WEIGHTED COVARIANCE MATCHING BASED SQUARE ROOT LASSO

Arash Owrang and Magnus Jansson

ACCESS Linnaeus Centre, School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm Sweden

ABSTRACT

We propose a method for high dimensional sparse estimation in the multiple measurement vector case. The method is based on the covariance matching technique and with a sparse penalty along the ideas of the square-root LASSO (sr-LASSO). The method not only benefits from the strong characteristics of sr-LASSO (independence of the hyper-parameter selection from the noise variance), but also offers a performance near maximum likelihood. It performs close to the Cramer-Rao bound even at low signal to noise ratios and it is generalized to manage correlated noise. The only assumption in this matter is that the noise covariance matrix structure is known. The numerical simulation provided in an array processing application illustrates the potential of the method.

***Index Terms*—** sparse estimation, multiple measurements, correlated noise, covariance matching

1. INTRODUCTION

Estimation problems, such as direction of arrival (DOA) estimation with sensor arrays [1] and channel estimation [2], are traditionally handled by dividing it into two tasks, the detection of number of signals and the estimation of signals. A precise detection of the number of signals [3] is the key to the accurate estimation of signals. Either over detection or under detection would cause large estimation error or overfitting. Alternatively, the whole framework can be unified as a sparse estimation problem, where the number of signals, the support set and the signal itself are estimated simultaneously. This under the condition of having an underdetermined system of equations with sparse solutions as in compressed sensing. Interestingly, solving an optimization problem under some assumption answers to the detection and estimation subproblems at once. The optimization problem is not unique, actually different methods consider rather different objectives [4–6]. Despite the elegance of these methods, they all suffer from two issues: dependency of the selection of hyper-parameter on the knowledge of noise variance, and inaccuracy of the estimation due to low signal to noise ratios (SNRs). These issues bring impracticality to sparse estimators when they face real world applications. Belloni et al. [7] proposed a method, sr-LASSO, to overcome the hyper-parameter dependency issue. However, it is still vulnerable to strong perturbation. For applications with multiple measurements, Stoica et al. proposed a free from hyper-parameter selection method, called SPICE [8]. SPICE utilizes the covariance matching technique [9] to handle the low SNR condition and it only considers uncorrelated noise. It turned out that SPICE under the assumption of white noise is indeed a sr-LASSO with a particular regularization parameter [10, 11]. Further investigation for accurate estimation of signals in the presence of strong contamination, independent and correlated, is under demand.

Inspired by covariance matching estimation techniques (COMET) [9], we propose a new method based on sr-LASSO which is well-performing in poor SNR situations. Another strong aspect of our method is its capability to manage correlated noise. Here, we only assume that the structure of the noise covariance matrix is known. A DOA estimation problem is considered in Section 3. Our method is compared with multiple, both standard and recently proposed, methods as well as the Cramer-Rao bound (CRB). The simulation results reveal that our method achieves the CRB in a wide range of SNRs for a reasonably low number of measurements.

This work was partially supported by the Swedish Research Council under contract 621-2011-5847.

2. PROBLEM FORMULATION AND METHOD

2.1. Data Model and Problem Formulation

Let $\mathbf{y}(t) \in \mathbb{C}^m$ represent the complex measurement vector obtained from the following model

$$\mathbf{y}(t) = \mathbf{\Phi}\mathbf{x}(t) + \mathbf{n}(t), \qquad t = 1, 2, \ldots, M. \tag{1}$$

Here, $\mathbf{x}(t) \in \mathbb{C}^N$ is the unknown vector parameter which is sparse such that $\|\mathbf{x}(t)\|_0 = k < m \ll N$. The $\|.\|_0$ norm denotes the cardinality of the support set of $\mathbf{x}(t)$ defined as $\mathcal{I} = \{i | x_i \neq 0\}$. It is assumed that signal instances $\{\mathbf{x}_{\mathcal{I}}(t)\}_{t=1,2,\ldots}$ are spatially and temporally independent and identically distributed (i.i.d) complex random variables with zero mean, and the support set $\mathcal{I}$ does not vary over time. The noise $\{\mathbf{n}(t)\}_{t=1,2,\ldots} \in \mathbb{C}^m$ is temporally i.i.d complex Gaussian distributed $\mathbf{n}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n)$, and assumed to be

uncorrelated with the signal $\mathbf{x}(t)$. The known sensing matrix $\mathbf{\Phi} \in \mathbb{C}^{m \times N}$ has full row rank.

Our main interest herein is to estimate the support set $\mathcal{I}$ from M number of measurements given the data model (1). The estimation proceeds under the assumption that $\mathbf{R}_n$ is linearly parameterized by some unknown parameters.

2.2. Background and Method

Straightforward calculation for the data model (1) shows that

$$\mathbf{R} \equiv \mathbb{E}(\mathbf{y}(t)\mathbf{y}^H(t)) = \mathbf{\Phi}\mathbf{P}\mathbf{\Phi}^H + \mathbf{R}_n, \tag{2}$$

where $\mathbb{E}(.)$ denotes the statistical expectation, $.^H$ denotes the matrix conjugate transpose and $\mathbf{P} = \mathbb{E}(\mathbf{x}(t)\mathbf{x}^H(t))$ is a diagonal matrix. Since $\mathbf{x}(t)$ is sparse, the diagonal of $\mathbf{P}$ is also sparse. Let the operator $\mathrm{vec}(.)$ stack the columns of a matrix into a vector, $\mathbf{r} \equiv \mathrm{vec}(\mathbf{R})$. Then vectorizing both sides of (2) yields

$$\begin{aligned}\mathbf{r} &= \mathrm{vec}(\mathbf{\Phi}\mathbf{P}\mathbf{\Phi}^H + \mathbf{R}_n) \\ &= (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\mathrm{vec}(\mathbf{P}) + \mathbf{r}_n = (\mathbf{\Phi}^* \circ \mathbf{\Phi})\mathbf{p} + \mathbf{r}_n,\end{aligned}$$

where $.^*$ is the matrix conjugate, $\otimes$ and $\circ$ are the Kronecker product and the Khatri-Rao product, respectively, $\mathbf{r}_n = \mathrm{vec}(\mathbf{R}_n)$ and $\mathbf{p} = \mathrm{diag}(\mathbf{P})$.

It is known that under weak conditions the maximum likelihood estimator is asymptotically efficient, but highly nonlinear. An alternative to the maximum likelihood estimator is COMET [9], also known as the generalized least squares [12]. COMET minimizes the weighted distance between the model $\mathbf{R}$ and the sample covariance matrix $\hat{\mathbf{R}}$, $\hat{\mathbf{R}} \equiv 1/M \sum_{t=1}^{M} \mathbf{y}(t)\mathbf{y}^H(t)$. The motivation for the weighting comes from the desire to emphasize on data that are more reliable (corrupted with less noise). An optimal choice for the weighting is the inverse of the covariance matrix of the residual.

COMET relies on the following objective function

$$g(\mathbf{P}, \mathbf{R}_n) = \left\| \hat{\mathbf{R}}^{-1/2} \left(\hat{\mathbf{R}} - \mathbf{R} \right) \hat{\mathbf{R}}^{-1/2} \right\|_F^2, \tag{3}$$

where $\hat{\mathbf{R}}^{-1/2}$ is the Hermitian matrix square root of $\hat{\mathbf{R}}^{-1}$. Further derivations reveal the relation with weighted least squares:

$$\begin{aligned} g(\mathbf{P}, \mathbf{R}_n) &= \mathrm{tr}\left(\hat{\mathbf{R}}^{-1}(\hat{\mathbf{R}} - \mathbf{R})\hat{\mathbf{R}}^{-1}(\hat{\mathbf{R}} - \mathbf{R}) \right) \\ &= \mathrm{vec}^H\left((\hat{\mathbf{R}} - \mathbf{R})\hat{\mathbf{R}}^{-1} \right) \mathrm{vec}\left(\hat{\mathbf{R}}^{-1}(\hat{\mathbf{R}} - \mathbf{R}) \right) \\ &= (\hat{\mathbf{r}} - \mathbf{r})^H \mathbf{\Gamma} (\hat{\mathbf{r}} - \mathbf{r}) \end{aligned} \tag{4}$$

where tr denotes the trace of a matrix and $\mathbf{\Gamma}$ is

$$\mathbf{\Gamma} \equiv \hat{\mathbf{R}}^{-*} \otimes \hat{\mathbf{R}}^{-1} \approx \frac{1}{M} \left(\mathrm{Cov}\left(\hat{\mathbf{r}} - \mathbf{r} \right) \right)^{-1}.$$

We first want to minimize $g(\mathbf{P}, \mathbf{R}_n)$ with respect to the noise covariance matrix $\mathbf{R}_n$, for a fixed $\mathbf{P}$. As $\mathbf{R}_n$ is assumed to be linearly parameterized we can write $\mathbf{r}_n = \mathbf{M}\boldsymbol{\beta}$. Elements of $\mathbf{M}$ are zero, one and the imaginary unit ($\pm$j) and $\boldsymbol{\beta}$ contains the real unknown parameters of the noise covariance. The reader should keep in mind that $\mathbf{M}$ is a known matrix as the structure of $\mathbf{R}_n$ is assumed to be known. Then,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} g(\mathbf{P}, \mathbf{R}_n(\boldsymbol{\beta})) = \left(\mathbf{M}^H \mathbf{\Gamma} \mathbf{M} \right)^{-1} \mathbf{M}^H \mathbf{\Gamma} \tilde{\mathbf{r}}_n \tag{5}$$

where $\tilde{\mathbf{r}}_n \equiv \mathrm{vec}(\tilde{\mathbf{R}}_n)$ and $\tilde{\mathbf{R}}_n \equiv \hat{\mathbf{R}} - \mathbf{\Phi}\mathbf{P}\mathbf{\Phi}^H$. We rewrite (4) as

$$g(\mathbf{P}, \mathbf{R}_n) = \left\| \mathbf{\Gamma}^{1/2} \left(\tilde{\mathbf{r}}_n - \mathbf{r}_n \right) \right\|_2^2, \tag{6}$$

and replace $\mathbf{r}_n$ by its estimate $\hat{\mathbf{r}}_n = \mathbf{M}\hat{\boldsymbol{\beta}}$. This gives

$$g(\mathbf{P}, \hat{\mathbf{R}}_n) = \left\| \mathbf{\Pi}^{\perp} \mathbf{\Gamma}^{1/2} \left(\hat{\mathbf{r}} - (\mathbf{\Phi}^* \circ \mathbf{\Phi})\mathbf{p} \right) \right\|_2^2, \tag{7}$$

where

$$\mathbf{\Pi}^{\perp} \equiv \mathbf{I} - \mathbf{\Gamma}^{1/2}\mathbf{M} \left(\mathbf{M}^H \mathbf{\Gamma} \mathbf{M} \right)^{-1} \mathbf{M}^H \mathbf{\Gamma}^{1/2}.$$

Matrix $\mathbf{\Pi}^{\perp}$ is a projection matrix and its eigenvalues are either one or zero. Hence, eigen-decomposition factorizes it to the canonical form of

$$\mathbf{\Pi}^{\perp} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}_{\Lambda}\mathbf{U}_{\Lambda}^H.$$

Here, the columns of $\mathbf{U}_{\Lambda}$ are the normalized eigenvectors corresponding to the non-zero eigenvalues. Exploiting the compact eigen-decomposition into function g in (7) reduces the dimensions of the problem to

$$\begin{aligned} g(\mathbf{P}, \hat{\mathbf{R}}_n) &= \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{\Phi}}\mathbf{p} \right\|_2^2, \\ \tilde{\mathbf{r}} &\equiv \mathbf{U}_{\Lambda}^H \mathbf{\Gamma}^{1/2} \hat{\mathbf{r}}, \\ \tilde{\mathbf{\Phi}} &\equiv \mathbf{U}_{\Lambda}^H \mathbf{\Gamma}^{1/2} (\mathbf{\Phi}^* \circ \mathbf{\Phi}). \end{aligned} \tag{8}$$

The final step to estimate $\mathbf{p}$ can be seen to be a classical sparse regression problem. sr-LASSO is a recently proposed sparse estimator, which provides a practical way to select the required regularization parameter. Unlike most methods [4, 5], sr- LASSO does not need the knowledge about the noise variance for this selection as will be discussed in some detail later in this section. Before introducing the sr-LASSO, we define a diagonal matrix $\mathbf{W}$ such that $w_{ii} = \|\tilde{\boldsymbol{\phi}}_i\|_2$ to normalize the length of the columns of the new sensing matrix $\tilde{\mathbf{\Phi}}$. Equalized contribution of the columns is essential to sr-LASSO. Finally, we suggest the following weighted sr-LASSO (wsr-LASSO) to estimate $\mathbf{p}$, or rather detect the support set $\mathcal{I}$:

$$\min_{\tilde{\mathbf{p}} \geq \mathbf{0}} \frac{\lambda}{\sqrt{m}} \|\tilde{\mathbf{p}}\|_1 + \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{\Phi}}\mathbf{W}^{-1}\tilde{\mathbf{p}} \right\|_2 \tag{9}$$

where $\tilde{\mathbf{p}} \equiv \mathbf{W}\mathbf{p}$.

The accuracy of the estimation by (9) depends on the regularization parameter λ. This has been studied by Belloni et al. in [7]. First, consider the subgradient optimality conditions which implies $|h_i| - \lambda/\sqrt{m} \leq 0$ [13], where $\mathbf{h}$ is defined as

$$\begin{aligned}\mathbf{h} &= \bigtriangledown \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{\Phi}}\mathbf{W}^{-1}\tilde{\mathbf{p}} \right\|_2 \Big|_{\tilde{\mathbf{p}}^{\text{true}}} \\ &= -\frac{\text{Re}\left\{ \mathbf{W}^{-1}\tilde{\mathbf{\Phi}}^H \mathbf{U}_\Lambda^H \mathbf{\Gamma}^{1/2} \tilde{\mathbf{r}}_n \right\}}{\left\| \mathbf{U}_\Lambda^H \mathbf{\Gamma}^{1/2} \tilde{\mathbf{r}}_n \right\|_2}. \end{aligned} \tag{10}$$

Note that the gradient $\mathbf{h}$ in (10) is given by the insertion of the true $\mathbf{p}$. From the previous definitions it can also be shown that $\mathbf{U}_\Lambda^H \mathbf{\Gamma}^{1/2} \tilde{\mathbf{r}}_n$ approximately is a zero mean complex Gaussian vector with covariance matrix $\frac{1}{M}\mathbf{I}$. Satisfaction of the optimality condition is granted if λ is chosen larger than $\sqrt{m}\max_i |h_i|$ with high probability. Here is the idea, we generate a data set $\mathcal{D}$ whose elements d_j are obtained by simulating $\sqrt{m}\max_i |h_i|$. Then, we choose λ such that it is larger than most d_j's with a large probability $1-\alpha$. In this sense, the optimality conditions would be statistically satisfied with a large probability. This is mathematically described as

$$\lambda = c \cdot \text{ quantile}(\mathcal{D}, 1-\alpha), \tag{11}$$

$$\mathcal{D} = \left\{ d_j | d_j = \sqrt{m} \frac{\left\| \text{Re}(\mathbf{W}^{-1}\tilde{\mathbf{\Phi}}^H \boldsymbol{\xi}_j) \right\|_\infty}{\left\| \boldsymbol{\xi}_j \right\|_2}, j = 1, 2, \dots \right\}. \tag{12}$$

The quantile function returns the data value where the cumulative distribution reaches the probability level of $1-\alpha$ over the data set $\mathcal{D}$, i.e.,

$$\text{quantile}(\mathcal{D}, 1-\alpha) = t \,|\, \Pr\{\mathcal{D} < t\} \leq 1-\alpha.$$

In (12), $\boldsymbol{\xi}_j$'s are independent and identically distributed complex zero mean Gaussian normalized random vectors, and $c > 1$ is a constant. It has been recommended to set $c = 1.1$ and $1-\alpha = .95$ [7]. Equations (11) and (12) are constructed to select the smallest hyper-parameter that satisfies the optimality condition for efficiency [7].

3. EMPIRICAL RESULTS

In this section the proposed method (9) is applied to a DOA estimation problem. The numerical results are compared with standard and newly proposed methods in two different setups. In order to measure its performance, the parametric CRB is also included to provide the fundamental limit [14].

Both setups consider a uniform linear array consisting of 10 sensors such that the distance between two adjacent sensors is half a wavelength of impinging wavefronts. The reason behind the choice of a ULA is the richness and diversity of the existing DOA estimation methods for such arrays. The vector of received signals at the array is given by

$$\mathbf{y}(t) = \mathbf{\Phi}(\tilde{\boldsymbol{\theta}})\mathbf{x}(t) + \mathbf{n}(t),$$

where $\mathbf{\Phi}$ is the $m \times N$ direction matrix and its columns are defined as

$$\boldsymbol{\phi}(\tilde{\theta}_l) = \left[1 \quad e^{-\mathrm{j}\pi\cos\tilde{\theta}_l} \quad \dots \quad e^{-\mathrm{j}\pi(m-1)\cos\tilde{\theta}_l} \right]^T.$$

Here, $\tilde{\boldsymbol{\theta}}$ is the complete set of possible directions obtained by gridding the interval of $[0°, 180°]$ with the resolution of $0.1°$. Two source signals with equal power are considered, located at $\{88.05°, 92.05°\}$, and the number of measurements M is equal to 200. The source signal and the additive noise are both zero mean random complex signals with Gaussian distribution. The source signal instances are independently generated with covariance matrix $\sigma_s^2\mathbf{I}$. The noise covariance matrix alters based on scenario. The signal power varies with SNR as $\sigma_s^2 = 10^{SNR/10}$.

Scenario 1: white noise

In this case the additive noise is white with covariance matrix $\mathbf{I}$. Fig. 1 illustrates the comparison between different DOA estimation techniques and the proposed method (9). For the sake of lucidity (prevention of indistinguishability in presence of many curves), the comparison criterion is the average root mean square of the estimated DOA's over 1000 Monte Carlo trials, i.e.

$$\text{RMSE} = \sqrt{\frac{1}{2000}\sum_{k=1}^{2}\sum_{l=1}^{1000}\left(\hat{\theta}_k^l - \theta_k^{true}\right)^2}.$$

Here, $\hat{\theta}_k^l$ is the DOA estimate of the k-th source for the l-th Monte Carlo trial. The competitive methods are ROOT-MUSIC, Spectral (ordinary)-MUSIC (see, e.g., [1]) and SPICE+AP [8]. SPICE+AP is an extension of SPICE+ for array processing applications and it minimizes a slightly different objective function than SPICE+, namely,

$$\left\| \mathbf{R}^{-1/2}(\hat{\mathbf{R}} - \mathbf{R})\hat{\mathbf{R}}^{-1/2} \right\|_F^2$$

The SPICE minimization can be implemented by different methods. The one which has been promoted most is the iterative one [8]. Hence, in this simulation the iterative implementation has been chosen, with the number of iterations fixed to 200. The CVX toolbox [15] was used to solve (9). Spectral-MUSIC, SPICE+AP and wsr-LASSO estimate a (pseudo-) spectrum in the grid points. For all these methods we have selected the locations of the two highest peaks as estimates of the DOAs. Note also that both MUSIC methods need to know the number of signals in advance to form the noise subspace. The sparse estimation methods wsr-LASSO and SPICE do

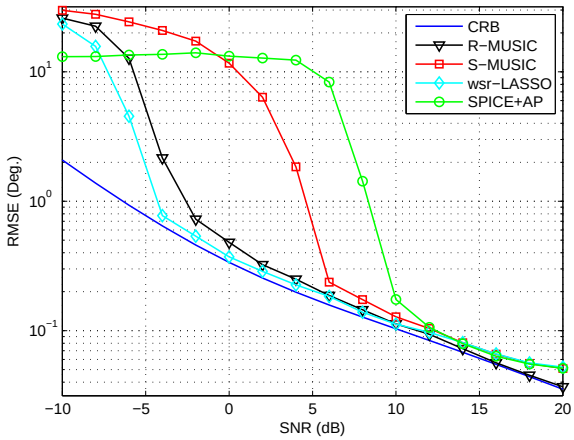

Fig. 1. The performance of DOA estimation for white additive noise described in the first scenario.

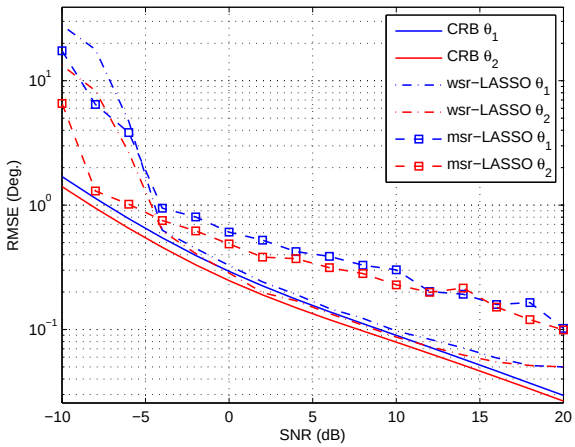

Fig. 2. The performance of DOA estimation for colored additive noise described in the second scenario.

not need or use this information in the estimation of the spectrum.

As can be seen in Fig. 1, wsr-LASSO outperforms all other methods and almost achieves the CRB bound already at an SNR equal to -4 dB. Eventually, its performance gets restricted by the grid resolution. Note that the true sources on purpose were selected to be off-the grid of the dictionary. Fig. 1 shows that SPICE+AP performs poorly when the source signals are closely separated. The only competitive method is ROOT-MUSIC, and it asymptotically obtains the CRB bound as SNR grows. The performance of ROOT-MUSIC is not restricted by the grid resolution as the other methods are since it is a parametric method. This justifies why ROOT-MUSIC does not get saturated at high SNRs. Note also that this scenario is precisely one in which ROOT-MUSIC performs at its best.

Scenario 2: colored noise

In this scenario the additive noise is colored. More precisely, the noise covariance matrix has a tri-diagonal structure. The elements of its main diagonal are all equal to one, the elements of the first diagonal above and below it are set to $-0.5\mathrm{j}$ and $0.5\mathrm{j}$, respectively. Although the true noise covariance parameters are equal along the diagonals, the estimators consider them to potentially be unequal. For this colored noise case, ROOT-MUSIC, SPECTRAL-MUSIC and SPICE are not directly applicable (they will give biased estimates). Instead, a numerical comparison is made with the newly proposed method [16], we name it as mapped sr-LASSO (msr-LASSO), is illustrated in Fig. 2. msr-LASSO minimizes the objective function

$$\|\mathbf{p}\|_1 + \lambda' \left\| \mathcal{P}(\hat{\mathbf{r}} - (\tilde{\mathbf{\Phi}}^* \circ \tilde{\mathbf{\Phi}})\mathbf{p}) \right\|_2,$$

with a positivity constraint on $\mathbf{p}$. Here, $\mathcal{P}$ is a projection that maps the support of the noise covariance matrix to zero. As can be seen in Fig. 2, wsr-LASSO performs poorly in very low SNRs, but as SNR grows it outperforms msr-LASSO. It almost achieves the CRB when SNR is adequate, and as SNR grows immensely it again saturates at the half of the grid resolution ($.05^\circ$ around 20 dB).

4. CONCLUSION

The wsr-LASSO method is introduced as an extension to sr-LASSO. It inherits all the strong characteristics of sr-LASSO and in addition it offers more. It performs near the CRB bound for low SNRs and it also considers correlated perturbations with merely the assumption on the knowledge of the structure of the noise covariance matrix.

Empirical results for DOA estimation are provided. They illustrate the superiority of wsr-LASSO to other existing, standard and newly proposed, methods.

Further theoretical investigations are needed. They are mainly the maximum number of detectable signal sources, the relation with MDL and asymptotic bounds for the error of estimation.

5. ACKNOWLEDGMENT

The authors would like to thank Mohammadreza Malek-Mohammadi for helpful discussions.

6. REFERENCES

[1] 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, Jul 1996.

[2] G.B. Giannakis and S.D. Halford, "Asymptotically optimal blind fractionally spaced channel estimation and performance analysis," *Signal Processing, IEEE Transactions on*, vol. 45, no. 7, pp. 1815–1830, Jul 1997.

[3] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 2, pp. 387–392, Apr 1985.

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

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

[6] Emmanuel Candes and Terence Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, pp. 2313–2351, 2007.

[7] A. Belloni, V. Chernozhukov, and L. Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[8] Petre Stoica, P. Babu, and Jian Li, "Spice: A sparse covariance-based estimation method for array processing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 629–638, Feb 2011.

[9] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185 – 210, 1998.

[10] C.R. Rojas, D. Katselis, and H. Hjalmarsson, "A note on the spice method," *Signal Processing, IEEE Transactions on*, vol. 61, no. 18, pp. 4545–4551, Sept 2013.

[11] Prabhu Babu and Petre Stoica, "Connection between spice and square-root lasso for sparse parameter estimation," *Signal Processing*, vol. 95, pp. 10–14, 2014.

[12] Theodore W Anderson, *An introduction to multivariate statistical analysis*, Wiley, 1984.

[13] Peter Bühlmann and Sara Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*, Springer, 2011.

[14] M. Jansson, B. Göransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *Signal Processing, IEEE Transactions on*, vol. 47, no. 4, pp. 945–956, Apr 1999.

[15] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx`, Mar. 2014.

[16] M. Malek-Mohammadi, M. Jansson, A. Owrang, A. Koochakzadeh, and M. Babaie-Zadeh, "Doa estimation in partially correlated noise using low-rank/sparse matrix decomposition," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*, June 2014, pp. 373–376.