christian.Jutten@inpg.fr

***Abstract*—Let $\mathbf{A}$ be an $n \times m$ matrix with $m > n$, and suppose the underdetermined linear system $\mathbf{As} = \mathbf{x}$ admits a unique sparse solution $\mathbf{s}_0$ (i.e. it has a solution $\mathbf{s}_0$ for which $\|\mathbf{s}_0\|_0 < \frac{1}{2}spark(\mathbf{A})$). Suppose that we have somehow a solution (sparse or non-sparse) $\hat{\mathbf{s}}$ of this system as an estimation of the true sparsest solution $\mathbf{s}_0$. Is it possible to construct an upper bound on the estimation error $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ without knowing $\mathbf{s}_0$? The answer is positive, and in this paper we construct such a bound which, in the case $\mathbf{A}$ has Unique Representation Property (URP), depends on the smallest singular value of all $n \times n$ submatrices of $\mathbf{A}$.**

I. Introduction and problem statement

Let $\mathbf{A}$ be an $n \times m$ matrix with $m > n$, and consider the Underdetermined System of Linear Equations (USLE) $\mathbf{As} = \mathbf{x}$. It has been shown [1], [2], [3] that if this linear system has a sparse enough solution, it would be its unique sparsest solution. More precisely:

Theorem 1 (Uniqueness Theorem [2], [3])**.** *Let 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 $\|\mathbf{s}_0\|_0 < \frac{1}{2}spark(\mathbf{A})$, it is its unique sparsest solution.*

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

Finding the sparsest solution of underdetermined linear systems is of significant importance in signal processing and statistics. It is used, for example, in compressed sensing [4], [5], [6], underdetermined Sparse Component Analysis (SCA) and source separation [7], [8], [9], [10], atomic decomposition on overcomplete dictionaries [11], [12], decoding real field codes [13], image deconvolution [14], [15], image denoising [16], electromagnetic imaging and Direction of Arrival (DOA) finding [1], etc. From the atomic decomposition viewpoint [17], the columns of $\mathbf{A}$ are called 'atoms' and the matrix $\mathbf{A}$ is called the 'dictionary' over which the signal $\mathbf{x}$ is to be decomposed.

This work has been partially funded by Iran National Science Foundation (INSF) under contract number 86/994, and also by ISMO and French embassy in Tehran in the framework of a GundiShapour collaboration program.

Finding the sparsest solution ($\mathbf{s}_0$) requires a combinatorial search and is generally NP-hard. Then, many different algorithms have been proposed to find an estimation of $\mathbf{s}_0$, for example, Basis Pursuit (BP) [11], Matching Pursuit (MP) [17], Smoothed L0 (SL0) [18], [19], etc. The solution $\hat{\mathbf{s}}$ found by some of these algorithms, *e.g.* BP, is sparse in the exact sense (that is, many of its components are exactly zero), while for some others, *e.g.* MP and SL0, $\hat{\mathbf{s}}$ is only approximately sparse (that is, many of its components have very small magnitudes but are not exactly zero).

Suppose now that by using any algorithm (or simply by a magic guess) we have found a solution $\hat{\mathbf{s}}$ of $\mathbf{As} = \mathbf{x}$, as an estimation of the true sparsest solution ($\mathbf{s}_0$). The question now is: "*Noting that $\mathbf{s}_0$ is unknown, is it possible to construct an upper bound on the estimation error $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ only from $\hat{\mathbf{s}}$*"? For example, if $\mathbf{A}$ has the URP, and $\hat{\mathbf{s}}$ has less than or equal $\lfloor n/2 \rfloor$ non-zero components, where $\lfloor x \rfloor$ stands for the largest integer smaller than or equal x, then the uniqueness theorem insures that $\hat{\mathbf{s}} = \mathbf{s}_0$. On the other hand, if all the components of $\hat{\mathbf{s}}$ are non-zero but its $(\lfloor n/2 \rfloor + 1)$'th largest magnitude component is very small, heuristically we expect to be close to the true solution $\mathbf{s}_0$, but the uniqueness theorem says nothing about this heuristic.

In this paper, we will see that the answer to the above question is positive, and we present an upper bound on $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ without knowing $\mathbf{s}_0$, which (in the case $\mathbf{A}$ has the URP) depends on the magnitude of the $(\lfloor n/2 \rfloor + 1)$'th largest component of $\hat{\mathbf{s}}$, and also to the matrix $\mathbf{A}$. Moreover, we will see that if the $(\lfloor n/2 \rfloor + 1)$'th largest component of $\hat{\mathbf{s}}$ is zero, then our upper bound vanishes, and hence $\hat{\mathbf{s}} = \mathbf{s}_0$. This is, in fact, the same result provided by the uniqueness theorem, and hence our upper bound can be seen as a generalization of the uniqueness theorem. In other words, from the classical uniqueness theorem, all we know is that if among m components of $\hat{\mathbf{s}}$, $m - \lfloor n/2 \rfloor$ components are '*exactly*' zero, then $\hat{\mathbf{s}} = \mathbf{s}_0$, but *if $\hat{\mathbf{s}}$ has more than $\lfloor n/2 \rfloor$ non-zero components (even if $m - \lfloor n/2 \rfloor$ of its components have very very small magnitudes) we are not sure to be close to the true solution*. However, our upper bound insures that in the second case, too, we are not far from the true solution. Moreover, our upper bound depends also on $\mathbf{A}$ which provides some explanations about the sensitivity of the error to the

properties of the matrix $\mathbf{A}$.

A related problem has already been addressed in [20], in which, for the noisy case $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e}$, an upper bound has been constructed for the error $\|\hat{\mathbf{s}} - \mathbf{s}_0\|$. However, in that paper it has been implicitly assumed that $\hat{\mathbf{s}}$ is sparse in the exact sense, that is, $\|\hat{\mathbf{s}}\|_0 \leq \lfloor n/2 \rfloor$, otherwise, the upper bound will go to infinity. On the other hand, if the noise power ($\|\mathbf{e}\|$) is set equal to zero, the upper bound of [20] for $\|\hat{\mathbf{s}} - \mathbf{s}_0\|$ vanishes, resulting again to the uniqueness theorem. In other words, [20] can be seen somehow as a generalization of the uniqueness theorem to the noisy case, while our paper can be seen as a generalization of the uniqueness theorem to the case $\hat{\mathbf{s}}$ is not sparse in the exact sense, but is approximately sparse.

The paper is organized as follows. In Section II we review a result already provided by Mohimani *et. al.* [19] during their analysis to the convergence of the SL0 algorithm. Then in Section III, we present a bound based on singular values of the submatrices of the dictionary by direct manipulating the result of Mohimani *et. al*. This bound will then be improved in Section IV. Finally, we discuss the case of random dictionaries in Section V.

II. The main idea

The main idea has been given by Mohimani *et. al.* (Corollary 1 of Lemma 1 of [19]). We re-state that result here (with a few changes in notations).

For the $n \times m$ matrix $\mathbf{A}$, let $\mathcal{P}_\ell(\mathbf{A})$ denote the set of all of its $n \times \ell$ submatrices, that is, submatrices of $\mathbf{A}$ which have been obtained by taking ℓ columns of $\mathbf{A}$ ($\ell \leq m$). Moreover, let:

$$\mathcal{M}_n(\mathbf{A}) = \mathcal{P}_1(\mathbf{A}) \cup \mathcal{P}_1(\mathbf{A}) \cup \cdots \cup \mathcal{P}_n(\mathbf{A}) \tag{1}$$

In other words, $\mathcal{M}_n$ is composed of all matrices that are formed by taking 1 or 2 $\ldots$ or n columns of $\mathbf{A}$. The number of these submatrices (the cardinality of $\mathcal{M}_n$, denoted by $|\mathcal{M}_n|$) is equal to:

$$|\mathcal{M}_n| = \binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{n}. \tag{2}$$

Define now:

$$G_{\mathbf{A}} \triangleq \max_{\mathbf{B} \in \mathcal{M}_n(\mathbf{A})} \|\mathbf{B}^\dagger\|_F, \tag{3}$$

where $\mathbf{B}^\dagger$ stands for the Moore-Penrose pseudo-inverse of $\mathbf{B}$, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. The constant $G_{\mathbf{A}}$ depends only on the dictionary matrix $\mathbf{A}$. Let also that for a vector $\mathbf{y}$ and a positive scalar α, $\|\mathbf{y}\|_{0,\alpha}$ denote the number of components of $\mathbf{y}$ which have magnitudes larger than α. In other words, $\|\mathbf{y}\|_{0,\alpha}$ denotes the ℓ^0 norm of a thresholded version of $\mathbf{y}$ (in which the components with magnitudes smaller than α are clipped to zeros).

The Corollary 1 of Lemma 1 of [19] states then:

Corollary 1 (of [19])**.** *Let $\mathbf{A}$ be an $n \times m$ matrix with unit ℓ^2 norm columns which satisfies the URP and let $\boldsymbol{\delta} \in \mathrm{null}(\mathbf{A})$. If for an $\alpha > 0$, $\boldsymbol{\delta}$ has at most n components with absolute values greater than α (that is, if $\|\boldsymbol{\delta}\|_{0,\alpha} \leq n$), then:*

$$\|\boldsymbol{\delta}\| < (G_{\mathbf{A}} + 1) m \alpha. \tag{4}$$

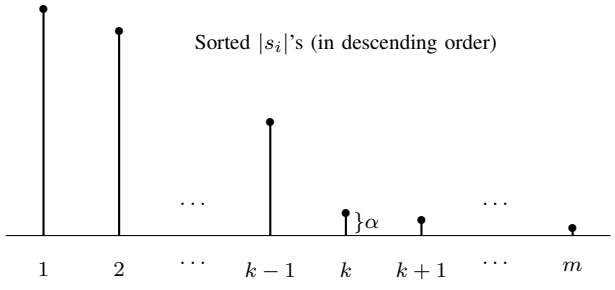

Fig. 1. The definition of $h(k, \mathbf{s})$: Sort the magnitudes of the components of $\mathbf{s}$ in descending order. Then, $h(k, \mathbf{s})$ is defined as the magnitude of the k's element (denoted by α in the figure).

Let now define the notation:

Definition 1. *Let $\mathbf{s}$ be a vector of length m. Then $h(k, \mathbf{s})$ denotes the magnitude of the k'th largest magnitude component of $\mathbf{s}$.*

In other words, $h(k, \mathbf{s})$ is calculated as follows: sort the magnitudes of the components of $\mathbf{s}$ in descending order, and return the magnitude of the k'th element. Figure 1 illustrates this definition.

Then, using the above corollary, Remark 5 of Theorem 1 of [19] states the following idea to construct an upper bound on $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$: Let $\alpha_{\hat{\mathbf{s}},n} \triangleq h(\lfloor \frac{n}{2} \rfloor, \hat{\mathbf{s}})$. Since the true sparsest solution ($\mathbf{s}_0$) has at most $\lfloor \frac{n}{2} \rfloor$ non-zero components, $\hat{\mathbf{s}} - \mathbf{s}_0$ has at most n components with absolute values greater than $\alpha_{\hat{\mathbf{s}},n}$, that is, $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_{0,\alpha_{\hat{\mathbf{s}},n}} \leq n$. Moreover, $(\hat{\mathbf{s}} - \mathbf{s}_0) \in \mathrm{null}(\mathbf{A})$ and hence the above Corollary implies that:

$$\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 \leq (G_{\mathbf{A}} + 1) m \alpha_{\hat{\mathbf{s}},n} \tag{5}$$

This result is consistent with the heuristic stated in the introduction: "if $\hat{\mathbf{s}}$ has at most $n/2$ 'large' components, the uniqueness of the sparsest solution insures that $\hat{\mathbf{s}}$ is close to the true solution".

III. A bound based on minimal singular values

The bound (5) is not easy to be analyzed and worked with. The dependence of the bound on the dictionary (through the constant $G_{\mathbf{A}}$) is very complicated. Moreover, calculating the $G_{\mathbf{A}}$ constant for a dictionary requires calculation of the pseudo-inverses of all of the $\binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{n}$ submatrices of $\mathbf{A}$. In this section, we modify (5) to obtain a bound that is easier to be analyzed and (in a statistical point of view) its dependence to (the statistics of) $\mathbf{A}$ is simpler.

To state the main theorem of this section, let define the notations:

- For an arbitrary matrix $\mathbf{M}$, $s_{\min}(\mathbf{M})$ denotes its smallest singular value.
- For an arbitrary matrix $\mathbf{M}$, $\sigma_{\min}(\mathbf{M})$ or $\sigma_{\min,\mathbf{M}}$ is defined as:

$$\sigma_{\min}(\mathbf{M}) = \sigma_{\min,\mathbf{M}} \triangleq \min_{\mathbf{x}} \frac{\|\mathbf{M}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \tag{6}$$

Note that if $\mathbf{M}$ has full column rank then $\sigma_{\min}(\mathbf{M}) = s_{\min}(\mathbf{M})$, otherwise $\sigma_{\min}(\mathbf{M}) = 0$. Hence, for the case $\mathbf{M}$ is tall and full rank, we use $\sigma_{\min}$ and $s_{\min}$ interchangeably.

- For an arbitrary matrix $\mathbf{M}$, $\sigma^{(\ell)}_{\min}(\mathbf{M})$ or $\sigma^{(\ell)}_{\min,\mathbf{M}}$ denotes the $\sigma_{\min}$ (defined above) of all submatrices of $\mathbf{M}$ that have been obtained by taking ℓ columns of $\mathbf{M}$. In other words, if $\mathbf{M}$ is $n \times m$ and $\mathcal{P}_\ell(\mathbf{M})$ denotes the set of all $n \times \ell$ submatrices ($\ell \leq m$) of $\mathbf{M}$, then:

$$\sigma^{(\ell)}_{\min}(\mathbf{M}) = \min_{\mathbf{N} \in \mathcal{P}_\ell(\mathbf{M})} \{\sigma_{\min}(\mathbf{N})\} = \min_{\|\mathbf{x}\|_0 \leq \ell} \frac{\|\mathbf{M}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}. \tag{7}$$

We can now state the following theorem:

Theorem 2. *Let $\mathbf{A}$ be an $n \times m$ matrix with unit ℓ^2 norm columns which satisfies the URP and let $\mathbf{s}_0$ be a solution of $\mathbf{A}\mathbf{s} = \mathbf{x}$ for which $\|\mathbf{s}_0\|_0 \leq \frac{n}{2}$. Suppose that $\hat{\mathbf{s}}$ is a solution of $\mathbf{A}\mathbf{s} = \mathbf{x}$, and $\alpha_{\hat{\mathbf{s}},n}$ is as defined in the previous section. Then:*

$$\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 \leq \left(\frac{\sqrt{n}}{\sigma^{(n)}_{\min,\mathbf{A}}} + 1 \right) m\, \alpha_{\hat{\mathbf{s}},n} \tag{8}$$

Before going to the proof, let state a few remarks on the consequences of the above theorem.

Remark 1. If $\|\hat{\mathbf{s}}\|_0 \leq \frac{n}{2}$, then $\alpha_{\hat{\mathbf{s}},n} = 0$, the upper bound (8) vanishes, and hence $\hat{\mathbf{s}} = \mathbf{s}$. In other words, the above theorem implies that a solution with $\|\mathbf{s}\|_0 \leq \frac{n}{2}$ is unique, that is, the above theorem implies the uniqueness theorem (for the URP case).

Remark 2. If the URP does not hold for $\mathbf{A}$, it means that there exists an $n \times n$ submatrix of $\mathbf{A}$ which is not invertible. This implies that $\sigma^{(n)}_{\min,\mathbf{A}} = 0$, and hence the upper bound (8) goes to infinity. In other words, when the URP does not hold, the uniqueness of a solution with $\|\mathbf{s}\|_0 \leq \frac{n}{2}$ does not hold.

Remark 3. If the URP holds, and if $\hat{\mathbf{s}}$ is only approximately sparse, that is, if $m - \lfloor \frac{n}{2} \rfloor$ components of it are of very small magnitudes, then $\alpha_{\hat{\mathbf{s}},n}$ is small, and the bound (8) states that we are probably (depending on the matrix $\mathbf{A}$) close to the true solution. Moreover, in this case, $\sigma^{(n)}_{\min,\mathbf{A}}$ determines some kind of sensitivity to the dictionary: if the URP holds but there exists a square submatrix of $\mathbf{A}$ which is ill-conditioned, then $\sigma^{(n)}_{\min,\mathbf{A}}$ is very small and hence for achieving a predetermined accuracy, $\alpha_{\hat{\mathbf{s}},n}$ should be very small, that is, the sparsity of $\hat{\mathbf{s}}$ should be held with a better approximation.

To prove Theorem 2, we state first a few lemmas. The first lemma is very well-known in matrix theory[1]:

[1]It is also very easy to be proved: Let $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ be the Singular Value Decomposition (SVD) of $\mathbf{M}$. Then since $\mathbf{U}$ and $\mathbf{V}$ are rotation matrices, it is easy to see $\|\mathbf{U}\mathbf{C}\|_F = \|\mathbf{C}\|_F$, and $\|\mathbf{C}\mathbf{V}\|_F = \|\mathbf{C}\|_F$, for any matrix $\mathbf{C}$ with appropriate dimensions. For example, if $\mathbf{D} = \mathbf{U}\mathbf{C}$, then: $\|\mathbf{D}\|_F^2 = \sum_i \|\mathbf{d}_i\|_2^2 = \sum_i \|\mathbf{U}\mathbf{c}\|_2^2 = \sum_i \|\mathbf{c}\|_2^2 = \|\mathbf{C}\|_F^2$, where $\mathbf{d}_i$ and $\mathbf{c}_i$ denote the columns of $\mathbf{D}$ and $\mathbf{C}$, respectively. Consequently, $\|\mathbf{M}\|_F = \|\mathbf{U}\mathbf{S}\mathbf{V}^H\|_F = \|\mathbf{S}\|_F$.

Lemma 1. *Let $\sigma_1, \sigma_2, \ldots, \sigma_r$ are the singular values of a matrix $\mathbf{M}$. Then:*

$$\|\mathbf{M}\|_F^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2 \tag{9}$$

$$\|\mathbf{M}^\dagger\|_F^2 = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_r^2}. \tag{10}$$

Lemma 2. *Let $\mathbf{A}$ be an $n \times m$ matrix and assume that any of its ℓ columns, $\ell \leq n$, are linearly independent. Let also $\mathbf{B}$ be an $n \times \ell$ submatrix of $\mathbf{A}$. Then:*

$$\|\mathbf{B}^\dagger\|_F^2 \leq \frac{\ell}{\sigma^{(\ell)}_{\min}(\mathbf{A})}. \tag{11}$$

Proof: Since the tall matrix $\mathbf{B}$ is full-rank, it has ℓ singular values, say $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_\ell > 0$. Hence, from (10):

$$\begin{aligned} \|\mathbf{B}^\dagger\|_F^2 &= \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \cdots + \frac{1}{\sigma_\ell^2} \\ &\leq \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\ell^2} + \cdots + \frac{1}{\sigma_\ell^2} \\ &= \frac{\ell}{\sigma_\ell^2}. \end{aligned} \tag{12}$$

From the definition of $\sigma^{(\ell)}_{\min}(\mathbf{A})$, $\sigma_\ell \geq \sigma^{(\ell)}_{\min}(\mathbf{A})$, and hence the above inequality results in (11). ∎

Now, we state the following two lemmas. Although Lemma 3 has already been known [21, p. 419], we prefer to present another proof in this paper. However, the proofs of both of these lemmas are left to the appendix.

Lemma 3. *Let $\mathbf{B}$ be an $n \times \ell$ matrix with $\ell < n$ (tall matrix). If we add a new column $\mathbf{b}$ to it to obtain $\mathbf{B}' = [\mathbf{B}, \mathbf{b}]$, then $s_{\min}(\mathbf{B}') \leq s_{\min}(\mathbf{B})$.*

Lemma 4. *Let $\mathbf{B}$ be an $n \times \ell$ matrix with $\ell \geq n$ (square or wide matrix). If we add a new column $\mathbf{b}$ to it to obtain the wide matrix $\mathbf{B}' = [\mathbf{B}, \mathbf{b}]$, then $s_{\min}(\mathbf{B}') \geq s_{\min}(\mathbf{B})$.*

Proof of Theorem 2: Let us form a submatrix $\mathbf{B}$ of $\mathbf{A}$ by taking one column of $\mathbf{A}$ and then sequentially append other columns of $\mathbf{A}$ to $\mathbf{B}$. The above lemmas show that the smallest singular value of $\mathbf{B}$ decreases until $\mathbf{B}$ becomes a square matrix, and then it increases. While $\ell < n$, by adding new columns of $\mathbf{A}$ to $\mathbf{B}$, the upper bound $\frac{\ell}{\sigma_\ell^2}$ of (12) increases (because its nominator increases and its denominator decreases) until $\frac{n}{\sigma_n^2}$. Therefore, since $\sigma^{(n)}_{\min}(\mathbf{A})$ is the smallest singular value among all the smallest singular values of all $n \times n$ submatrices of $\mathbf{A}$, then:

$$\forall \mathbf{B} \in \mathcal{M}_n, \quad \|\mathbf{B}^\dagger\|_F^2 \leq \frac{n}{(\sigma^{(n)}_{\min}(\mathbf{A}))^2} \tag{13}$$

and hence $G_{\mathbf{A}} \leq \frac{\sqrt{n}}{\sigma^{(n)}_{\min}(\mathbf{A})}$. Combining this with (5) proves (8). ∎

Remark 1. The discussions after Theorem 2 show that $\sigma^{(n)}_{\min}(\mathbf{A})$ is an important parameter of a dictionary (which has the URP). The calculation of $\sigma^{(n)}_{\min}(\mathbf{A})$ requires examination of all $\binom{m}{n}$ submatrices of $\mathbf{A}$. Although this should be done only once for a dictionary $\mathbf{A}$, it is still very time consuming and is probably NP-hard.

Remark 2. Since calculation of $\sigma^{(n)}_{\min}(\mathbf{A})$ is not easy for a dictionary, if we can find a lower bound for $\sigma^{(n)}_{\min}(\mathbf{A})$ that can be calculated easily, then we would have an upper bound which can be calculated easier. Note also that the minimum singular value of the whole dictionary $\mathbf{A}$ says nothing about $\sigma^{(n)}_{\min}(\mathbf{A})$, because from Lemma 4, if after a square $\mathbf{B}$ we continue adding columns from $\mathbf{A}$ to $\mathbf{B}$, its minimum singular value increases.

Moreover, in Lemma 2.2 of [12] it has been shown that if $\mathbf{B}_{n\times\ell}$ is a submatrix of $\mathbf{A}$, with $\ell \leq n$, then $\sigma^2_{\min}(\mathbf{B}) \geq 1 - M(\ell-1)$, where M denotes the mutual coherence of $\mathbf{A}$, *i.e.* the maximum of absolute values of correlations between the columns of $\mathbf{A}$. This result seems useless for our problem ($\ell = n$), because for example for the case of concatenation of two orthonormal bases, it has been shown that $M \geq \frac{1}{\sqrt{n}}$, and hence for $\ell = n$ the right side of this inequality is negative.

IV. Modifying the bound

In this section, we modify the bound (8) in two senses: Firstly we show that the bound can tightened by dropping the constant $\sqrt{n}$. Secondly, we restate the theorem in a form that covers the case *spark*$(\mathbf{A}) < n+1$, that is, the case the URP does not hold for $\mathbf{A}$.

Theorem 3. *Let $\mathbf{A}$ be an $n \times m$ matrix with unit ℓ^2 norm columns, and assume that any $n \times q$ submatrix of $\mathbf{A}$ has full column rank ($q \leq n$). Let also that $\mathbf{s}_0$ be a solution of $\mathbf{A}\mathbf{s} = \mathbf{x}$ for which $\|\mathbf{s}_0\|_0 \leq \frac{q}{2}$. Suppose that $\hat{\mathbf{s}}$ is a solution of $\mathbf{A}\mathbf{s} = \mathbf{x}$, and set $\alpha_{\hat{\mathbf{s}},q} = h(\lfloor \frac{q}{2} \rfloor, \hat{\mathbf{s}})$. Then:*

$$\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 < \left(\frac{1}{\sigma^{(q)}_{\min,\mathbf{A}}} + 1 \right) m\, \alpha_{\hat{\mathbf{s}},q} \tag{14}$$

To prove this theorem, we first state a modified version of (4):

Lemma 5. *Let $\mathbf{A}$ be an $n \times m$ matrix with unit ℓ^2 norm columns and assume that any $n \times q$ submatrix of $\mathbf{A}$ has full column rank, and let $\boldsymbol{\delta} \in \mathrm{null}(\mathbf{A})$. If for an $\alpha > 0$, $\|\boldsymbol{\delta}\|_{0,\alpha} \leq q$, then:*

$$\|\boldsymbol{\delta}\|_2 < \left(\frac{1}{\sigma^{(q)}_{\min,\mathbf{A}}} + 1 \right) m\alpha. \tag{15}$$

The proof of this lemma is left to Appendix.

Proof of Theorem 3: $\mathbf{s}_0$ has at most $\lfloor \frac{q}{2} \rfloor$ non-zero components and $\hat{\mathbf{s}}$ has at most $\lfloor \frac{q}{2} \rfloor$ components with magnitudes larger than α. Therefore, $\hat{\mathbf{s}} - \mathbf{s}_0$ has at most q components with magnitudes larger than α. Moreover, $(\hat{\mathbf{s}} - \mathbf{s}_0) \in \mathrm{null}(\mathbf{A})$. Hence, the conditions of Lemma 5 hold for $\boldsymbol{\delta} = \hat{\mathbf{s}} - \mathbf{s}_0$ and $\alpha = \alpha_{\hat{\mathbf{s}},q}$, which proves the theorem. ∎

Remark 1. Similar to the reasoning after Theorem 2, it can be seen that the above theorem implies that a solution with $\|\mathbf{s}\|_0 \leq \frac{q}{2}$ is unique. The largest value of q for which any $n \times q$ submatrix of $\mathbf{A}$ has full column rank is exactly what is defined as *spark*$(\mathbf{A}) - 1$. Consequently, a solution with $\|\mathbf{s}\|_0 \leq \frac{1}{2}$*spark*$(\mathbf{A})$ is unique (which is again the uniqueness theorem).

Remark 2. Having in mind Remark 3 after Theorem 2 about the sensitivity, and from Lemma 3, it can be seen that (15) states also some kind of 'sensitivity' to the degree of sparseness of the sparsest solution ($\mathbf{s}_0$). Let $p \triangleq \|\mathbf{s}_0\|_0$, and set $q = 2p$, and suppose that $q \leq$ *spark*$(\mathbf{A}) - 1$ (that is, any $n \times q$ submatrix of $\mathbf{A}$ has full column rank). Then the conditions of Theorem 3 have been satisfied and hence:

$$\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 \leq \left(\frac{1}{\sigma^{(2p)}_{\min,\mathbf{A}}} + 1 \right) m\, \alpha_{\hat{\mathbf{s}},2p} \tag{16}$$

In other words, whenever $\mathbf{s}_0$ is sparser, p is smaller, hence $\sigma^{(2p)}_{\min,\mathbf{A}}$ is larger, and therefore a larger $\alpha_{\hat{\mathbf{s}},2p}$ is tolerable (that is, we have less sensitivity to exact sparseness of $\hat{\mathbf{s}}$) .

V. Random dictionaries

As stated in the remarks of Section III, $\sigma^{(n)}_{\min}(\mathbf{A})$ (or after (14) and (16), $\sigma^{(q)}_{\min}(\mathbf{A})$ for a $q \leq n$) is an important parameter of a dictionary. However, estimating $\sigma^{(q)}_{\min}(\mathbf{A})$ for a deterministic matrix is probably NP-hard and intractable for large dictionaries (although it has to be computed only once for a given dictionary). This is because it seems that for calculating $\sigma^{(q)}_{\min}(\mathbf{A})$ we would need to examine all $n \times q$ submatrices of $\mathbf{A}$ which is a combinatorial and intractable task. Even finding a computationally tractable lower bound for $\sigma^{(q)}_{\min}(\mathbf{A})$ would provide us a computable upper bound for the error.

On the contrary, for random matrices, there are already many works for finding a lower bound on their least singular values [22], [23]. Moreover, for a random $\mathbf{A}$ with independently and identically distributed (iid) entries, we need no more to examine all of its $\binom{n}{q}$ submatrices, because all $n \times q$ submatrices are statistically identical. Random dictionaries are also practically important, because they are frequently used in compressed sensing [6]. We note also that random matrices have the URP with probability 1.

In random matrix theory, it is easier [24] to bound below the smallest singular value of a random $n \times q$ matrix for the case $q/n = c < 1$ than the case $q = n$. A famous result by Marčhenko and Pastur [23] states that if the entries of $\mathbf{B}_{n\times q}$, $q < n$ are iid and Gaussian (or any other distribution with fourth order moment of order $O(\frac{1}{n^2})$) with mean zero and variance[2] $\frac{1}{n}$, as $n, q \to \infty$ and $\frac{q}{n} \to c < 1$, the empirical distribution of singular values of $\mathbf{B}$ converges almost surely to a distribution bounded between $1-\sqrt{c}$ and $1+\sqrt{c}$. Moreover, it has been shown (in [25] for the Gaussian case and in [26] for iid entries with finite fourth moment) that almost surely its minimum singular value converges to $1-\sqrt{c}$. Another result from Davidson and Szarek [27, Eq. (4.36)], [28] states that for the Gaussian case:

$$\mathbb{P}\left\{ \sigma_{\min} < 1 - \sqrt{\frac{q}{n}} - r \right\} \leq e^{-nr^2/2}, \tag{17}$$

where $\mathbb{P}\{\cdot\}$ stands for the probability of the event $\{\cdot\}$.

[2]Note that the condition on variance is satisfied if the entries of $\mathbf{B}$ are zero-mean and its columns are of unit ℓ^2 norm.

For square random matrices, it has been shown that if the entries of an $n \times n$ random matrix are iid Gaussian with zero-mean and variance $1/n$, where n is large enough, then its smallest singular value satisfies [23, Theorem 2.36] [22, Theorem 5.1], [29]:

$$\mathbb{P}\{\sigma_{\min} \leq \frac{r}{n}\} = 1 - e^{-r-r^2/2} \tag{18}$$

In fact, even an n of order 4 or 5 is considered large enough for the validity of the above formula [22, Fig. 5.2]. From the Taylor expansion $1 - e^{-r-r^2/2} = r - r^3/3 + O(r^4)$ the above formula states that for small ϵ's [22, Lemma 7.4]:

$$\mathbb{P}\{\sigma_{\min} \leq \frac{\epsilon}{n}\} \approx \epsilon \tag{19}$$

These results can provide us stochastic upper bounds for $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$, as stated below.

Theorem 4. *If there exists a solution $\mathbf{s}_0$ for $\mathbf{As} = \mathbf{x}$ such that $\|\mathbf{s}_0\|_0 \leq \lfloor n/2 \rfloor$, then for any other solution $\hat{\mathbf{s}}$:*

$$\mathbb{P}\left\{\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 < (\frac{n}{\epsilon} + 1) m \alpha_{\hat{\mathbf{s}},n}\right\} \geq 1 - \binom{m}{q}(1 - e^{-\epsilon-\epsilon^2/2}) \tag{20}$$

Note again that for small ϵ's, $1 - e^{-\epsilon-\epsilon^2/2} \approx \epsilon$.

Proof: From (15), for $q = n$, if $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 \geq (n/\epsilon + 1) m \alpha_{\hat{\mathbf{s}},n}$ then $(1/\sigma^{(n)}_{\min,\mathbf{A}} + 1) m \alpha_{\hat{\mathbf{s}},n} \geq (n/\epsilon + 1) m \alpha_{\hat{\mathbf{s}},n}$, and hence:

$$\begin{aligned}
&\mathbb{P}\Big\{\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 \geq (\frac{n}{\epsilon} + 1) m \alpha_{\hat{\mathbf{s}},n}\Big\} \\
&\quad \leq \mathbb{P}\left\{\left(\frac{1}{\sigma^{(n)}_{\min,\mathbf{A}}} + 1\right) m \alpha_{\hat{\mathbf{s}},n} \geq (\frac{n}{\epsilon} + 1) m \alpha_{\hat{\mathbf{s}},n}\right\} \\
&\quad = \mathbb{P}\left\{\sigma^{(n)}_{\min,\mathbf{A}} \leq \frac{\epsilon}{n}\right\} = \mathbb{P}\left\{\bigcup_{\mathbf{B} \in \mathcal{P}_n(\mathbf{A})} \sigma_{\min}(\mathbf{B}) \leq \frac{\epsilon}{n}\right\} \\
&\quad \leq \sum_{\mathbf{B} \in \mathcal{P}_n(\mathbf{A})} \mathbb{P}\left\{\sigma_{\min}(\mathbf{B}) \leq \frac{\epsilon}{n}\right\} = \binom{m}{q}(1 - e^{-\epsilon-\epsilon^2/2})
\end{aligned}$$

■

Theorem 5. *If there exists a solution $\mathbf{s}_0$ for $\mathbf{As} = \mathbf{x}$ such that $\|\mathbf{s}_0\|_0 \leq \lfloor cn/2 \rfloor$ for a fixed $0 < c < 1$ (such that cn is an integer less than n), then for any other solution $\hat{\mathbf{s}}$:*

$$\begin{aligned}
\mathbb{P}\Big\{\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 < (\frac{1}{1 - \sqrt{c} - r} + 1) m \alpha_{\hat{\mathbf{s}},(cn)}\Big\} \\
\geq 1 - \binom{m}{cn} e^{-nr^2/2}
\end{aligned} \tag{21}$$

Proof: Is similar to the proof of the previous theorem using (14) for $q = cn$. ■

For example, by taking $c = 0.5$ and $r = 0.2$, the above theorem states that if $\mathbf{s}_0$ is sparser than $n/4$, that is, if $\|\mathbf{s}_0\|_0 \leq \lfloor n/4 \rfloor$, then:

$$\mathbb{P}\left\{\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2 < (11.765)\, m\, \alpha_{\hat{\mathbf{s}},(n/2)}\right\} \geq 1 - \binom{m}{0.5n} e^{-0.02n}$$

Remark. A closer look (*e.g.* via a few numerical examples) at (20), (21), reveals that it is easier to construct upper bounds for error $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ where $c < 1$ than $c = 1$. In other words, although the uniqueness theorem guarantees that a solution with less than $n/2$ non-zero entries is unique, constructing an upper bound for $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ with a good confidence using (20) is difficult. On the contrary, for the cases where the number of non-zero components of the sparsest solution is less than a fraction of $n/2$, for example $\|\mathbf{s}_0\| \leq n/4$, highly better bounds are possible (via (21)), and the more sparsity, the better bounds. This can be seen as a statistical explanation to the fact that in numerical algorithms of computing the sparsest solution, the estimation quality decreases where the sparsity of $\mathbf{s}_0$ approaches $n/2$ [30], [19].

VI. Conclusions

In this paper, we showed that it is possible to construct an upper bound on the estimation error $\|\hat{\mathbf{s}} - \mathbf{s}_0\|_2$ without knowing $\mathbf{s}_0$, which is based on $\sigma^{(n)}_{\min}$, the smallest singular value of all $n \times q$ submatrices of $\mathbf{A}$. Moreover, we saw that for the case the URP does not hold, or the case the ℓ^0 norm of the sparse solution is smaller than the theoretical limit $n/2$, other bounds are possible based on $\sigma^{(q)}_{\min}$, the smallest singular value of all $n \times n$ submatrices of $\mathbf{A}$, for a $q < n$. For random dictionaries, this result suggests that the problem of finding sparsest solutions of linear systems and the problem of finding lower bounds on smallest singular values are closely related. Moreover, we developed two stochastic error bounds for random dictionaries. From these bounds, we saw that it is possible to statistically explain why, in algorithms computing the sparsest solution, the estimation accuracy decreases where the sparsity approaches $n/2$.

Appendix

Proof of Lemma 3: We know:

$$s_{\min}(\mathbf{B}) = \min_{\mathbf{x} \in \mathbb{R}^\ell} \frac{\|\mathbf{Bx}\|_2}{\|\mathbf{x}\|_2}, \quad s_{\min}(\mathbf{B}') = \min_{\mathbf{x} \in \mathbb{R}^{\ell+1}} \frac{\|\mathbf{B}'\mathbf{x}'\|_2}{\|\mathbf{x}'\|_2} \tag{22}$$

Let $\mathbf{x}' = [\mathbf{x}, y]^T$, and define:

$$f(\mathbf{x}, y) \triangleq \frac{\|\mathbf{B}'\mathbf{x}'\|_2}{\|\mathbf{x}'\|_2} = \frac{\|\mathbf{Bx} + \mathbf{b}y\|_2}{\sqrt{\|\mathbf{x}\|_2^2 + y^2}} \tag{23}$$

■

Let $\mathbf{x}_1$ be the minimizer of $\frac{\|\mathbf{Bx}\|_2}{\|\mathbf{x}\|_2}$, and $(\mathbf{x}_2, y_2)$ be the minimizer of $f(\mathbf{x}, y)$. Then since $(\mathbf{x}_2, y_2)$ is the minimizer of $f(.,.)$:

$$s_{\min}(\mathbf{B}') = f(\mathbf{x}_2, y_2) \leq f(\mathbf{x}_1, 0) = s_{\min}(\mathbf{B}) \tag{24}$$

Proof of Lemma 4: In this case, since $\mathbf{B}$ is wide ($\|\cdot\|$ stands for the ℓ^2 norm):

$$s_{\min}(\mathbf{B}) = \min_{\|\mathbf{x}\|=1, \mathbf{x} \in \mathbb{R}^\ell} \|\mathbf{B}^T\mathbf{x}\|, s_{\min}(\mathbf{B}') = \min_{\|\mathbf{x}\|=1, \mathbf{x} \in \mathbb{R}^\ell} \|\mathbf{B}'^T\mathbf{x}\|$$

We write:

$$\|\mathbf{B}'^T\mathbf{x}\|_2^2 = \left\| \begin{bmatrix} \mathbf{B}^T\mathbf{x} \\ \mathbf{b}^T\mathbf{x} \end{bmatrix} \right\| = \|\mathbf{B}'^T\mathbf{x}\|_2^2 + \|\mathbf{b}^T\mathbf{x}\|_2^2$$

and hence $\forall \mathbf{x}$, $\|\mathbf{B}'^T\mathbf{x}\|_2 \geq \|\mathbf{B}^T\mathbf{x}\|_2$, which proves the lemma.

■

Proof of Lemma 5: The proof is based on some modifications to the proof of Lemma 1 of [19].

Let m_l be the number of components of $\boldsymbol{\delta}$ which are larger than α, and m_s be the number of components of $\boldsymbol{\delta}$ that are smaller than or equal to α. Note that $m_s + m_l = m$. We discriminate two cases:

Case 1 ($m_l \geq 1, m_s \leq m-1$): In this case, there is at least one component of δ larger than α. Let $\boldsymbol{\delta}_l$ be composed of the components of $\boldsymbol{\delta}$ which have magnitudes larger than α, and $\mathbf{A}_l$ be composed of the corresponding columns of $\mathbf{A}$. Similarly, let $\boldsymbol{\delta}_s$ be composed of the components of $\boldsymbol{\delta}$ which have magnitudes less than or equal to α, and $\mathbf{A}_s$ is composed of the corresponding columns of $\mathbf{A}$. Since $\boldsymbol{\delta} \in \mathrm{null}(\mathbf{A})$, $\mathbf{0} = \mathbf{A}\boldsymbol{\delta} = \mathbf{A}_l\boldsymbol{\delta}_l + \mathbf{A}_s\boldsymbol{\delta}_s$, and hence:

$$\mathbf{b} \triangleq \mathbf{A}_l\boldsymbol{\delta}_l = -\mathbf{A}_s\boldsymbol{\delta}_s \tag{25}$$

From $\mathbf{b} = -\mathbf{A}_s\boldsymbol{\delta}_s$:

$$\|\mathbf{b}\|_2 = \|\mathbf{A}_s\boldsymbol{\delta}_s\|_2 = \|\sum_i \delta_{s,i}\mathbf{a}_{s,i}\|_2 \leq \sum_i \underbrace{|\delta_{s,i}|}_{\leq\alpha}\underbrace{\|\mathbf{a}_{s,i}\|_2}_{1}$$
$$\Rightarrow \|\mathbf{b}\|_2 \leq m_s\alpha \leq (m-1)\alpha < m\alpha \tag{26}$$

From $\mathbf{b} = \mathbf{A}_l\boldsymbol{\delta}_l$, and since from the assumption $m_l \leq q$ and hence is of full column rank:

$$\|\mathbf{b}\|_2 = \|\mathbf{A}_l\boldsymbol{\delta}_l\|_2 \geq \sigma_{\min}(\mathbf{A}_l)\|\boldsymbol{\delta}_l\|_2$$
$$\Rightarrow \|\boldsymbol{\delta}_l\|_2 \leq \frac{\|\mathbf{b}\|_2}{\sigma_{\min}(\mathbf{A}_l)} \tag{27}$$

Note that in the above, the assumption $m_l \leq q$ was essential, otherwise $\|\mathbf{A}_l\boldsymbol{\delta}_l\|_2$ and $\sigma_{\min}(\mathbf{A}_l)$ could be zero. Combining now (27) and (28), we will have:

$$\|\boldsymbol{\delta}_l\|_2 < \frac{m\alpha}{\sigma_{\min}(\mathbf{A}_l)} \tag{28}$$

Moreover, $\|\boldsymbol{\delta}_s\|_2 \leq m_s\alpha \leq m\alpha$. Therefore:

$$\|\boldsymbol{\delta}\|_2 \leq \|\boldsymbol{\delta}_l\|_2 + \|\boldsymbol{\delta}_s\|_2 < \frac{m\alpha}{\sigma_{\min}(\mathbf{A}_l)} + m\alpha \tag{29}$$

Now, from the definition (7) and Lemma 3, $\sigma_{\min}(\mathbf{A}_l) \geq \sigma^{(m_l)}_{\min}(\mathbf{A}_l) \geq \sigma^{(q)}_{\min}(\mathbf{A}_l)$, which proves the lemma.

Case 2 ($m_l = 0, m_s = m$): In this case, all the components of $\boldsymbol{\delta}$ have magnitudes less than or equal α, and hence we can simply write $\|\boldsymbol{\delta}\|_2 \leq m\alpha$ which satisfies also (15). ■

REFERENCES

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

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

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

[4] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[6] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–124, July 2007.

[7] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges," in *Proceedings of ESANN'06*, April 2006, pp. 323–330.

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

[9] P. G. Georgiev, F. J. Theis, and A. Cichocki, "Blind source separation and sparse component analysis for over-complete mixtures," in *Proceedinds of ICASSP'04*, Montreal (Canada), May 2004, pp. 493–496.

[10] Y. Li, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *ICA2003*, 2003, pp. 89–94.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[12] 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.

[13] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

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

[15] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *IEEE Internation Conference on Image Processing (ICIP)*, August 2005, pp. II–782–5.

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

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

[18] H. Mohimani, M. Babaie-Zadeh, and Ch. Jutten, "Fast sparse representation based on smoothed l0 norm," in *Proceedings of 7th International Conference on Independent Component Analysis and Signal Separation (ICA2007), Springer LNCS 4666*, London, UK, September 2007, pp. 389–396.

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

[20] R. Gribonval, R.M. Figueras i Ventura, and P. Vandergheynst, "A simple test to check the optimality of a sparse signal approximation," *Signal Processing (Elsevier)*, vol. 86, pp. 496–510, 2006.

[21] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.

[22] A. Edelman, *Eigenvalues and condition number of random matrices*, Ph.D. thesis, Department of Mathematics, MIT, Cambridge, MA, 1989.

[23] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Now Publishers, 2004.

[24] M. Rudelson and R. Vershynin, "The smallest singular value of a random rectangular matrix," 2008, available online at arXiv:0802.3956v1.

[25] J.W. Silverstein, "The smallest eigenvalue of a large dimensional wishart matrix," *Ann. Probab*, vol. 13, pp. 1364–1368, 1985.

[26] Z.D. Bai and Y.Q. Yin, "Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, no. 3, pp. 1275–1294, 1993.

[27] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, December 2006.

[28] S.J. Szarek, "Condition numbers of random matrices," *J. Complexity*, vol. 7, pp. 131–149, 1991.

[29] J. Shen, "On the singular values of gaussian random matrices," *Linear Algebra and its Applications*, vol. 326, no. 1–3, pp. 1–14, 2001.

[30] A. A. Amini, M. Babaie-Zadeh, and Ch. Jutten, "A fast method for sparse component analysis based on iterative detection-projection," in *AIP Conference Proceeding (MaxEnt2006)*, 2006, vol. 872, pp. 123–130.