An Improved Smoothed l_0-norm Algorithm Based on Multiparameter Approximation Function

Zhifu Cui
Institute of Communication Engineering
PLA Univ. of Sci. & Tech.
Nanjing, China
czf_e@hotmail.com

Hang Zhang
Institute of Communication Engineering
PLA Univ. of Sci. & Tech.
Nanjing, China
hangzh@126.com

Wei Lu
Institute of Communication Engineering
PLA Univ. of Sci. & Tech.
Nanjing, China
luwei@hotmail.com

***Abstract*-In this paper, by representing the sources with subspace method, it is proved that the iteration and projection process of SL0 (Smoothed L0-norm Algorithm) is equivalent to search the sparsest solution in solution space of the system equation. Then, an improvement for over-complete sparse decomposition is introduced, which replaces the Gaussian function used in SL0 algorithm with Sigmoid one to improved its noise tolerance. And, the algorithm could convergence in less iteration. It is experimentally shown that the improved algorithm is superior to the original one on the performance of noise tolerance and speed while keeping its performance related to the probability of activity of the sources.**

Keywords-Overcomplete sparse decomposition, underdetermined blind sources separation, subspace representation, smoothed l_0-norm, sparse component analysis (SCA)

I. INTRODUCTION

In general linear BSS problem, it is assumed that $\mathbf{x} = \mathbf{As} + \mathbf{n}$, where $\mathbf{s}$, $\mathbf{x}$ and $\mathbf{n}$ are the $m \times 1$, $n \times 1$ and $n \times 1$ vectors of sources, recorded signals and noise respectively, and $\mathbf{A}$ is the unknown mixing matrix. If the system is noiseless, then $\mathbf{n} = \mathbf{0}$. The purpose of BSS is to estimate the unknown sources from the recorded signals. When the system is underdetermined, which means $m > n$, there exist an infinite number of possible solutions even though the mixing matrix $\mathbf{A}$ was estimated accurately. Further information is therefore needed in order to uniquely define the solution of this system. In case the sources are sparse, the underdetermined problem could be resolved by introducing the method of sparse component analysis (SCA). There are always two steps of such kind of algorithms: first estimating the mixing matrix in virtue of the feature of recorded signals when only one source is active[1]~[3], and then recovering the sources on the assumption of a known $\mathbf{A}$. In this paper, we research how to recover the sparse sources fast and accurately in the situation that the mixing matrix has been already estimated exactly.

For a given underdetermined system $\mathbf{x} = \mathbf{As} + \mathbf{n}$ on the condition that the sources are sparse, the problem of source recovery is equal to the one of finding the sparsest solution of the equations $\mathbf{x} = \mathbf{As}$ in noisy case. Usually, the sparsest solution is defined as the one with least nonzero component. According to such a definition, the problem is converted to such a regularization problem:

$$P : \min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t. } \mathbf{As} = \mathbf{x} \tag{1}$$

where $\|\mathbf{s}\|_0$, the l_0 norm of $\mathbf{s}$, is defined as

$$\|\mathbf{s}\|_0 = \sum I_i \,,\; I_i = \begin{cases} 1 & \text{if } s_i \neq 0 \\ 0 & \text{if } s_i = 0 \end{cases} \tag{2}$$

In such an approach, searching the minimum l_0-norm solution is a combinatorial problem, which is non-convex and highly non-smoothed. Further more, l_0 norm of a vector is extremely sensitive to noise, and its value will be completely changed even though the noise is very weak. Therefore, other noise tolerated and smooth definitions of sparsity are proposed as replacer of $\|\mathbf{s}\|_0$. In the literature [4]~[7], basis pursuit(BP) is utilized to find the minimum l_1-norm solution of system equation. And the literature [8] extends this method to l_q norm situation, in which the l_q norm is defined as

$$\|\mathbf{s}\|_q = \left(\sum_i |s_i|^q \right)^{1/q}, \quad 0 < q \leq 1 \tag{3}$$

In the experiment of processing the speech signals, it is approved that best performance is got when q is between 0 and 0.4, but without a universal method to fix its value. Such representation of sparsity, like l_1 or l_q norm, is continuous and noise-tolerated, and the linear programming (LP) algorithms, especially the improved ones, enable the large scale problems with thousands of sources and mixtures to be resolved. However, it is still very slow. Contrary to previous approaches, the literature [9] proposed an algorithm to minimize the l_0 norm directly. By choosing a continuous function to approximate l_0 norm, the algorithm, called SL0, searches the optimal solution based on gradient ascent method. It is shown that this algorithm is about two to three orders of magnitude faster than the interior-point LP solvers, while providing the same or better accuracy. However, the function used in [9] could not conciliate the conflict between the quality of approximation and the tolerance of noise due to its single parameter. Furthermore, the iteration of SL0 is carried in the full space with m dimensions, the result of which, not the solution of system equation, is therefore needed to be modified after iteration. Consequently, whether such a process could insure the final result of iteration to be optimal one is unknown. In this paper, the superiority of the functions with double parameters over the single-parameter ones is discussed, and the Sigmoid function is introduced to replace the Gaussian one. Then, we utilize a subspace-representation method, which iterates on the null space of mixing matrix, to search the hidden component of the sparsest solution, and ultimately prove that the result .

The paper is organized as follows. Section II introduces two quantified guidelines to describe the functions used to approximate l_0 norm, then, compares the Sigmoid function with the Gaussian one in the frame of such guidelines. An iterative algorithm is proposed in section III to search the hidden component of sparsest solution in the null space of mixing matrix. Section VI provides some experimental results of our algorithm and its comparison with SL0.

II. An Approximative Approch For Representing The Sparsity With Smoothed L_0-Norm

There are two advantages to approximate l_0 norm with a continuous function. Firstly, such a representation enable the smoothed l_0 norm to be continuous versus the value of each element. Secondly, such a definition of sparsity could tolerate noise in some extend. To obtain these benefits, functions must have the following properties:

The function, $f(x)$, must be monotone and derivable; $f(x)$ must subject to

$$\begin{cases} \lim\limits_{x\to 0} f(x) = 0 \text{ or } 1 \\ \lim\limits_{x\to \infty} f(x) = 1 \text{ or } 0 \end{cases} \tag{4}$$

Thus, for a single sample x, its l_0 norm is defined as $\|x\|_0 \approx f(x)$ or $\|x\|_0 \approx 1 - f(x)$. For a vector $\mathbf{x}$, its l_0 norm is defined as $\|\mathbf{x}\|_0 \approx \sum_i f(x_i)$. The performance of a function in approximating the l_0 norm could be described by two guidelines, the tolerance of noise and the approximation accuracy to its real l_0 norm. In this paper, α and β are used to quantify these two guidelines, which are respectively defined as $\alpha = g(\omega)|_{\omega=0.5}$ and $\beta = g(\omega)|_{\omega=\omega_c}$, where $g(\omega) = f^{-1}(x)$, equal the value of variable x when $f(x) = \omega$. And α denotes the noise tolerance of the function, similar to the half-power bandwidth of a filter; β denotes the transition width from 0 to 1, similar to the end bandwidth of a filter, which determines how smooth the function is, as well as its approximation accuracy. When the scale of system is large, we could choose $\omega_c = (m-1)/m$, where m is the dimension of source vector. The ideal l_0 norm is a unit step function in origin. In noiseless case, α is expected to be close to 0 as possible, and β is expected to be appropriate to conciliate the confliction between approximation accuracy and smoothness of function. In noisy case, α is expected to be adjustable, and β is expected to be independent of variable α . For the Gaussian function $f(x) = 1 - e^{-x^2/2\sigma^2}$ which is used in [9]

$$\begin{cases} \alpha = \sqrt{2\sigma^2 \ln 2} \approx 1.4\sigma \\ \beta = \sigma\sqrt{2\ln[(m-1)/m]} \end{cases} \tag{5}$$

Apparently, both α and β are dependent on σ . That is to say, the adjustment of α will inevitably affect β . When SNR is high, a smaller α is needed, which means a smaller σ . In such a circumstance, the function $F(\mathbf{x}) = \sum_i f(x_i)$ is highly nonsmooth. On the other hand, when SNR is low, a bigger α is therefore needed, which means worse approximation to l_0 norm because of the bigger σ . In either case, the performance of algorithm will be adversely affected.

By choosing multi-parameter function, the disadvantage of single-parameter function could be overcome. In this paper, Gaussian function is replaced by Sigmoid function to approximate l_0 norm.

The family of Sigmoid functions is defined as

$$f_{\sigma,\mu}(x) = 1\Big/\left(1 + e^{-\frac{x-\mu}{\sigma}}\right) \tag{6}$$

And approximately, the sparsity of source vector is therefore defined by

$$F_{\sigma,\mu}(\mathbf{s}) = m - \sum_{i=1}^{m} f_{\sigma,\mu}(s_i) \tag{7}$$

In such a circumstance

$$\alpha|_{\omega=0.5} = \sqrt{\mu}\ ,\ \beta|_{\omega_c=(m-1)/m} = \sqrt{\sigma \ln(m-1) + \mu} \tag{8}$$

Apparently, α is only determined by parameter μ , and β could be adjusted by changing parameter σ , which means the noise tolerance of the Sigmoid function and its smoothness (or its approximation to l_0 norm) could be controlled independently by different parameter. β changes with values of different σ , while α always remains the same. The smaller value of β brings the better approximation to l_0 norm, but the worse smoothness to the function. How to determine the values of α and β will be discussed in section VI. Furthermore, consider the function defined by (7)

$$f(0) = 1/(1 + \exp(\mu/\sigma)) \neq 0 \tag{9}$$

When the value of μ/σ is large enough, $f(0)$ is sufficiently small and its influence on the performance of algorithm is negligible.

III. The Model Of Sparse Decomposition Algorithm Based On Subspace Representation

In this section, a subspace iterative method will be induced. Comparing with SL0 algorithm, proposed one searches the sparsest solution of system equation on its solution space but not the full m-dimension space. And we will prove that they are equivalent to each other, and both ensure the sparsest solution.

First of all, we define the solution space of equation $\mathbf{As} = \mathbf{x}$ and the null space of mixing matrix $\mathbf{A}$:

$$\begin{cases} \mathbf{U} : \mathbf{U} = \{\mathbf{u} \mid \mathbf{Au} = \mathbf{x}\} \\ \mathbf{N} : \mathbf{N} = \{\mathbf{v} \mid \mathbf{Av} = \mathbf{0}\} \end{cases} \tag{10}$$

Apparently $\forall \mathbf{u} \in \mathbf{U}, \mathbf{v} \in \mathbf{N}$, $\mathbf{u} + \mathbf{v} \in \mathbf{U}$,and $\forall \mathbf{u}_1, \mathbf{u}_2 \in \mathbf{U}$, $\mathbf{u}_1 - \mathbf{u}_2 \in \mathbf{N}$ .Therefore, any solution of $\mathbf{As} = \mathbf{x}$ is the summation of another solution and some linear combination of the basis of space $\mathbf{N}$

$$\mathbf{s} = \mathbf{s}_p + \mathbf{V}\boldsymbol{\alpha} = \mathbf{s}_d + \mathbf{s}_h \tag{11}$$

where $\mathbf{s}_p$ is a particular solution of system equation, $\mathbf{V}$ is a set of basis of space $\mathbf{N}$, and $\boldsymbol{\alpha}$ is the coefficient vector of this linear transformation. That being the case, we may wish to choose the minimum l_2-norm solution $\mathbf{A}^T\left(\mathbf{AA}^T\right)^{-1}\mathbf{x}$ as the particular one, called dominant component of source vector and recorded as $\mathbf{s}_d$. Furthermore, the column vector of $\mathbf{V}$ is expected to be orthogonal with each other. Their linear combination is called hidden component of source vector and recorded as $\mathbf{s}_h$. Because mixing matrix, size of which is $n \times m$, is full rank in row, its null space is therefore (m-n)-dimensional. Assumed that $\mathbf{V} = \left[\mathbf{v}_1, \ \mathbf{v}_2, \cdots, \mathbf{v}_{m-n}\right]$ is a set of orthonormal basis, then $\mathbf{s}_h = \mathbf{V}\boldsymbol{\alpha} = \sum_{j=1}^{m-n} \alpha_j \mathbf{v}_j$ .And

$$\mathbf{s} = \mathbf{s}_d + \mathbf{s}_h = \mathbf{A}^\dagger \mathbf{x} + \sum_{j=1}^{m-n} \alpha_j \mathbf{v}_j \, , \mathbf{A}^\dagger = \mathbf{A}^T\left(\mathbf{AA}^T\right)^{-1} \tag{12}$$

where $\boldsymbol{\alpha} = \left[\alpha_1, \ \alpha_2, \cdots, \alpha_{m-n}\right]^T$, $\mathbf{A}^\dagger$ denotes the Moor-Penrose inverse of $\mathbf{A}$. By doing this, the problem in SL0 to search sparsest solution in full space is converted to the one to search it just in solution space.

As the definition in (8)

$$F_{\sigma,\mu}(\mathbf{s}) = m - \sum_{i=1}^{m} f_{\sigma,\mu}\left(\sum_{j=0}^{m-n} \alpha_j v_{ij}\right) \tag{13}$$

where $\alpha_0 = 1$ and $\mathbf{v}_0 = \mathbf{A}^T\left(\mathbf{AA}^T\right)^{-1}\mathbf{x}$. In (16), the variables have been already converted to $\left[\alpha_1, \ \alpha_2, \cdots, \alpha_{m-n}\right]$. And starting from such an expression, we derive a new gradient descent algorithm based on Gaussian function and Sigmoid function respectively. Iterative equation is as follows (see Appendix for details)

$$\mathbf{s}^l = \mathbf{s}^{l-1} - \lambda \mathbf{T}\boldsymbol{\omega} \tag{14}$$

where $\mathbf{T} = \mathbf{VV}^T, \mathbf{V}^{m\times(m-n)} = null(\mathbf{A})$, $\boldsymbol{\omega}$ is gradient vector, and λ is iterative step. Comparing with SL0

$$\hat{\mathbf{s}}^l = \mathbf{s}^{l-1} - \lambda\boldsymbol{\omega} \tag{15}$$

$$\mathbf{s}^l = \hat{\mathbf{s}}^l - \mathbf{A}^\dagger\left(\mathbf{A}\hat{\mathbf{s}}^l - \mathbf{x}\right) \tag{16}$$

Substituting (18) to (19)

$$\mathbf{s}^l = \mathbf{s}^{l-1} - \lambda\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\boldsymbol{\omega} \tag{17}$$

If choosing the last $m-n$ vectors of the right singular-vector matrix of $\mathbf{A}$ to be $\mathbf{V}$, then we note that $\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}$ is just $\mathbf{T} = \mathbf{VV}^T$, which means the two method is equivalent to each other, and the result got from SL0 is just the sparsest solution of system equation.

As proved in [9], there is unique sparsest solution only when $k < n/2$, where k denotes the amount of source active at the same time, and n is the amount of recorded signal. The convergence of this steepest ascent algorithm has been considered in [10]. Referring to the approach to accelerate convergence by decreasing parameter[10].

IV. Experiment Results

In this section, the performance of presented algorithm based on Gaussian function and Sigmoid function is verified respectively, then compared with SL0 algorithm. The effect of parameters (μ and σ), sparsity, and noise on the performance are also experimentally discussed.

In all experiments, the sources $s_1, \cdots, s_m$ are assumed to be independent and identically distributed random variables. And the active probability of each source is equal to p . When active, the source obeys the Gaussian distribution $N(1, 0.1)$ and $N(-1, 0.1)$ with the same probability of 0.5 respectively. The noise, obeyed the distribution $N\left(0, \sigma_n^2\right)$, is AWGN and independent of each other.

Experiment 1. Robustness against Noise of Different Approximation Functions

This section will analyse the performance of the presented algorithm for different approximation functions, the Gaussian one and the Sigmoid one, then compare their performance with the SL0 in different noise condition. To utilize the Sigmoid function, we first have to determine the parameter value of it. As mentioned earlier, the parameter μ determines the noise tolerance. On known the distribution of noise and signal, the optimal value of μ could be chosen as (22) to best distinguish between signal and noise in the sense of expectation.

$$\begin{aligned} \mu_{op} = \arg\min_{\mu} \sum_{i=1}^{m} \Big[& p_i - p_i E\left(f\left(s_i^{act} + \tilde{n}_i\right)\right) \\ & -\left(1 - p_i\right) E\left(f\left(s_i^{act} + \tilde{n}_i\right)\right)\Big] \end{aligned} \tag{18}$$

where p_i denotes the active probability of s_i , $\tilde{n}_i$ denotes the i th component of the vector $\mathbf{A}^\dagger\mathbf{n}$. In this experiment, the sources are i.i.d.. (22) is therefore simplified to

$$\mu_{op} = \arg\min_{\mu} \Big[p - pE\Big(f\big(s_i^{act} + \tilde{n}_i \big) \Big) - (1-p) E\Big(f\big(s_i^{act} + \tilde{n}_i \big) \Big) \Big] \tag{19}$$

Fixing μ, there are three factor needed to be considered when choosing σ. Firstly, it must ensure the approximation accuracy of the Sigmoid function to l_0 norm. Secondly, enough smoothness, determined by σ, is needed to guarantee the convergence of the algorithm. Finally, as discussed before, $f_{\sigma,\mu}(0) \neq 0$, therefore, μ/σ must be chosen to ensure $f(0)$ is small enough. In this experiment, we choose a decreasing sequence for σ, and the least one is fixed to be 0.01.

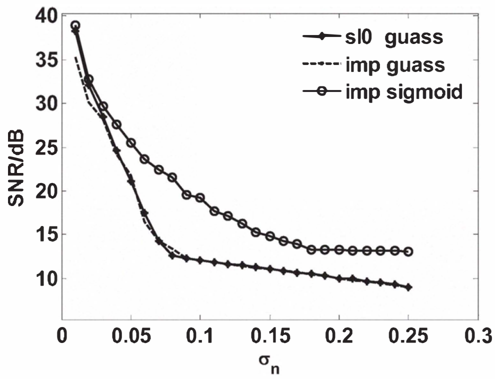

Figure 1. Averaged SNRs(over 10 runs of the algorithms) versus the noise variance σ_n

Fig. 2 shows the averaged separation SNR of each algorithm versus the noise variance. When the value of noise is small, the averaged SNR of the SL0, the presented algorithm using Gaussian function and Sigmoid function is almost equal. While σ_n is increasing, our algorithm using Sigmoid function performs better than both the one using Gaussian function and the SL0. On such assumption of the signal and noise distribution, when the value of σ_n is bigger than 0.05, the advantage of using Sigmoid function is remarkable, and more than 5dB extra SNR could be got.

Experiment 2. Effect of Sparsity on the Performance

As discussed in [8], there is a theoretical limit of $n/2$ on the maximum number of active sources to insure the uniqueness of the sparsest solution. But practically, most algorithms cannot achieve this limit.

In this experiment, we set σ_n to be 0.01, and the scale factor δ is chosen as 0.9. Fig.3 shows the averaged SNR of each algorithm versus the active probability of source. It is obvious that all algorithms work well when active probability of source is small, and then breakdown when p is more than a critical value. Considering the influence of randomicity of sources and noise, the critical value for each method is almost the same, and the performance of the presented algorithm using Gaussian and Sigmoid function is very close to SL0 for different value of active probability.

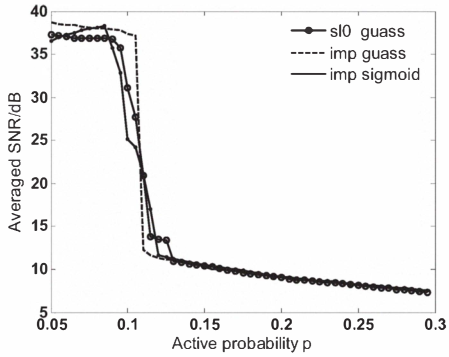

Figure 2. Averaged SNRs(over 3 runs of the algorithms) versus the active probability of sources.

V. Conclusion

In this paper, we derived an algorithm based on the subspace representation for the solution of the system equation. And then, we proved that such one is the same to SL0, which searches the sparsest solution on the whole space and then projects this solution to the solution space of system equation. Furthermore, in order to coordinate the contradiction between noise tolerance and accuracy of the function used to approximate the l_0 norm, we introduced Sigmoid function with two parameters instead of Gaussian one. Numerical experiments show that such an improvement is more effective when dealing with high noise power situation. However, this method is very sensitive to the active probability of sources. As its value is bigger than a critical value (usually less than 0.15), the performance will breakdown quickly. This paper does not make progress in this aspect. Next, we will focus on research in this area.

References

[1] Xiao M., Xie S. L., Fu Y. L.. Underdetermined Blind Delayed Source Separation Based on Single Source Intervals in Frequency Domain[J].ACTA ElECTRONICA SINICA, 2007, 35(12): p. 2279-2283.

[2] SangGyun Kim. Underdetermined Blind Source Separation Based on Subspace Representation [J]. IEEE Trans. Signal Processing, vol. 57(7), pp. 2604-2614, Jul. 2009.

[3] D. L. Donoho. Stable recovery of sparse over-complete representation in the presence of noise[J]. IEEE Trans. Inf. Theory, vol. 52, no. 1, pp. 6-18, Jan. 2006.

[4] Ichigaku Takigawa. Performance Analysis of Minimum l_1-Norm Solutions for Underdetermined Source Separation[J]. IEEE Trans. Signal Processing, vol. 52(3), pp. 582-591, Mar. 2004.

[5] Yuanqing Li, Shun-Ichi Amari. Underdetermined Blind Source Separation Based on Sparse Representation[J]. IEEE Trans. Signal Processing, vol.54, NO.2, pp. 423-437, Feb. 2006.

[6] D. L. Donoho. For most large underdetermined systems of linear equation the minimal l_1-norm solution is also the sparsest solution. Tech. Rep., 2004[Online]. http:/www-stat. stanford. edu/~donoho/ reports.html.

[7] Rayan Saab. Underdetermined Anechoic Blind Source Separation via l_q-basis-pursuit with q <1[J]. IEEE Trans. Signal Processing, vol. 55(8), pp. 4004-4016, Aug. 2007.

[8] Hosein Mohimani. A Fast Approach for Over-complete Sparse Decomposition Based on Smoothed l_0-Norm[J]. IEEE Trans. Signal Processing, vol. 57(1), pp. 289-301, Jan. 2009.

[9] Irina F. Gorodnitsky. Sparse Signal Reconstruction from Limited Data Using FOCUSS:A Re-weighted Minimum Norm Algorithm[J]. IEEE Trans. Signal Processing, vol. 45(3), pp. 600-616, Mar. 1997.