MAXIMUM A POSTERIORI BASED REGULARIZATION PARAMETER SELECTION

Ashkan Panahi, Mats Viberg

Chalmers University of Technology
Department of Signals and Systems
Email: {*ashkanp@ , viberg@*} *chalmers.se*

ABSTRACT

The ℓ_1 norm regularized least square technique has been proposed as an efficient method to calculate sparse solutions. However, the choice of the regularization parameter is still an unsolved problem, especially when the number of nonzero elements is unknown. In this paper we first design different ML estimators by interpreting the ℓ_1 norm regularization as a MAP estimator with a Laplacian model for data. We also utilize the MDL criterion to decide on the regularization parameter. The performance of these new methods are evaluated in the context of estimating the Directions Of Arrival (DOA) for the simulated data and compared. The simulations show that the performance of the different forms of the MAP estimator are approximately equal in the one snapshot case, where MDL may not work. But for the multiple snapshot case both methods can be used.

Index Terms— Linear regression, Sparse analysis, DOA estimation, LASSO, Model order selection

I. INTRODUCTION

The problem of estimating the Direction of Arrival (DOA) of the signals transmitted by a set of sources is a well studied problem. This estimation is based on the data received by a sensor array. It has also become an efficient tool in a variety of applications from radar detection to multi-user communication. Many different methods have been proposed for such a problem depending on the physical model of generating the data. Here, we focus on the far field model of signals received by a Uniform Linear Array (ULA). In this case, the methods such as Maximum Likelihood (ML) and subspace techniques have been in the center of research for a long time. However, it is shown recently that the so-called parameter selection method of LASSO [1] can be used as an estimation tool for the DOA estimation problem [2].

The Least Absolute Shrinkage and Selection Operator (LASSO), also known as *Basis Pursuit*, is originally a parameter selection technique for linear regression problems. Looking for a linear representation of a set of data by the least possible parameters, it is shown that solving a least square problem regularized by the ℓ_1 norm of the unknown parameters gives a proper solution with many zero parameters. In [2] and [4] it is shown that the narrowband model of a far-field signal can be expressed by a linear overcomplete expression with a sparse parameter space as a solution. This allows the LASSO method to be used for estimating the DOAs.

The ℓ_1 regularization in the LASSO method can be viewed as an approximation to the exact Maximum Likelihood (ML), which can be expressed by the ℓ_0 optimization. In fact, any regularization with l_p ($p < 1$) can be an approximation. However, with ℓ_1 we benefit from the convexity property. This may help us to implement the whole method with the fast and robust methods of convex optimization [2]. In [3] it is shown that for sufficiently sparse true parameters and in the noiseless case the solution of the ℓ_1 regularization is identical to this exact ℓ_0 regularization. As a regularized optimization, the LASSO criterion contains a regularization parameter controlling the importance of the shrinkage term (ℓ_1 norm) versus the error term. Note that for the problem of sparse linear estimation the goal is to minimize the error, while the number of active parameters is kept as small as possible. However, to decrease the error level, using more parameters is inevitable. In this case the regularization parameter controls the number of active parameters. The choice of this parameter has been the main difficulty in every implementation of the sparse estimation, especially when the true number of active parameters is unknown. A full discussion of the different technical aspects of this method is given in [4].

From one point of view, the LASSO regularization is a representation of the Maximum A posteriori Probability (MAP) assuming a Laplacian prior distribution over the data. Such a method has been studied in [5] and [6]. In [7] the Bayesian LASSO (BLASSO) method is introduced as an EM method of estimating the LASSO regularization parameter. Here, we first derive a different direct ML estimator by regarding the number of sources as a deterministic parameter. Next, we generalize the method for a different realization of the LASSO method. This results in a mismatched MAP estimator which we empirically show to give a better estimation in the case of one snapshot. From another point of view, choosing the regularization parameter is equivalent to the model order selection problem which has been studied extensively. Many methods such as General Likelihood Ratio Test (GLRT) [8], General Information Criterion [9], and Minimum Description Length [10] have been proposed for this purpose. The relation between these different methods has also been discussed in [8]. In [11] the application of different information criteria is also motivated and discussed. Here, we apply MDL to choose of the regularization parameter and compare the different methods. Although we presented the proposed approach in the context of DOA estimation, it is equally applicable to any sparse reconstruction problem.

II. SYSTEM MODEL

Consider an array with m receiving elements arranged in a Uniform Linear Array (ULA), and a scenario with n objects that transmit energy from angles $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$. As described in [12], assuming narrowband signals arriving from the far-field, the complete received data set $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T \in \mathbb{C}^m$ over the ULA can be modeled as

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

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_n)]$ and $\mathbf{a}(\theta_i)$ is the steering vector describing the signal phase shift of the source i at each antenna. Further, $\mathbf{n}(t)$ is an additive noise vector, which is assumed to be zero mean Gaussian with the covariance matrix $\sigma^2\mathbf{I}$. This noise term describes the receiver noise, deviations from the signal model and other unmodeled phenomena.

II-A. One Snapshot Case

Assume that the directions are to be estimated from a finite (but sufficiently large) grid, $G = \{\frac{\pi}{N}, \frac{2\pi}{N}, \ldots\}$, of all possible directions. Denoting the matrix of all possible steering vectors by $\mathbf{A}^g = \left[\mathbf{a}(\frac{\pi}{N}), \mathbf{a}(\frac{2\pi}{N}), \ldots\right]$, (1) can be written as

$$\mathbf{x}_{m\times 1} = \mathbf{A}^g_{m\times N}\mathbf{s}^g_{N\times 1} + \mathbf{n}_{m\times 1} \quad , \tag{2}$$

where $\mathbf{x}$, $\mathbf{n}$ are short notations of $\mathbf{x}(1)$, $\mathbf{n}(1)$ respectively, and $\mathbf{s}^g$ is the sparse source vector consisting of the true source components of $\mathbf{s}(1)$ at the indexes corresponding to $\theta_1, \theta_2, \ldots, \theta_n$ and zero everywhere else. Since (2) does not possess a unique solution, the task is to minimize the number of nonzero elements ($= \|\mathbf{s}^g\|_0$) while the magnitude of the noise component, $\|\mathbf{n}\|_2 = \|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2$ is small enough. This is a regularization problem, and it can be formulated in many equivalent forms. The following form is often used.

$$\mathscr{L}(\lambda) = \min_{\mathbf{s}^g} \|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2 + \lambda\|\mathbf{s}^g\|_0 \quad , \tag{3}$$

where λ is a proper regularization parameter. Choosing λ is discussed in more details in Section IV. It controls the number of detected active basis vectors in (3). The greater the value of λ, the smaller $\|\mathbf{s}\|_0$ is chosen by (3). After estimating the complete source vector, $\mathbf{s}^g$, one can form a vector of direction parameters, $\theta(\mathbf{s}) = [\theta_1, \theta_2, \ldots, \theta_n] \subset G$, corresponding to the nonzero elements of $\mathbf{s}$ known as the *active basis*,

$$\boldsymbol{\theta}(\mathbf{s}^g) = \{\theta_i | s^g_i \neq 0, i = 1, \ldots, N\} \quad . \tag{4}$$

II-B. Multiple Snapshots Case

The model of (2) can be easily extended to the case of many snapshots due to the linear model of (1). The sequence of received data, $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(T)]$, and the transmitted one, $\mathbf{S}^g = [\mathbf{s}^g(1), \mathbf{s}^g(2), \ldots, \mathbf{s}^g(T)]$, are related by a linear model

$$\mathbf{X} = \mathbf{A}^g\mathbf{S}^g + \mathbf{N} \quad , \tag{5}$$

where $\mathbf{N}$ is a sequence of independent noise vectors with variance σ^2. The matrix $\mathbf{A}^g$ is the same dictionary matrix as in (2). In this case the active basis indexes can be defined as follows:

$$\boldsymbol{\theta}(\mathbf{S}^g) = \{\theta_i | \exists t, s^g_i(t) \neq 0\} \quad , \tag{6}$$

which means that a certain base is generally active if it is active at least in one snapshot. Denoting the number of elements of $\boldsymbol{\theta}(\mathbf{S})$ by $n(\mathbf{S})$ the goal is to minimize $n(\mathbf{S})$ while the total noise, $\sum_{t=1}^{T} \|\mathbf{n}(t)\|_2^2$, is kept as small as possible. Note that minimizing $n(\mathbf{S})$, the optimization has a tendency to choose the same basis vectors in different snapshots. Therefore, this algorithm is suitable wherever the actual basis vectors do not change by time. Now the regularized criterion can be formed as follows:

$$\mathscr{L}_S(\lambda) = \min_{\mathbf{S}} \|\mathbf{X} - \mathbf{A}^g\mathbf{S}\|_2 + \lambda n(\mathbf{S}) \quad . \tag{7}$$

III. ℓ_1 REGULARIZATION SOLUTION

Because of the complexity of solving (3) and (7) as ℓ_0 regularization problems, it is proposed to use the ℓ_1 norm which gives a sparse solution as well. Furthermore, because of the convexity of the ℓ_1 norm the optimization can be performed in relatively low computational time using convex optimization methods. The following criterion results from substituting ℓ_0 by the ℓ_1 norm in (3).

$$\mathscr{K}(\lambda) = \min_{\mathbf{s}^g} \|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2 + \lambda\|\mathbf{s}^g\|_1 \quad , \tag{8}$$

or

$$\mathscr{K}_2(\lambda) = \min_{\mathbf{s}^g} \|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2^2 + \lambda\|\mathbf{s}^g\|_1 \quad , \tag{9}$$

where $\|\mathbf{s}^g\|_1 = \sum_{i=1}^{N} |s^g_i|$. There are also many different equivalent regularizations. We here only introduced the most popular versions. For the case of multiple snapshots, in [3] it is proposed to use the following regularization instead of (7)

$$\mathscr{K}_S(\lambda) = \min_{\mathbf{S}^g} \|\mathbf{X} - \mathbf{A}^g\mathbf{S}^g\|_2 + \lambda \sum_{i=1}^{N} \sqrt{\sum_{t=1}^{T} s_i^2(t)} \quad . \tag{10}$$

It is observed that (9) and (10) give results for $\mathbf{s}$ different from (3) and (7). However, as discussed in [2], the indexes of the nonzero elements in (3) and (9) are likely to be identical. In summary, although (9) does not give a good estimation of the sparse vector $\mathbf{s}^g(t)$, it is a very good estimator of $\boldsymbol{\theta}(\mathbf{s}^g(t))$. In practice, the numerical optimization methods introduce a small noise to the solution which decreases the number of nonzero elements. Thus a thresholding method is needed to decide on the set of active indexes. However, because of the small magnitude of the computational error and a great difference between the values of active and non-active components, choosing a proper threshold is straightforward in most practical cases.

Both techniques introduced in (9) and (10) give a consistent estimation of the active basis. Knowing the index of the nonzero elements of the unknown source vector, $\mathbf{A}(\theta)$ and $\mathbf{s}(\theta)$ in (1) can be found as a sub matrix and sub vector of $\mathbf{A}^g$ and $\mathbf{s}^g$ respectively. For a small number of objects, $\mathbf{A}(\theta)$ will be a full column rank matrix, and from (5) an improved estimation of the nonzero elements of $\mathbf{S}$ can be given by

$$\hat{\mathbf{S}}(\theta(\mathbf{S})) = \mathbf{A}^{\dagger}(\theta)\mathbf{X} \quad . \tag{11}$$

where $\mathbf{A}^{\dagger}(\theta)$ is the pseudo inverse of $\mathbf{A}(\theta)$. The procedure in (11) is often referred to as a *debiasing* step as introduced in [13]. Accordingly, we consider the following overall procedure for estimating the sources.

1) The active basis indexes are estimated using (10). The regularization parameter should be chosen properly. We choose it using a model order selection criterion explained in Section IV.
2) Using the estimated active basis indexes, $\hat{\theta}$, the sources are estimated by (11).

IV. REGULARIZATION PARAMETER SELECTION

IV-A. Maximum A-posteriori Probability interpretation

Now, we introduce a specified technique of selecting the regularization parameter by interpreting (9) as a Maximum A-posteriori Probability (MAP) estimation assuming the following prior distribution to the data.

$$f_{\mathbf{S}^g}(\mathbf{s}^g) = \begin{cases} \left(\frac{\mu^2}{2\pi}\right)^n e^{-\mu\|\mathbf{s}^g\|_1} & \text{the active directions of } \mathbf{s}^g \text{ is } \boldsymbol{\theta} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Where $\boldsymbol{\theta} \subset G$ is the set of true active basis directions and n is the number of elements of $\boldsymbol{\theta}$, which is also the number of sources. Then,

$$\begin{aligned} -\ln f_{\mathbf{X},\mathbf{S}^g}(\mathbf{x}, \mathbf{s}^g; \mu, \boldsymbol{\theta}) &= \frac{\|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2^2}{\sigma^2} + \mu\|\mathbf{s}^g\|_1 \\ &+ m\ln\pi\sigma^2 - n\ln\frac{\mu^2}{2\pi} \quad . \end{aligned} \tag{13}$$

In this case the ML estimation of the unknown variable μ is given by minimizing (13), which can be written as follows

$$\begin{aligned} \hat{\mu} = \arg\min_{\mu} \min_{\mathbf{s}^g} \frac{\|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2^2 + \lambda\|\mathbf{s}^g\|_1}{\sigma^2} + m\ln\sigma^2 \\ -n\ln\frac{\lambda^2}{2\sigma^4\pi} \quad , \end{aligned} \tag{14}$$

where we introduce $\lambda = \sigma^2\mu$ to take advantage of the definition (9). Note that the term $-n\ln\frac{\lambda^2}{2\sigma^4\pi}$ is also a function of $\mathbf{s}^g$. This can be illustrated by substituting $n = \|\mathbf{s}^g\|_0$. However, as a first

order approximation it is treated as a constant while minimizing (14) with respect to $\mathbf{s}^g$. The claim is that the coefficient $\ln \frac{\lambda^2}{2\pi}$ is not strong enough to make a jump in the estimated number of sources n. Furthermore, note that the other unknown parameters can be substituted by a robust estimation. As we discussed before, for the active directions $\boldsymbol{\theta}$, and consequently $n = \|\boldsymbol{\theta}\|_0$ solving the LASSO optimization in (9) gives a robust estimation, $\hat{\boldsymbol{\theta}}(\lambda)$ and $\hat{n}(\lambda)$. However, the noise level σ^2, can not be estimated directly by minimizing (14), because the LASSO method is not a robust estimator of the waveforms, $\mathbf{s}$. Instead, we propose an ML estimator using the source estimation in (11)

$$\hat{\sigma}^2(\lambda) = \frac{\|\mathbf{x} - \mathbf{A}(\hat{\theta})\hat{\mathbf{s}}\|_2^2}{m} = \frac{\|\mathbf{P}^{\perp}_{\mathbf{A}(\hat{\theta})}\mathbf{x}\|_2^2}{m} \quad . \tag{15}$$

Substituting this result to the new Log likelihood function and using the definition (9) we achieve the following result as an estimator of λ.

$$\hat{\lambda} = \arg\min_{\lambda} \frac{\mathscr{K}(\lambda)}{\hat{\sigma}^2(\lambda)} + (m + 2\hat{n}(\lambda)) \ln \hat{\sigma}^2(\lambda) - \hat{n}(\lambda) \ln \frac{\lambda^2}{2\pi}. \tag{16}$$

In the multiple snapshot case the result can be written as

$$\hat{\lambda} = \arg\min_{\lambda} \frac{\mathscr{K}(\lambda)}{\hat{\sigma}(\lambda)} + 2(m+n)T \ln \hat{\sigma} - nT \ln \frac{K(T)\lambda^2}{2\pi} \quad , \tag{17}$$

where

$$K(T) = \frac{1}{\sqrt[T]{1 \times 3 \times \ldots \times (2T-1)}} \approx \frac{1}{T} \tag{18}$$

is a normalization factor for likelihood function and

$$\hat{\sigma}^2(\lambda) = \frac{\|\mathbf{X} - \mathbf{A}(\hat{\theta})\hat{\mathbf{S}}\|_2^2}{mT} = \frac{\|\mathbf{P}^{\perp}_{\mathbf{A}(\hat{\theta})}\mathbf{X}\|_2^2}{mT} \quad . \tag{19}$$

IV-B. Mismatched MAP estimator

The exact MAP estimator introduced in the last section is closely related to the LASSO realization in (9). However, for many applications the LASSO form in (8) is prefered. To get an estimator for the regularization parameter in this case we have to neglect the Gaussian model of the noise and reconstruct the model based on the LASSO form. This might be a good choice since with one snapshot it is easy to reinterpret the data. In this case we can assume that

$$f_{\mathbf{N}}(\mathbf{n}) \propto e^{-\alpha\|\mathbf{n}\|_2} \quad . \tag{20}$$

Then, the mismatched log-likelihood function can be written as

$$\begin{aligned} -\ln f_{\mathbf{X},\mathbf{S}^g}(\mathbf{x}, \mathbf{s}^g; \mu, \boldsymbol{\theta}) &= \alpha\|\mathbf{x} - \mathbf{A}^g\mathbf{s}^g\|_2 + \mu\|s\|_1 \\ &- 2m \ln \frac{\alpha}{2\pi} - n \ln \frac{\mu^2}{2\pi}. \end{aligned} \tag{21}$$

Introducing $\lambda = \frac{\mu}{\alpha}$ and following the same procedure as in (16) we find out the following result as an estimator of the regularization parameter in (8)

$$\begin{aligned} \hat{\lambda} = \arg\min_{\lambda} \mathscr{K}_2(\lambda)\hat{\alpha}(\lambda) &- (m + 2\hat{n}(\lambda)) \ln \hat{\alpha}(\lambda) \\ &- \hat{n}(\lambda) \ln \frac{\lambda^2}{2\pi} \quad , \end{aligned} \tag{22}$$

where

$$\hat{\alpha}(\lambda) = \frac{m}{\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2} \quad . \tag{23}$$

The new criteria (17) and (22) constitute the main contribution of this paper.

IV-C. Minimum Description Length

The Minimum Description Length principle (MDL) is developed as a rule for measuring the fitness of a model to the received data [10]. The MDL criterion for a certain model order and data in (1) can be approximated as ([14]),

$$L(x) = mT \ln \sigma^2 + \frac{\|\mathbf{X} - \mathbf{A}(\theta)\mathbf{S}\|_2^2}{\sigma^2} + \frac{1}{2} nT \ln (mT) \quad . \tag{24}$$

If the noise variance is not known a robust estimation of it is used. This estimation can be obtained by maximizing the Log Likelihood function. The result for (24) will be

$$L(x) = m \ln \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2 + n \ln mT \quad . \tag{25}$$

The criterion for one snapshot case can be found by putting $T = 1$.

V. NUMERICAL RESULTS

The algorithm introduced in Section III is applied to the simulated data. First the performance of the direct MAP criterion in section IV is compared to the mismatched one. It is shown that the mismatched method gives approximately the same results but at lower computational cost. Then the mismatched one is compared to MDL for varying number of snapshots. In all simulation cases the number of the receiving sensors is equal to 8.

To discuss the performance of different parameter estimation methods, we use zero mean independent Gaussian random variables with variance one as the sources. The number and the directions of the sources are fixed, and the probability of success in detecting the number of sources is computed by different criteria for many different realizations of sources.

Figure 1 shows the result of applying the mismatched MAP to a one snapshot data of 8 sensors, which is generated randomly with different number of sources. As can be seen the performance of all methods decrease as the number of sources grows. With the MAP criterion it is more likely to detect the number of sources correctly. However, the probability of correctness decreases as the noise power increases. Figure 1 also shows the result for the direct MAP. It should be noted that because of the rapid change in the a-posteriori probability with λ, the simulation needs much more time for direct MAP because a higher resolution of λ is needed. The performance of the mismatched MAP criterion is almost the same as that of the direct MAP.

Figure 2 is related to the detection of the number of sources for different number of snapshots for a low SNR of 0 dB. As expected, the performance of both criteria increases for increasing number of snapshots as well as increasing SNR. However, the MAP approach can work properly with less number of snapshots or lower SNRs.

VI. CONCLUSION

In this paper, the problem of finding the regularization parameter in the LASSO-based DOA estimator is discussed. We introduce two methods by first considering LASSO as a Bayesian estimation method, and then by interpreting the problem as a model order selection one. The results show that the Bayesian interpretation may lead to better results as compared to MDL. Next we showed that the more popular form of (8) can also be used in a Bayesian setting, although it leads to a mismatched interpretation of the data. It might be beneficial, because for the LASSO form of (10) the DOA estimation is unaffected over a long interval of the regularization parameter. This means that the method can be performed with less resolution on this parameter, which results in less computational time. For a large number of snapshots case there is no remarkable difference between the introduced methods, and also MDL yields a satisfactory performance.

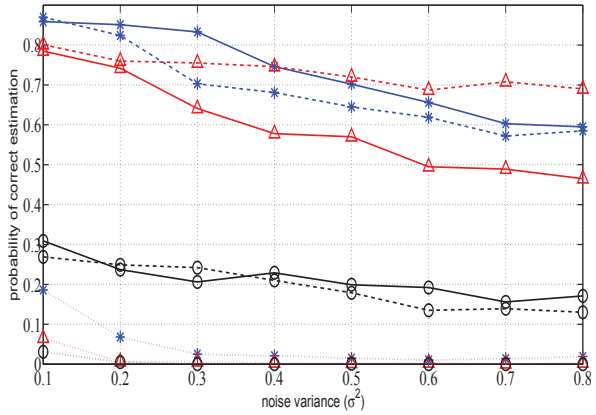

Fig. 1. Probability of correct number of signals estimation v.s. the noise variance for MDL and mismatched MAP. The star, triangle, and circle marks are related to 1,2, and 3 sources respectively. The solid, dashed and dotted line styles are related to mismatched MAP, direct MAP, and MDL respectively.

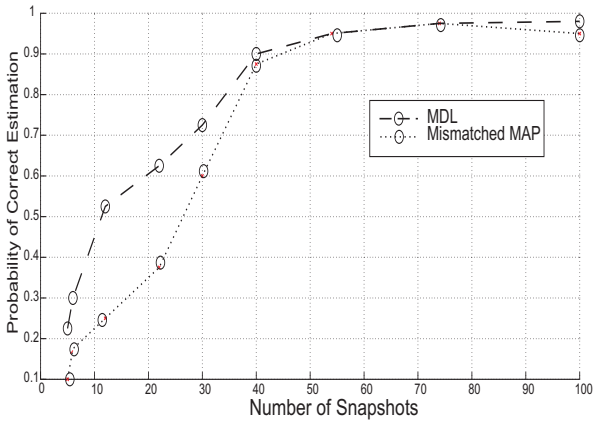

Fig. 2. Probability of correct number of signals estimation v.s. the number of snapshots for the mismatched MAP and MDL. SNR=0 dB

VII. 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] 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.

[3] 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.

[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] M.A.T. Figueiredo, "Adaptive sparseness for supervised learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1150 – 1159, sep. 2003.

[6] M. Yuan and Y. Lin, "Efficient empirical bayes variable selection and esimation in linear models," *Journal of the American Statistical Association*, vol. 100, pp. 1215 1225, 2005.

[7] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, pp. 681 686, 2008.

[8] P. Stoica, Y. Selen, and Jian Li, "On information criteria and the generalized likelihood ratio test of model order selection," *Signal Processing Letters, IEEE*, vol. 11, no. 10, pp. 794 – 797, oct. 2004.

[9] H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Proc. 2nd Int. Syrnp. IriJbrrn. Theory*, pp. 267–281, 1973.

[10] J. Rissanen, "Information and complexity in statistical modeling," *Information Theory Workshop, ITW Punta del Este. IEEE*, pp. 351 –351, march 2006, 10.1109/ITW.2006.1633845.

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

[12] M. Buhren and B. Yang, "Simulation of automotive radar target lists using a novel approach of object representation," *IEEE Intelligent Vehicles Symposium (IV),Tokyo, Japan*, June 2006.

[13] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586 –597, 2007.

[14] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the mdl principle," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 8, pp. 1190 –1196, aug 1989.