Chalmers Publication Library

On the resolution of the LASSO-based DOA estimation method

(article starts on next page)

On The Resolution of The LASSO-Based DOA Estimation Method

Ashkan Panahi and Mats Viberg

Signal Processing Group

Chalmers University of Technology, Gothenburg, Sweden

Email: {*Ashkanp , Viberg*}*@chalmers.se*

Abstract—**This paper investigates the consistency of the LASSO-based DOA estimation of the narrow-band signals in infinitely high SNR. Such a method provides a robust and accurate approximation of the Maximum Likelihood estimation. However, as we show, unlike the standard techniques such as subspace methods the LASSO-based estimation is generally not consistent in high SNRs. In return, considering the true DOA's, we show that the method is consistent for certain configuration of the sources. This approach leads us to relate such a conditional consistency to the resolution concept. We next give a condition to verify the consistency of a given set of directions and simplify it to a computationally fast equivalent algorithm. The results show that the resolution in infinitely high SNR case for m sensors decreases by speed $\frac{1}{m}$.**

Index Terms—**DOA estimation, LASSO, performance analysis, consistency analysis, resolution.**

Fig. 1. The configuration of the sources and the sensors. The direction angles are measured from the array axis.

I. INTRODUCTION

The Least Absolute Shrinkage and Selection Operator (LASSO) [1] method is a well known tool in compressive sensing and sparse linear regression application fields. It is also being widely used as a reliable technique in many estimation problems. This method consists of a linear least square optimization regularized by an additional penalty term of the 1-norm as a measure of sparseness. The LASSO method provides a low computational cost algorithm due to its convex nature. The convexity property also ensures a robust estimation because there exists only a unique local minimum.

The LASSO technique can particularly, be utilized to estimate the Directions of Arrivals (DOA) of narrow-band signals transmitted by far-field sources and received by an array of antennas. The received signals for such a scenario is modeled as a linear combination of the *steering vectors*. The estimation problem is conventionally solved by the techniques such as Non-Linear Least Square (NLLS) method, Maximum Likelihood (ML) [2], and subspace-based methods [3]. The DOA estimation problem can be reformulated as a selection problem from a dictionary of sample steering vectors of the continuous one-dimensional array manifold ([4] and [5]). This representation allows us to apply the LASSO method to estimate the directions.

Many questions arise concerning the performance of the LASSO technique as such an estimation tool. These are generally discussed in the performance analysis. It is commonly expected that a good estimator gives a nearly *unbiased* estimation of the true unknown parameters with a relatively
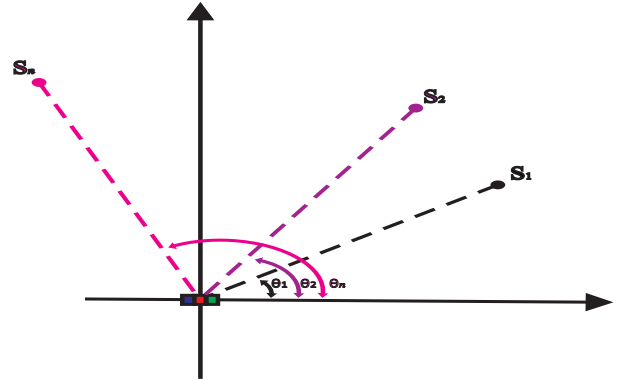
small error variance. This can be studied by introducing an asymptotic consistent situation in which the estimated parameters converge to the true values[6]. In [7] a condition is introduced for the LASSO method to be generally consistent in the asymptotic noiseless case. The condition is generally hard to verify because it includes a non-convex optimization with many singular points. For the LASSO based DOA estimation, we observe that the method is not generally consistent. This can be interpreted as a resolution limit in the method. In this case, we modify the condition of [7] to verify the consistency for each particular configuration of the target sources. Next, we introduce an auxiliary convex condition to verify the former condition, which results in a computable lower bound on the resolution. Later, in Section IV we empirically observe that the lower bound is in fact the exact resolution limit. Thus, we claim that the latter auxiliary condition is equivalent to the former consistency condition, although a complete theoretical proof is not given.

Equipped with a fast computable tool to verify the consistency of the estimation we investigate the behavior of the LASSO technique in the noiseless case in Section IV.

II. PROBLEM STATEMENT

A. Direction of Arrival Estimation

Consider a Uniform Linear Array (ULA) of m omni-directional sensors with separation d receiving narrow-band signals at frequency f_0 from n far-field sources at angles

$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]$ as shown in Figure 1. Due to the plane-wave model of the field around the array the received signals complex envelope $\mathbf{x}(t)$ can be written as [8]

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \ , \tag{1}$$

where $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are the transmitted data and the noise vector respectively. The noise is assumed to be white, zero-mean, and Gaussian with covariance matrix $\sigma^2\mathbf{I}$. The matrix $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_n)]$ is the collection of the steering vectors corresponding to the directions of the sources. A steering vector from a direction θ is given by

$$\mathbf{a}(\theta) = \begin{bmatrix} 1 \\ e^{j\phi} \\ e^{j2\phi} \\ \vdots \\ e^{j(m-1)\phi} \end{bmatrix}, \tag{2}$$

where for simplicity we introduce the *electrical angle* $\phi = 2\pi\frac{d}{\lambda}\cos\theta$ with λ as the wavelength. For a full-resolution and non-ambiguous estimation, d is normally half of the wavelength ($d = \frac{\lambda}{2}$). Later, we may use the notation $\mathbf{a}(\phi)$ to represent the steering vector as an explicit function of the electrical angle.

Such a DOA estimation problem can be reformulated as a sparse regression one. First, we discretize the problem so that we are only allowed to choose from a fine grid of directions $G = \{\theta_1^g, \theta_2^g, \ldots, \theta_N^g\}$. Suppose the true source directions are in G as well. Next, we introduce the extended "time" source vector $\mathbf{s}^g(t)$ as the $1 \times N$ spatial spectrum of the waveform signals received from each direction of the space. Ideally, we have

$$s_i^g(t) = \begin{cases} s_k(t) & \theta_i^g = \theta_k \in \boldsymbol{\theta} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Finally, introducing $\mathbf{A}^g = [\mathbf{a}(\theta_1^g) \ \mathbf{a}(\theta_2^g) \ \ldots \ \mathbf{a}(\theta_N^g)]$ as the large dictionary of all steering vectors, we can write the model as

$$\mathbf{x}(t) = \mathbf{A}^g\mathbf{s}^g(t) + \mathbf{n}(t) \ . \tag{4}$$

In the following, we will only consider the single snapshot case, so the time dependence of $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{n}$ is dropped.

B. Least Absolute Shrinkage and selection Operator

It can be seen that the model introduced in (4) is under-determined. It means that the problem does not have a unique solution until a sparsity limitation is introduced. The LASSO method is a least square optimization regularized by the ℓ_1 norm as a measure of sparsity [1]. In this manner, it combines the model in (4) with the sparsity assumption. The ℓ_1 norm ensures a sparse solution because it is a piecewise linear function with the singular points at sparse points. Because of the nature of the regularization, the optimal point occurs on one of these singular vertices[1].

The LASSO optimization can be written in a variety of equivalent forms. However, in our special noiseless, one-snapshot case. The following form is preferred ([1])

$$\hat{\mathbf{s}}^g = \arg\min_{\mathbf{s}^g} \|\mathbf{s}^g\| \quad \text{subject to} \quad \mathbf{x} = \mathbf{A}^g\mathbf{s}^g \ . \tag{5}$$

This problem is convex with a linear constraint and it can be solved efficiently using convex optimization techniques. The question is now "under which constraints are the estimated vector $\hat{\mathbf{s}}^g$ and the true vector $\mathbf{s}_0^g$ identical ?". Note that in the noiseless case $\mathbf{x} = \mathbf{A}^g\mathbf{s}_0^g$. Although we are not able to give an analytical solution, in the next chapter we provide some observations leading to a computable method of searching for such consistent cases.

III. CONSISTENCY CONDITIONS

Here, we give some theorems which we later use to give the consistency criterion. In fact, such a criterion has been introduced in [7]. We will first paraphrase and complete it for our application. However, this theorem can not be solved neither analytically, nor with a computationally fast method. For the following theorem we introduce the *active set* $I = \{i_1, i_2, \ldots, i_n\}$ as the set of all indexes in $\mathbf{s}_0^g$ with nonzero components. In other words, the elements of I are the indexes corresponding to true DOAs.

Theorem 1: Consider a set of active basis $I = \{i_1, i_2, \ldots, i_n\}$. For every nonzero realization of the original source vector $\mathbf{s}_0$ with I as the active set, the solution to the noiseless LASSO problem, $\hat{\mathbf{s}}^g$, is identical to the original source if and only if $\psi(I) \leq \frac{1}{2}$, where

$$\psi(I) = \max_{\boldsymbol{\delta} \in \mathcal{N}_{\mathbf{A}^g} - \{\mathbf{0}\}} \frac{\sum\limits_{i \in I} |\delta_i|}{\|\boldsymbol{\delta}\|_1} \ . \tag{6}$$

The $\mathcal{N}_{\mathbf{A}^g} - \{\mathbf{0}\}$ is the null space of the matrix $\mathbf{A}^g$ without the zero vector.

Proof: The proof is given in Appendix A.

∎

We call the cost function in (6) the *active portion* of the null vector $\boldsymbol{\delta}$. The following result can immediately be seen from this theorem.

Corollary 1: The LASSO based DOA is always consistent for the one-source case.

Proof: assume that $I = \{i\}$. For each $\boldsymbol{\delta} \in \mathcal{N}_{\mathbf{A}} - \{\mathbf{0}\}$ we have $\mathbf{A}\boldsymbol{\delta} = \mathbf{0}$. From the first row equation of this matrix relation we have $\sum\limits_{j=1}^{N} \delta_j = 0$ which can also be written as $\delta_i = -\sum\limits_{j \neq i} \delta_j$. Taking the absolute value and using the the triangle inequality we have $|\delta_i| = |\sum\limits_{j \neq i} \delta_j| \leq \sum\limits_{j \neq i} |\delta_j|$. Finally, adding the term $|\delta_i|$ to both sides of the equation we get $2|\delta_i| \leq \|\boldsymbol{\delta}\|_1$. This shows that for each arbitrary null vector $\boldsymbol{\delta}$ and active index i, the active portion is less than half which proves the corollary.

∎

The condition introduced in (6) gives a criterion for each possible active basis set to be consistent. However, it is

computationally hard to be verified directly. Thus a simpler equivalent form should be introduced.

First, note that for the optimum point in (6) there exists a component δ_r ($r \in I$) which is nonzero. Without loss of generality, we can assume that $\delta_r = 1$, because the active portion is invariant under scaling transformation. Next, we observe that maximizing the active portion is equivalent to maximizing $\frac{\sum_{i \in I} |\delta_i|}{\sum_{i \notin I} |\delta_i|}$. A high value of the latter can be obtained by minimizing the denominator while keeping the numerator high enough. This will be automatically done by adding the constraint $\delta_r = 1$. Accordingly, we introduce the following optimization to approximate the solution of (6).

$$\boldsymbol{\delta}_0 = \arg \min_{\boldsymbol{\delta} \in \mathscr{N}_{\mathbf{A}}, r \in I} \sum_{i \notin I} |\delta_i| \text{ subject to } \delta_r = 1 . \quad (7)$$

Computing $\boldsymbol{\delta}_0$ is fast because for each $r \in I$ the optimization can be performed by a convex optimization technique. If the active portion for $\boldsymbol{\delta}_0$ is more than half, due to the condition in (6) the active set is non-consistent. The converse is not necessarily true. However, in Section IV we present some empirical results implying the equivalence between two conditions in (6) and (7).

IV. EMPIRICAL OBSERVATIONS

The results shown in this section are generated by implementing the aforementioned algorithms by MATLAB software. We use the CVX [9] toolbox to perform the convex optimization parts.

A. Two-source case consistency

When there are two active indexes, the concept of consistency can be viewed as the ability of the method to resolve close sources. This idea leads to the definition of the *fundamental resolution* as the minimum angular separation between two close sources for which the LASSO based DOA estimation is consistent in the noiseless case. The term "fundamental" is related to the fact that the minimum separation may increase by adding noise to the problem. From this point of view the fundamental resolution is the resolution in the infinitely high Signal to Noise Ratio (SNR) case and the best one.

Due to the circular symmetry of the cost function in (6), when expressed in terms of electrical angle the resolution is independent of the direction of the sources. In this case, it is only a function of the number of sensors m. Let us denote this resolution by $\Delta\phi_0(m)$. Then, for a small value of $\Delta\phi_0(m)$ we have

$$\Delta\phi_0(m) = 2\pi \frac{d}{d_0} \Delta \cos\theta \approx -2\pi \frac{d}{d_0} \sin\theta \Delta\theta . \quad (8)$$

this shows that in terms of the direction angle, the resolution at direction θ is proportional to $\frac{1}{\sin(\theta)}$.
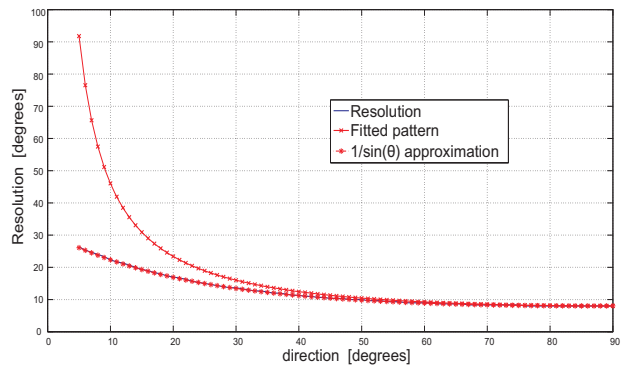

Fig. 2. The resolution of a half-wavelength ULA with 8 sensors for different directions. The exact theoretical and the true resolutions coincide.
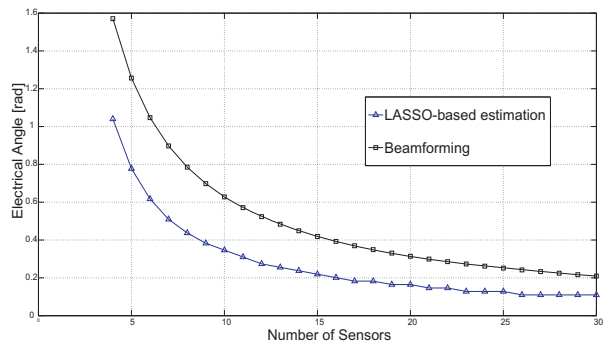

Fig. 3. The minimum resolvable electrical angle Vs the number of sensors for the LASSO-based and the beamforming techniques.

B. The equivalence of the auxiliary optimization

In Section III, we propose the auxiliary optimization (7) to verify the condition (6). As we explained, the auxiliary one gives a necessary condition. In the case of two sources, this implies that we get a lower resolution bound by applying this optimization. Figure 2 shows the result of examining various active sets of two directions at angles θ and $\theta + \Delta\theta$. For each θ we look for the minimum $\Delta\theta$ for which the active set is consistent. The consistency is verified by (7). Although this minimum separation is supposed to be a lower bound for the resolution, direct application of LASSO for different realizations of the sources shows that LASSO-based DOA estimation for separations higher than the values given in Figure 2 is consistent. This indicates that the auxiliary optimization is actually equivalent to the exact condition in (6). Furthermore, it can be seen that although we associate the concept of non-consistency to the existence of at least one waveform with a wrong solution, in practice, in a non-consistent situation we get a wrong result for almost all realizations of the waveform $\mathbf{s}$. Also note that, as we previously explained, the empirical curve in Figure 2 fits to the pattern in (8), and also the approximate curve of $\frac{A}{\sin(\theta)}$ with a proper value of A. Convinced by the equivalence between the two conditions, we can compare the resolution of the LASSO-based method to that of the conventional beamforming technique. Figure
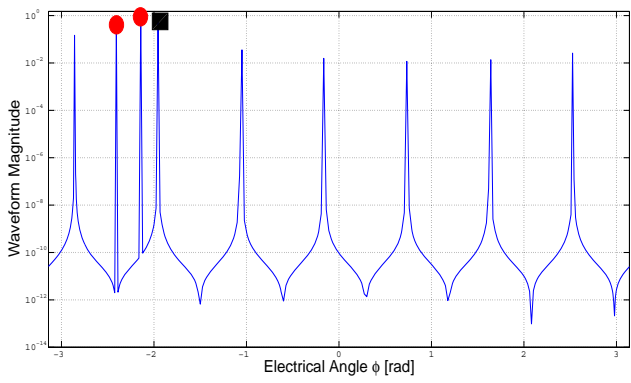
Fig. 4. A typical non-consistent situation for two sources at $40°$ and $47°$ in a half-wavelength ULA of 8 sensors. The null waveform is plotted in the logarithmic scale. The circles show the active indexes and the rectangle is related to the dominant third component. The other components are relatively small.
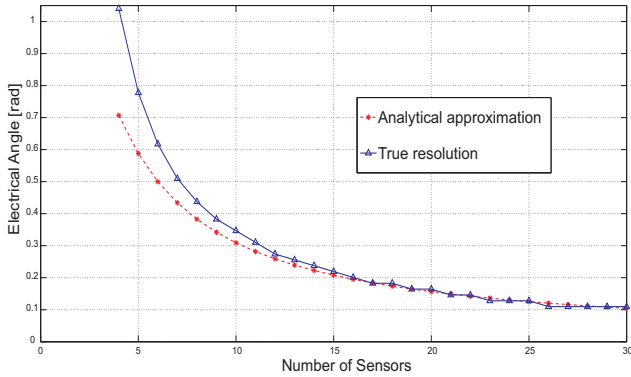

Fig. 5. The analytical approximation compared to the exact resolution for different number of sensors.

3 compares the fundamental resolution of the LASSO-based estimation and beamforming for different numbers of sensors in a half-wavelength ULA. The resolution is expressed in terms of the electrical angle. Thus, it is independent of the direction. The fundamental resolution of the beamforming method is approximately equal to half of the main-lob beam width ($\Delta\phi_0 = \frac{2\pi}{m}$). Obviously, the LASSO-based estimation has a better resolution.

C. Non consistent case analysis and an approximate resolution formula

When the active set I is non-consistent we can learn the form of the null vector $\boldsymbol{\delta}$ which introduces ambiguity by looking at the solution of (7). Figure 4 shows such a vector. As can be seen, the solution is sparse itself. This can be explained by a reasoning similar to the work of [10]. For the optimum point in (7), the vector $\boldsymbol{\delta}$ consists of two equally large peaks at the position of active indexes and one other smaller peak. In the non-consistent scenario these three dominant components are linearly combined so that the other components become negligible. Limiting our search to the class of such vectors, we get the following approximation for the resolution $\Delta\phi_0(m)$.

Theorem 2: In terms of the electrical angle ϕ, the fundamental resolution can be approximated by

$$\Delta\phi_0(m) \approx \frac{1}{\max \frac{d}{d\phi}\left(\frac{\sin m\phi}{\sin \phi}\right)} . \tag{9}$$

For a large m it can be further simplified as

$$\Delta\phi_0(m) \approx \sin(\frac{\pi}{m}) \approx \frac{\pi}{m} . \tag{10}$$

The result can be seen in figure 5 where the exact resolution curve is compared to the curve of $\Delta\phi_0(m) = \sin(\frac{\pi}{m})$. As can be seen, the two curves coincide for $m \geq 15$ say.

D. More Than Two Source Case

We also examine the condition (6) by the optimization (7) for some cases in which the active set has more than two elements. The results firstly emphasis on the equivalence between the two conditions introduced in Section III. Furthermore, we find out that adding a new active index will decrease the resolution by a negligible value. Thus, as a rule of thumb, we can state that a general set of active indexes is consistent if any two elements of this set are separated by more than approximately the fundamental resolution.

V. CONCLUDING REMARKS

A. Summary

In this work we discussed the consistency of the LASSO-based DOA estimation method. We first observed that the method is not consistent in all cases even with infinitely high SNR. This can be expressed as the finite resolution of the DOA estimation. It is a disadvantage for the method as compared to the subspace methods with infinite resolution in high SNRs. However, we show that the resolution is still better than that of classical beamforming.

Next, we discussed the consistency for each true configuration of the sources and gave a criterion for such a configuration to be consistent. Because the condition is hard to verify we introduced an alternative using a convex optimization technique. From the results, we claim that the tool is equivalent to the exact criterion. Finally, observing the results, we find and approximate analytical formula for the minimum resolvable angle separation which is half of classical beamforming.

B. Why Non-consistency?

As a final conclusion, we give a discussion about the non-consistent cases. It is based on the results obtained from applying the auxiliary optimization. For a non-consistent case there exists a nonzero vector $\boldsymbol{\delta}$ in the null space of $\mathbf{A}^g$ with an active portion higher than half. Roughly speaking, such a "neutral" vector can be added to the true extended source vector to remove a big share of ℓ_1 norm from the true direction indexes, and introduce a smaller share at some "fake" indexes. The result is a wrong solution with smaller ℓ_1 norm and wrong positions. But why does such a vector exist? The reason is that the huge dictionary contains samples of a continuous manifold,

which means that close indexes correspond to very close basis vectors. If we choose such close indexes i and j, as the active ones then $\mathbf{a}_i^g - \mathbf{a}_j^g + \mathbf{v} = 0$ where $\mathbf{v}$ is small. This forms a null vector with values 1 and -1 at indexes i and j respectively, and a small ϵ share of ℓ_1 norm at other indexes. As can be seen, for such a vector the active portion is $\frac{2}{2+\epsilon}$. If $\mathbf{v}$ is small enough, so is ϵ, and the active portion can be bigger than half. By the same reasoning we can understand that any estimation using a least square regularized by a continuous norm has a finite resolution. It should also be pointed out that the resolution is independent of the sampling method used to generate the dictionary matrix.

VI. APPENDIX A: PROOF OF THEOREM 1

For simplicity we first introduce the notation $\|\boldsymbol{\delta}\|_I = \sum_{i \in I} |\delta_i|$ for any index set I and null vector $\boldsymbol{\delta}$. Now, suppose for an index set I the condition is satisfied. Also, suppose there exists a signal vector $\mathbf{s}^g$ with active vectors at I for which the solution is some different vector $\mathbf{s}^{g'}$. Then we have $\mathbf{A}^g \mathbf{s}^g = \mathbf{A}^g \mathbf{s}^{g'}$ and $\|\mathbf{s}^{g'}\|_1 < \|\mathbf{s}^g\|_1$. Introducing $\boldsymbol{\delta} = \mathbf{s}^{g'} - \mathbf{s}^g$, we note that $\boldsymbol{\delta} \in \mathcal{N}_{A^g}$ and

$$\|\boldsymbol{\delta} + \mathbf{s}^g\|_1 < \|\mathbf{s}^g\|_1 \ . \tag{11}$$

Since $\|\mathbf{s}^g\|_1 = \|\mathbf{s}^g\|_I$, this can be also written as

$$\|\mathbf{s}^g\|_I > \|\boldsymbol{\delta} + \mathbf{s}^g\|_I + \|\boldsymbol{\delta}\|_{I^c} > \|\mathbf{s}^g\|_I - \|\boldsymbol{\delta}\|_I + \|\boldsymbol{\delta}\|_{I^c} \ , \tag{12}$$

where $I^c = \{i \in G | i \notin I\}$ is the complement index set of I. The last inequality is the result of the triangle inequality. This means that $\|\boldsymbol{\delta}\|_{I^c} < \|\boldsymbol{\delta}\|_I$ so that $\boldsymbol{\delta}\|_1 < 2\boldsymbol{\delta}\|_I$ which violates $\psi(I) \leq \frac{1}{2}$, see (6).

Second, suppose there exists $\boldsymbol{\delta} \in \mathcal{N}_{A^g} - \{\mathbf{0}\}$ so that $\|\boldsymbol{\delta}\|_{I^c} < \|\boldsymbol{\delta}\|_I$. Let us denote the active basis for this vector by J. Assume a vector $\mathbf{s}^g$ with I as active indexes, and the additional property that its elements at the indexes in $I \cap J$ are the negative of the corresponding elements in $\boldsymbol{\delta}$ so that $\|\boldsymbol{\delta} + \mathbf{s}^g\|_{I \cap J} = 0$ and also $\|\mathbf{s}^g\|_{I \cap J} = \|\boldsymbol{\delta}\|_I$. Then

$$
\begin{aligned}
\|\boldsymbol{\delta} + \mathbf{s}^g\|_1 &= \|\boldsymbol{\delta} + \mathbf{s}^g\|_{I-J} + \|\boldsymbol{\delta} + \mathbf{s}^g\|_{J-I} \\
&= \|\mathbf{s}^g\|_{I-J} + \|\boldsymbol{\delta}\|_{I^c} \\
&< \|\boldsymbol{\delta}\|_I + \|\mathbf{s}^g\|_{I-J} \\
&= \|\mathbf{s}\|_{I \cap J} + \|\mathbf{s}^g\|_{I-J} = \|\mathbf{s}^g\|_1 \quad (13)
\end{aligned}
$$

This shows that there exists a signal vector for which the estimated extended source vector is different to the true one both in directions and values.

REFERENCES

[1] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B, (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[2] P. Stoica and K.C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 7, pp. 1132 –1143, July 1990.

[3] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *Signal Processing Magazine, IEEE*, vol. 13, no. 4, pp. 67 –94, July 1996.

[4] J.-J. Fuchs, "Detection and estimation of superimposed signals," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, May 1998, vol. 3, pp. 1649 –1652 vol.3.

[5] D.M. Malioutov, M. Cetin, and A.S. Willsky, "Source localization by enforcing sparsity through a laplacian prior: an svd-based approach," *IEEE Workshop on Statistical Signal Processing*, pp. 573 – 576, sept.-1 oct. 2003.

[6] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1, Prentice Hall, 1993.

[7] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845 –2862, Nov. 2001.

[8] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing - Concepts and Techniques*, Prentice Hall, 1993.

[9] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version," http://cvxr.com/cvx, Jan. 2011.

[10] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2558 – 2567, sep 2002.