Fast Candidate Points Selection in the LASSO Path

Ashkan Panahi, *Student Member, IEEE*, and Mats Viberg, *Fellow, IEEE*

Abstract—The LASSO sparse regression method has recently received attention in a variety of applications from image compression techniques to parameter estimation problems. This paper addresses the problem of regularization parameter selection in this method in a general case of complex-valued regressors and bases. Generally, this parameter controls the degree of sparsity or equivalently, the estimated model order. However, with the same sparsity/model order, the smallest regularization parameter is desired. We relate such points to the nonsmooth points in the path of LASSO solutions and give an analytical expression for them. Then, we introduce a numerically fast method of approximating the desired points by a recursive algorithm. The procedure decreases the necessary number of solutions of the LASSO problem dramatically, which is an important issue due to the polynomial computational cost of the convex optimization techniques. We illustrate our method in the context of DOA estimation.

Index Terms—Homotopy, LARS, LASSO, linear regression, stagewise regression.

I. INTRODUCTION

THE Least Absolute Shrinkage and Selection Operator (LASSO) [1] has been developed as a computational tool to find sparse solutions of a linear regression problem. It has been used in an expanding field of applications from statistics ([2], [3]) to estimation scenarios ([4]) with remarkably good results. As a robust approximation of the well know Maximum Likelihood (ML) estimator, the LASSO can be realized robustly and efficiently through convex Second Order Cone (SOC) programming techniques [5], and unlike the efficient subspace methods [6], the LASSO technique is reliable even with one data measurement realization (snapshot). There has been also an extensive research on very fast enumeration techniques in real valued case such as [7] and [8].

The regularization parameter in LASSO is a mathematical tool to implement the compromise between the model fit and the estimated model order given by the number of nonzero parameters. As the regularization parameter evolves, the LASSO solution changes continuously, forming a continuous trajectory in a very high dimensional space which we will refer to as the *LASSO path*. However, in general, we are only interested in one estimate for each degree of sparsity. This gives a finite set of points. Thus, the goal is to determine a suitable sample point in the LASSO path for each level of sparsity. Following the ML rule, the candidate point normally corresponds to the smallest regularization parameter value with the given order. We call such a point a *desired singular point*. The reason for such a

terminology is that, as we will show, above mentioned desired points are among the nondifferentiable points in the path which we generally call the singular points.

In this paper, we give a homotopy method [9] to find the desired singular points by an iterative technique. We refer to this new method as *Singular Point Selection (SPS)-LASSO*. We first give a representation of the path by deriving a new optimality condition. The condition is actually a new representation of the Karush–Kuhn–Tucker KKT conditions in [9] using inequalities. This enables us to characterize the singular points in the path. Finally, we introduce an iterative method to choose the closest singular point starting from a higher value of the regularization parameter. The new representation also enables us to generalize the method to the case of complex numbers and even more general cases, which has not been possible by other techniques such as Least Angle Regression (LARS) [7]. We also address the continuous estimation case in which some undesired singular points may appear during the continuous evolution of the estimated parameter in the path. For this case we propose a slight modification of method by omitting the improper ones.

II. PROBLEM FORMULATION

Consider a vector $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T \in \mathbb{C}^m$ of m measurements. Suppose that we are to approximately represent $\mathbf{x}$ by the basis set given by the columns of an $m \times N$ dictionary matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$. We assume that N, $N > m$, is a large number of basis vectors in $\mathbf{A}$. In this case, the representation is not unique. However, we are only interested in those for which the coordinate vector has a small number of nonzero components. Such a vector and the index of its nonzero components may be referred to as a *sparse vector* and *active indexes* respectively.

For such a problem, the LASSO method has been developed as a linear least square optimization regularized by the ℓ_1 norm which is the sum of the absolute values of the components. Although different equivalent forms of the regularization have been proposed, we focus on the following basic formulation of the LASSO

$$\hat{\mathbf{s}}(\lambda) = \arg\min_{\mathbf{s}} \|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2 + \lambda\|\mathbf{s}\|_1 \qquad (1)$$

where the solution gives a sparse coordinate $\hat{\mathbf{s}}(\lambda)$. Note that the LASSO criterion function is a uniformly continuous function of λ. Thus, the LASSO path is also a continuous function of λ. However, due to the singularity of the magnitude function at zero this path is not smooth. During the gradual change of λ, the singularity happens when either a coordinate is grown from zero or disappears. This means that the LASSO path consists of some smooth pieces during which the active indexes do not change, because there is no coordinate creation or annihilation, connected at some singular points.

To form the candidate sampled set in the LASSO path we focus on the ones with the least error magnitude, $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda)\|_2$. It is not hard to see that these are the points with the smallest

regularization parameter λ, since the error magnitude is an increasing function of λ. We conclude that the desired points are given by some singular points in the path.

For the sake of simplicity, we introduce the following definitions and notations. For each solution, $\hat{\mathbf{s}}(\lambda)$ we define the active basis set, $I(\lambda)$ as the set of all indexes with nonzero elements i.e., $I(\lambda) = \{i \mid \hat{s}_i(\lambda) \neq 0\}$. We also denote the matrix of the basis vectors corresponding to the active indexes I by $\mathbf{A}_I$. If $I = \{i_1, i_2, \ldots, i_n\}$ then $\mathbf{A}_I = [\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \ldots, \mathbf{a}_{i_n}]$.

A. LASSO as an Estimation Tool

Many estimation problems, such as DOA and frequency estimations can also be solved by first discretizing the solution space and selecting the closest point to the true parameters. For such problems, LASSO can be used as a global matched filter [10] estimation tool [4]. In this cases, $\{\mathbf{a}_i\}_{i=1}^N$ is an indexed set of the points sampled from an r-dimensional manifold $\mathbf{a}(\theta)$ (e.g., array manifold). Due to the manifold's continuity, for a sufficiently fine grid the consecutive columns in $\mathbf{A}$ are close and the optimization in (1) can be interpreted as a discretized version of a "continuous LASSO" method, in which the active component indexes can vary continuously. Thus, we expect the LASSO path to mimic such a gradual change in the active set I. What actually happens is a fast annihilation of an active index and a simultaneous creation of a new neighbor, which approximates a continuous change in large scale. This shows that the LASSO path, in this case, contains many undesired close singular points. Later, we show how such singular points can be avoided.

III. THE LASSO SOLUTION

In this section we give a closed form solution for the LASSO optimization in (1) as follows.

Theorem 1: For a certain value of λ, if $\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda) \neq 0$, the solution to (1) is characterized by

$$\frac{\mathbf{a}_i^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}})}{\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}\|_2} = \lambda \frac{\hat{s}_i}{|\hat{s}_i|} \, i \in I, \frac{|\mathbf{a}_i^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}})|}{\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}\|_2} < \lambda \quad i \notin I. \tag{2}$$

The singular points occur when the second condition is changed to $\lambda = \max_{i \notin I}(|\mathbf{a}_i^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}})|)/(\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}\|_2)$.

Proof: The proof is given in Appendix A. ∎

Note that because the term $\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda)\|_2$ is a nondecreasing function of λ, the condition $\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda) \neq 0$ means that λ should be large enough. On the other hand, for a very large value of λ the solution will be $\hat{\mathbf{s}}(\lambda) = \mathbf{0}$. Then we can conclude the following statement.

Corollary 1: The first singular point, related to the first active base, is given by

$$\lambda_1 = \max_i \frac{|\mathbf{a}_i^H \mathbf{x}|}{\|\mathbf{x}\|_2}. \tag{3}$$

Once a singular point parameter λ_i is computed, the solution at that point can be found by solving the LASSO. This solution includes the active set and the waveforms at λ_i. However, the active set by definition does not change between two consecutive singular points. This means that for the next singular point the active set is known. The next singular point λ_{i+1} is then

given by the smallest λ satisfying (2) with the known active set I.

IV. RECURSIVE SINGULAR POINT SELECTION ALGORITHM

As we explain in this section, the singular points can be found by a recursive method of solving LASSO at a singular point and then predicting the next one. In this section we provide an approximate prediction of the next singular point. Later in Section V we show that this is a sufficiently good prediction for the iterative method with a low computational price. First, assume that for each set of active bases I, we can solve the first equation in (2) to obtain $\hat{\mathbf{s}}$ as a function of λ. Following Theorem 1, knowing I at λ_i, the next singular point λ_{i+1}, is given by

$$\lambda_{i+1} = \inf \left\{ \lambda \mid \forall i \notin I \quad \frac{|\mathbf{a}_i^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda))|}{\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda)\|_2} < \lambda \right\}. \tag{4}$$

For every index $j \notin I$, let us introduce $S_j = \{\lambda \mid (|\mathbf{a}_j^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda))|)/(\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda)\|_2) < \lambda\}$. In fact, S_j is an interval of the form $[\Lambda_j \, \infty]$, where Λ_j is the solution to the equation $(|\mathbf{a}_j^H(\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda))|)/(\|\mathbf{x} - \mathbf{A}\hat{\mathbf{s}}(\lambda)\|) = \lambda$. On the other hand, from (4) we have

$$\lambda_{i+1} = \inf \bigcap_j S_j = \max_{j \notin I} \Lambda_j. \tag{5}$$

Accordingly, the process of finding the next singular point can be summarized as follows.
1) The active set $I = \{i_1, i_2, \ldots, i_n\}$ is determined by solving LASSO at λ_i.
2) for each $j \notin I$ solve the following system of equations for a vector $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_n]$ and Λ_j.

$$\left\{ \frac{\mathbf{a}_{i_l}^H(\mathbf{x} - \mathbf{A}_I\hat{\mathbf{s}})}{\|\mathbf{x} - \mathbf{A}_I\hat{\mathbf{s}}\|_2} = \Lambda_j \frac{\hat{s}_l}{|\hat{s}_l|} \right\}_{l=1}^n, \frac{|\mathbf{a}_j^H(\mathbf{x} - \mathbf{A}_I\hat{\mathbf{s}})|}{\|\mathbf{x} - \mathbf{A}_I\hat{\mathbf{s}}\|_2} = \Lambda_j. \tag{6}$$

If the system is infeasible it means that $S_j = [0 \, \infty]$. Thus, in this case we put $\Lambda_j = 0$.
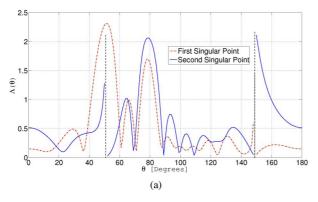3) The next singular point λ_{i+1} is given by $\lambda_{i+1} = \max_j \Lambda_j$.
The above steps give an iterative method to find the singular points. However, it needs a numerical method to solve the system of equations in (6) for each $j \notin I$. Note that the former $\hat{\mathbf{s}}(\lambda_i)$ which is obtained by solving LASSO at λ_i can be used as an initial guess for the $\hat{\mathbf{s}}$ parameters in (6). We introduce the following iterative root finding method. For convenience we introduce the function $R(a, b)$ of two complex numbers a, b as the positive root r of the equation $r = |a + rb|$.

Theorem 2: Starting from $\gamma_l^{(1)} = (\hat{s}_l(\lambda_i))/(|\hat{s}_l(\lambda_i)|)$ for $l = 1, 2, \ldots, n$, the following iterative equations may only converge to the true roots of the system of equations in (6) for each j.

$$\nu^{(k)} = R\left(\mathbf{a}_j^H P_{\mathbf{A}_I}^\perp \mathbf{x}, \, \mathbf{a}_j^H \mathbf{A}_I \left(\mathbf{A}_I^H \mathbf{A}_I\right)^{-1} \boldsymbol{\gamma}^{(k)}\right)$$

$$\hat{\mathbf{s}}^{(k)} = \left(\mathbf{A}_I^H \mathbf{A}_I\right)^{-1} \left(\mathbf{A}_I^H \mathbf{x} - \nu^k \boldsymbol{\gamma}^{(k)}\right)$$

$$\Lambda_j^{(k)} = \frac{\nu^{(k)}}{\left\| P_{\mathbf{A}_I}^\perp \mathbf{x} + \nu^{(k)} \mathbf{A}_I \left(\mathbf{A}_I^H \mathbf{A}_I\right)^{-1} \boldsymbol{\gamma}^{(k)} \right\|_2}$$
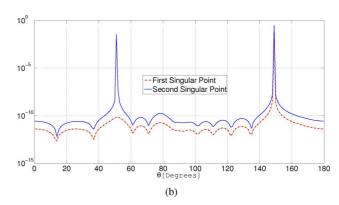
Fig. 2. Result of detecting 3 first singular points. Fig. (a) shows $\Lambda(\theta)$ for the first and second singular point prediction while Fig. (b) shows the LASSO solution at first and second singular points respectively. The vertical lines in (a) correspond to active indexes in (b).
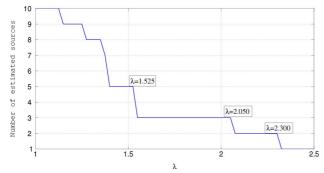

Fig. 1. Number of estimated sources VS the regularization parameter for a certain measurement $\mathbf{x}$. The discontinuity points are the singular ones.

$$\gamma_l^{(k+1)} = \frac{\hat{s}_l^{(k)}}{\left|\hat{s}_l^{(k)}\right|}. \tag{7}$$

Proof: The proof is given in Appendix B. ∎

A. The Iterative Method for LASSO Based Estimators

As we discussed in Section II-A, when LASSO is used as an estimation tool the LASSO path includes many undesired singular points. At each such point, the next one is very close due to a gradual change in the active set. This can be justified by the fact that for a large and fine grid, the values Λ_j are the samples of a continuous function $\Lambda(\boldsymbol{\theta})$ which can be found by substituting $\mathbf{a}_j$ by the general vector $\mathbf{a}(\boldsymbol{\theta})$ in (6). Thus, the index of maximum is close to an active index with an infinitesimally close value to λ_i. These undesired values can be avoided by first finding the peaks in $\Lambda(\theta)$ and then choosing the maximum peak value. Thus, the last step will be modified as

- Find the peak set $L = \{j \mid \Lambda_j > \Lambda_{j-1}, \ \Lambda_j > \Lambda_{j+1}\}$ and set $\lambda_{i+1} = \max_{j \in L} \Lambda_j$.

Note that we still assume that the active set does not change between two consecutive singular points, which may not be true anymore. However, it is still a good approximation as will be illustrated in the next section.

V. RESULTS

To examine the proposed method we apply the LASSO technique to the well known Direction Of Arrival (DOA) estimation problem. Such an application has been proposed and discussed widely ([11], [6]). Note that the problem for a linear array is identical to the frequency estimation problem. The DOA estimation problem for narrow-band, far-field sources is formulated by a linear regression model of the following form.

$$\mathbf{x} = \sum_i \mathbf{a}(\theta_i)s_i + \mathbf{n} \tag{8}$$

where $\theta_1, \theta_2, \ldots, \theta_n$ are the set of true DOAs and assuming a uniform linear array of m sensors spaced a half-wavelength apart, the $\mathbf{a}(\theta) = [1 \ e^{j\pi\cos\theta} \ e^{j2\pi\cos\theta} \ldots e^{j(m-1)\pi\cos\theta}]^T$ represents the steering vector corresponding to the direction θ. Further, s_i represents the signal value and $\mathbf{n}$ is additive noise. As explained in [4] the DOA parameters in (8) can be estimated using a LASSO optimization with $\mathbf{A} = [\mathbf{a}(\theta_1^g) \ \mathbf{a}(\theta_2^g) \ldots \mathbf{a}(\theta_N^g)]$, where $\{\theta_1^g, \theta_2^g, \ldots, \theta_N^g\}$ is a sample grid of the continuous DOA manifold $\theta \in [0 \ \pi]$.

We show the results of applying the proposed method to a LASSO-based DOA estimator with a certain realization of the measurement vector $\mathbf{x}$. For this realization $\|\mathbf{s}\|_2 = 1$, $\|\mathbf{n}\|_2 = .1$, and $m = 12$. The LASSO method is applied using the CVX toolbox [12]. Fig. 1 shows the true places of the singular points where new sources are introduced.

According to the steps in Section IV, we find the exact value of the first singular point corresponding to the creation of the first source by (3) which gives $\lambda_1 = 2.53$. Fig. 2(a) shows the solution at this point while the solution for any greater value is $\hat{\mathbf{s}}(\lambda) = \mathbf{0}$. Then, the next singular point is predicted through forming $\Lambda(\theta)$ by the method introduced in Theorem 2 and finding the greatest peak due to the modified step in Section IV-A. Fig. 2(a) shows the computed $\Lambda(\theta)$ for one iteration of the numerical method in Theorem 2. The next singular point is the greatest peak value in this figure ($\lambda_2 = 2.31$). Fig. 2(b) shows the solution at this point. The figure shows that a new source has just been created at this point. The method can be continued by the same procedure of predicting the next singular point. The new $\Lambda(\theta)$ is shown in Fig. 2(b) and the next predicted singular point is given by $\lambda = 2.06$. These results can be compared to the exact points in Fig. 1, which shows an excellent agreement between the introduced method and the true values.

VI. CONCLUDING REMARKS

In this paper we introduced a method to solve the LASSO regularization parameter selection problem by limiting the search to a finite set of singular points. The results show that the technique gives remarkably close approximations of the true candidate singular points even with one iteration of the numerical method in Theorem 2. The reason is that Theorem 2 is based on a consecutive update of the phases and amplitudes of the

LASSO solution. We observe that the phase component of the LASSO solution is almost constant between two consecutive singular points, so that any further phase update in the next iterations does not have a prominent effect on the numerical solution. Furthermore, the constant active index assumption in Section IV.A is acceptable, especially when LASSO underestimates the model order (first singular points). For highly overestimated model orders (the last singular points) the assumption does not hold anymore. However, such points are not desirable.

The numerical method can also be combined with the following LASSO optimization step so that the computational speed grows further. Note that the maximization in (5) not only gives the next singular point, but also its maximum index argument $\hat{j}$ is the newly created active index in the next continuous piece of the LASSO path. Furthermore, it can be simply seen that the solution to the numerical method in Theorem 2 actually gives the relaxed LASSO solution. This shows that one may modify the method so that it does not need for LASSO step. In the real case this gives a complexity of order Nmn^3. Note that n is typically very small. The optimization over directions can also be added in the continuous case. Finally, we remark that the singular-point set gives the best candidate solutions for different model orders. Finding this set by the proposed method, one can simply utilize a model order selection method such as the Akaike Information Criterion (AIC) [13] or [14] for the case that the degree of sparsity is not known.

APPENDIX A

A. Proof of Theorem 1

For a certain value of λ let us first introduce

$$V(\mathbf{s}) = \|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2 + \lambda\|\mathbf{s}\|_1. \tag{9}$$

Suppose at the global minimum the active set is given by I. Now, making an infinitesimal change ds_i at the ith element we have

$$dV(\mathbf{s}) = \begin{cases} -\frac{\mathrm{Re}\left(ds_i^* \mathbf{a}_i^H (\mathbf{x}-\mathbf{A}\mathbf{s})\right)}{\|\mathbf{x}-\mathbf{A}\mathbf{s}\|_2} + \lambda\frac{\mathrm{Re}(ds_i^* s_i)}{|s_i|} & i \in I \\ -\frac{\mathrm{Re}\left(ds_i^* \mathbf{a}_i^H (\mathbf{x}-\mathbf{A}\mathbf{s})\right)}{\|\mathbf{x}-\mathbf{A}\mathbf{s}\|_2} + \lambda|ds_i| & i \notin I \end{cases}. \tag{10}$$

The minimality condition is that the variations at this point should satisfy $dV \geq 0$ for every i and ds_i. Due to (10), if $i \in I$ the minimality condition is satisfied if and only if $-(\mathbf{a}_i^H (\mathbf{x} - \mathbf{A}\mathbf{s}))/(\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2) + \lambda(s_i)/(|s_i|) = 0 \quad i \in I$. On the other hand, if $i \notin I$ it is easy to see that the second row in (10) is alway nonnegative if and only if

$$\frac{|\mathbf{a}_i^H (\mathbf{x} - \mathbf{A}\mathbf{s})|}{\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2} < \lambda \quad i \notin I. \tag{11}$$

This forms the analytical LASSO solution in Theorem 1.

B. Proof of Theorem 2

The numerical method is to solve (6) by the following recursive relations.

$$\begin{aligned}
\frac{\mathbf{A}_I^H \left(\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)}\right)}{\left\|\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)}\right\|} &= \Lambda^{(k)} \boldsymbol{\gamma}^{(k)}, \\
\frac{\left|\mathbf{a}_j^H \left(\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)}\right)\right|}{\left\|\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)}\right\|} &= \Lambda^{(k)} \\
\left\{\gamma_l^{(k+1)} = \frac{\hat{s}_l^{(k)}}{\left|\hat{s}_l^{(k)}\right|}\right\}_{l=1}^{n} &.
\end{aligned} \tag{12}$$

Obviously, if the sequence $\{\mathbf{s}^{(k)}, \Lambda^{(k)}\}_{k=1}^{\infty}$ converges, the fixed point is the solution to (6). To solve the system of the two first equations in (12) we introduce $\nu^{(k)} = \|\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)}\|\Lambda^{(k)}$. From the first equation we have $\hat{\mathbf{s}}^{(k)} = (\mathbf{A}_I^H \mathbf{A}_I)^{-1}(\mathbf{A}_I^H \mathbf{x} - \nu^{(k)}\boldsymbol{\gamma}^{(k)})$. Then, we have

$$\begin{aligned}
\mathbf{x} - \mathbf{A}_I \hat{\mathbf{s}}^{(k)} &= \mathbf{x} - \mathbf{A}_I \left(\mathbf{A}_I^H \mathbf{A}_I\right)^{-1} \left(\mathbf{A}_I^H \mathbf{x} - \nu^{(k)}\boldsymbol{\gamma}^{(k)}\right) \\
&= \mathbf{P}_{\mathbf{A}_I}^{\perp} \mathbf{x} + \nu^{(k)} \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \boldsymbol{\gamma}^{(k)} \tag{13}
\end{aligned}$$

where $\mathbf{P}_{\mathbf{A}_I}^{\perp} = \mathbf{I} - \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1} \mathbf{A}_I^H$ is the orthogonal projection matrix of $\mathbf{A}_I$. From the second equation of (12) and (13) we get $\nu^{(k)} = |\mathbf{a}_j^H (\mathbf{P}_{\mathbf{A}_I}^{\perp} \mathbf{x} + \nu^{(k)} \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1}\boldsymbol{\gamma}^{(k)})|$ which gives the first equation of (7). Finally, by the definition of $\nu^{(k)}$ and (13) we get $\Lambda^{(k)} = (\nu^{(k)})/(\|\mathbf{P}_{\mathbf{A}_I}^{\perp} \mathbf{x} + \nu^{(k)} \mathbf{A}_I (\mathbf{A}_I^H \mathbf{A}_I)^{-1}\boldsymbol{\gamma}^{(k)}\|)$, which completes the proof.

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal Statist. Soc. B, (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[2] B. P. Meinshausen and N. , "High-dimensional graphs and variable selection with the LASSO," *Ann. Statist.*, vol. 34, no. 3, pp. 1436–1462, 2006.

[3] N. Simon, J. H. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for cox's proportional hazards model via coordinate descent," *J. Statist. Softw.* vol. 39, no. 5, pp. 1–13, 3, 2011 [Online]. Available: http://www.jstatsoft.org/v39/i05

[4] D. Malioutov, M. Cetin, and A. Willsky, "Source localization by enforcing sparsity through a Laplacian prior: An SVD-based approach," in *IEEE Workshop on Statistical Signal Processing*, Sept.–1 Oct. 2003, pp. 573–576.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

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

[7] B. Efron, T. Hastie, L. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.* vol. 32, pp. 407–499, 2004 [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.8168

[8] H. H. J. Friedman, T. Hastie, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, 2007.

[9] M. R. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–403, 2000.

[10] J.-J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 702–709, Apr. 2001.

[11] J.-J. Fuchs, "Detection and estimation of superimposed signals," in *Proc. 1998 IEEE Int. Conf. Acoustics, Speech and Signal Processing, 1998*, May 1998, vol. 3, pp. 1649–1652, vol. 3.

[12] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, Version 1.21 [Online]. Available: http://cvxr.com/cvx Apr. 2011

[13] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, 2004.

[14] A. Panahi and M. Viberg, "Maximum aposteriory based regularization parameter selection," in *Int. Conf. Acoustics, Speech, and Signal Processing*, 2011.