Cramer-Rao Bound for a Sparse Complex Model

Anisia Florescu, Emilie Chouzenoux, Jean-Christophe Pesquet, Silviu Ciochina

Cramer-Rao Bound for a Sparse Complex Model

Anisia Florescu[1,*], Emilie Chouzenoux[2], Jean-Christophe Pesquet[2], and Silviu Ciochină[3]
[1]*Dunărea de Jos University, Electronics and Telecommunications Dept., Galaţi, România*
[2]*Université Paris-Est, LIGM, UMR CNRS 8049, Marne-la-Vallée, France*
[3]*Politehnica University of Bucharest, Telecommunications Dept., România*
[*]Corresponding author (E-mail : anisia.florescu@ugal.ro)

***Abstract*—Complex-valued data play a prominent role in a number of signal and image processing applications. The aim of this paper is to establish some theoretical results concerning the Cramer-Rao bound for estimating a sparse complex-valued vector. Instead of considering a countable dictionary of vectors, we address the more challenging case of an uncountable set of vectors parameterized by a real variable. We also present a proximal forward-backward algorithm to minimize an ℓ_0 penalized cost, which allows us to approach the derived bounds. These results are illustrated on a spectrum analysis problem in the case of irregularly sampled observations.**

***Keywords* - sparsity; estimation; Cramer-Rao bound; complex signals; proximal methods; nonconvex optimization; spectrum estimation**

I. Introduction

Many efforts have been dedicated over the past few years to the development of methods for estimating sparse signals based on a linear observation model. The problem of finding a sparse solution to an undetermined system of linear equations is *NP*-complete [1] so that it requires an exhaustive search. In practice, the problem can be relaxed because it is sufficient to obtain an approximately optimal solution. Theoretically, finding such a sparse solution requires establishing conditions under which it is possible to determine consistent and efficient estimators as well as evaluating their quality. As a general characteristic, to evaluate the performance of an estimator, one has to compare the mean square estimation error with theoretical bounds. From a statistical viewpoint, this also means that it is useful to determine an estimator whose dispersion is close to these bounds [2]. Although several lower limits for the dispersion exist in the literature, the Cramer-Rao bound (CRB) is often preferred because it is easier to calculate and, under some conditions, it can be asymptotically attained by using the maximum likelihood method [3]. This motivated the study of the CRB under constraints [4], in particular, for the estimation of real-valued sparse signals [5].

A large number of application areas of signal and image processing (e.g. digital communications, spectroscopy, Magnetic Resonance Imaging) involve complex-valued data. However, in this context, statistical estimator bounds have not been as extensively studied as for real-valued data, especially for sparse signals. The main contribution of this work is to calculate the CRB for sparse complex-valued signals. In addition, while most of existing works in the sparse and compressive sensing literature [6] focus on countable dictionaries, we will be interested in the case when an uncountable dictionary of signals depending on a real-valued parameter is employed. Another contribution of this work is to propose a proximal method which yields a performance close to the CRB by minimizing a nonconvex cost function including an ℓ_0 penalty term.

The rest of the paper is organized as follows: in Section II, we introduce the complex-valued statistical model which is investigated in this work. We then provide a general expression of the Fisher Information Matrix (FIM) for the related parameter estimation problem. Section III describes a forward-backward optimization algorithm to estimate sparse complex-valued signals in an uncountable dictionary. Simulation results are then given in Section IV to validate our results from irregularly sampled observations for the estimation of cisoids in the presence of Gaussian circular noise. The performance of our algorithm is shown to be close to the CRB. Some conclusions are drawn in Section V.

II. Theoretical results

A. Statistical model

A classical problem in statistical signal processing consists of recovering a signal x from a vector of Q observations

$$y = x + w \tag{1}$$

where w is a noise vector. Here, we assume that $w \in \mathbb{C}^Q$ is a realization of a circular Gaussian random noise vector W with zero-mean and covariance matrix $\Gamma = \mathsf{E}[WW^H] \in \mathbb{C}^{Q\times Q}$ ($(\cdot)^H$ denotes the transconjugate operation). We assume that the signal $x \in \mathbb{C}^Q$ admits a sparse representation in a finite dictionary $\mathcal{E} = \{e_\upsilon \mid \upsilon \in \mathbb{R}\}$ of vectors of $\mathbb{C}^Q$ which are parameterized by a scalar variable $\upsilon \in \mathbb{R}$. More precisely, there exist $M \in \mathbb{N}^*$, $c = [c_1, \ldots, c_M]^\top \in (\mathbb{C}^*)^M$ and $\nu = [\nu_1, \ldots, \nu_M]^\top \in \mathbb{R}^M$ such that

$$x = \sum_{n=1}^{M} c_n e_{\nu_n} = [e_{\nu_1} \ldots e_{\nu_M}]c = Ec. \tag{2}$$

The vector $y \in \mathbb{C}^Q$ is thus a realization of a random vector Y with probability density function

$$\mathsf{p}_{Y|c,\nu}(y) = \frac{1}{(\pi)^M \big(\det(\Gamma)\big)^{1/2}} \exp\big(-(y-Ec)^H \Gamma^{-1}(y-Ec)\big). \tag{3}$$

In the following, it is assumed that $\upsilon \mapsto e_\upsilon$ is a twice differentiable function.

B. Calculation of the CRB

Up to an additive constant, the negative-log-likelihood is equal to

$$\mathcal{L}(y \mid c, \nu) = w^H \Gamma^{-1} w. \tag{4}$$

In the following w_R and w_I denote the real part and the imaginary part of w, a similar notation being used for other complex-valued vectors and matrices.

Let us first look at the expression of the Wirtinger's derivative [7] of the neg-log-likehood with respect to the conjugate of c:

$$\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial c^*} = \frac{1}{2}\Big(\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial c_R} + \imath \frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial c_I}\Big) = -E^H \Gamma^{-1} w. \tag{5}$$

We have then

$$\begin{aligned}\frac{\partial^2 \mathcal{L}(y \mid c, \nu)}{\partial c_R \partial c_R^\top} &= -2 \frac{\partial \big(w_R^\top (\Gamma^{-1} E)_R + w_I^\top (\Gamma^{-1} E)_I\big)}{\partial c_R} \\ &= 2 \operatorname{Re}\{E^H \Gamma^{-1} E\}\end{aligned} \tag{6}$$

and, by similar calculations,

$$\frac{\partial^2 \mathcal{L}(y \mid c, \nu)}{\partial c_R \partial c_I^\top} = -2 \operatorname{Im}\{E^H \Gamma^{-1} E\} \tag{7}$$

$$\frac{\partial^2 \mathcal{L}(y \mid c, \nu)}{\partial c_I \partial c_I^\top} = 2 \operatorname{Re}\{E^H \Gamma^{-1} E\}. \tag{8}$$

On the other hand, the neg-log-likelihood can be reexpressed as

$$\mathcal{L}(y \mid c, \nu) = \Big(y - \sum_{n=1}^M c_n e_{\nu_n}\Big)^H \Gamma^{-1} \Big(y - \sum_{n=1}^M c_n e_{\nu_n}\Big). \tag{9}$$

For every $n \in \{1, \dots, M\}$, this leads to

$$\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial \nu_n} = -2 \operatorname{Re}\{c_n^* (e'_{\nu_n})^H \Gamma^{-1} w\} \tag{10}$$

where e'_{ν_n} is the gradient of $\upsilon \mapsto e_\upsilon$ at ν_n. For the second-order derivatives, we deduce that, for every $(n, m) \in \{1, \dots, M\}^2$,

$$\begin{aligned}\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial \nu_n \partial \nu_m} &= 2\Big(\operatorname{Re}\{c_n^* c_m (e'_{\nu_n})^H \Gamma^{-1} e'_{\nu_m}\} \\ &\quad - \operatorname{Re}\{c_n^* (e''_{\nu_n})^H \Gamma^{-1} w\} \delta_{n-m}\Big)\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial \nu_n \partial c_m^*} &= \frac{1}{2}\Big(\frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial \nu_n \partial c_{R,m}} + \imath \frac{\partial \mathcal{L}(y \mid c, \nu)}{\partial \nu_n \partial c_{I,m}}\Big) \\ &= c_n e_{\nu_m}^H \Gamma^{-1} e'_{\nu_n} - (e'_{\nu_n})^H \Gamma^{-1} w\, \delta_{n-m}\end{aligned} \tag{12}$$

where e''_{ν_n} is the second-order derivative of $\upsilon \mapsto e_\upsilon$ at ν_n. Let $p = [c_R^\top \; c_I^\top \; \nu^\top]^\top \in \mathbb{R}^{3M}$ be the vector of parameters to be estimated. The Fisher Information Matrix is defined as

$$\mathcal{F}_p = \mathsf{E}\Big[\frac{\partial^2 \mathcal{L}(Y \mid c, \nu)}{\partial p \partial p^\top}\Big] \in \mathbb{R}^{3M \times 3M}. \tag{13}$$

Since W is zero-mean, (11) and (12) yield, for every $(n, m) \in \{1, \dots, M\}^2$,

$$\mathsf{E}\Big[\frac{\partial \mathcal{L}(Y \mid c, \nu)}{\partial \nu_n \partial \nu_m}\Big] = 2 \operatorname{Re}\{f_n^H \Gamma^{-1} f_m\} \tag{14}$$

$$\mathsf{E}\Big[\frac{\partial \mathcal{L}(Y \mid c, \nu)}{\partial \nu_n \partial c_m^*}\Big] = e_{\nu_m}^H \Gamma^{-1} f_n \tag{15}$$

where $f_n = c_n e'_{\nu_n}$. Hence, we end up with the following property:

Proposition 1. *The FIM is equal to*

$$\mathcal{F}_p = 2 \begin{bmatrix} \operatorname{Re}\{E^H \Gamma^{-1} E\} & -\operatorname{Im}\{E^H \Gamma^{-1} E\} & \operatorname{Re}\{E^H \Gamma^{-1} F\} \\ \operatorname{Im}\{E^H \Gamma^{-1} E\} & \operatorname{Re}\{E^H \Gamma^{-1} E\} & \operatorname{Im}\{E^H \Gamma^{-1} F\} \\ \operatorname{Re}\{F^H \Gamma^{-1} E\} & -\operatorname{Im}\{F^H \Gamma^{-1} E\} & \operatorname{Re}\{F^H \Gamma^{-1} F\} \end{bmatrix} \tag{16}$$

where $F = [c_1 e'_{\nu_1} \dots c_M e'_{\nu_M}] \in \mathbb{C}^{Q \times M}$.

Remark 1.

(i) When a sum of distinct cisoids is considered, i.e., for every $\upsilon \in \mathbb{R}$,

$$e_\upsilon = \big(\exp(i\upsilon\tau_q)\big)_{1 \le q \le Q} \tag{17}$$

where $\tau_1, \dots, \tau_Q$ are distinct sampling times, we have

$$F = \imath T E C \tag{18}$$

with $T = \operatorname{Diag}(\tau_1, \dots, \tau_Q)$ and $C = \operatorname{Diag}(c_1, \dots, c_M)$. The resulting expression of $\mathcal{F}_p$ was already obtained in [8] for the irregular sampling case when $M = 1$, but the authors consider a more general polynomial phase complex-valued signal and they adopt a different parameterization by considering the estimation of the modulus and phase of c_1 instead of its real and imaginary parts.

(ii) When a sum of distinct cisoids is considered as previously, a regular sampling is performed by choosing $(\forall q \in \{1, \dots, Q\})\ \tau_q = q - 1$, and the noise components are uncorrelated with variance σ^2, (16) can be simplified by performing approximations of the involved matrices. As a consequence, for any unbiased estimates $\widehat{c}$ and $\widehat{\nu}$, we have, for every $n \in \{1, \dots, M\}$,

$$\operatorname{Var}[\widehat{c}_n] \ge \frac{\sigma^2}{2Q} \tag{19}$$

$$\operatorname{Var}[\widehat{\nu}_n] \ge \frac{6\sigma^2}{Q(Q^2 - 1)|c_n|^2}. \tag{20}$$

This means that the performance is then similar to that resulting from the estimations of M single cisoids at angular frequencies $(\nu_n)_{1 \le n \le M}$. Such an approximation is valid when $Q \gg 1$ [9].

(iii) In the above derivations, the parameter M must be set to the exact model order, otherwise $\mathcal{F}_p$ is singular.

III. Proposed estimation algorithm

We now propose an efficient algorithm to estimate the parameters $(\nu_n)_{1 \le n \le M}$ and $(c_n)_{1 \le n \le M}$ in Model (2). Imposing that the values of the complex amplitudes $(c_n)_{1 \le n \le M}$ are all nonzero is however equivalent to assuming that the support of the coefficients is known. In a more realistic manner, we consider here a larger set of parameters $(\nu_n, c_n)_{1 \le n \le N}$, but we assume that the so-defined representation is sparse in the sense that only a small subset of coefficients $(c_n)_{1 \le n \le N}$ of unknown size $M \le N$ is nonzero.

A main difficulty in this context is to appropriately choose the parameters $(\nu_n)_{1 \le n \le N}$. To do so, we consider that the

parameters $(\nu_n)_{1\leq n\leq N}$ are known in an imprecise manner, i.e. they are such that, for every $n \in \{1,\ldots,N\}$,

$$\nu_n = \theta_n + \delta_n \tag{21}$$

where $\theta_n \in \mathbb{R}$ is some given value and $\delta_n \in \mathbb{R}$ is an unknown error on the parameter ν_n to be estimated. Typically, $(\theta_n)_{1\leq n\leq N}$ may correspond to values on a search grid and $(\delta_n)_{1\leq n\leq N}$ are possible shifts with respect to these values. Such an approach was followed in [10] but it was however restricted to real-valued signals, which makes it much simpler.

Under the considered differentiability assumptions for $\upsilon \mapsto e_\upsilon$, if we suppose that the perturbations $(\delta_n)_{1\leq n\leq N}$ are small, we can perform the following first-order Taylor expansion:

$$(\forall n \in \{1,\ldots,N\}) \qquad e_{\nu_n} \simeq e_{\theta_n} + \delta_n e'_{\theta_n}. \tag{22}$$

For this approximation to be valid and also for avoiding some possible ambiguities in the estimation process, we contrain the perturbation parameter vector $\delta = (\delta_n)_{1\leq n\leq N}$ to belong to an N-dimensional box $B = [-\Delta_1,\Delta_1]\times\cdots\times[-\Delta_N,\Delta_N]$, where $(\Delta_n)_{1\leq n\leq N} \in [0,+\infty)^N$. Using Approximation (22), the model under consideration takes the bilinear form:

$$x = \sum_{n=1}^{N} \big(c_n e_{\theta_n} + c_n\delta_n e'_{\theta_n}\big). \tag{23}$$

In order to estimate the parameters $(\nu_n)_{1\leq n\leq N}$ and $(c_n)_{1\leq n\leq N}$, while taking into account the underlying sparsity assumptions, we propose to solve the following optimization problem:

$$\underset{\substack{c=(c_n)_{1\leq n\leq N}\in\mathbb{C}^N\\ \delta=(\delta_n)_{1\leq n\leq N}\in B}}{\text{minimize}} \Big\| \sum_{n=1}^{N} \big(c_n e_{\theta_n} + c_n\delta_n e'_{\theta_n}\big) - y\Big\|^2_{\Gamma^{-1}} + \lambda\ell_0(c) \tag{24}$$

where $\ell_0(c)$ denotes the number of nonzero coefficients in c, $\|\cdot\|^2_{\Gamma^{-1}}$ is the weighted squared norm equal to $(\cdot)^H\Gamma^{-1}(\cdot)$, and $\lambda \in (0,+\infty)$ is a regularization constant used to impose the sparsity of the representation. (One procedure was devised in order to automatically set the regularization parameter.) Note that the first data-fidelity term in the expression of the cost function corresponds to the neg-log-likelihood of the noise corrupting the observations.

Let us introduce the variables $(d_n)_{1\leq n\leq N} = (c_n\delta_n)_{1\leq n\leq N}$, let $\tilde{E} = [e_{\nu_1},\ldots,e_{\nu_N}]$ and $E' = [e'_{\nu_1},\ldots,e'_{\nu_N}]$, and let the function $(\psi_n)_{1\leq n\leq N}$ be defined as

$$(\forall (c_n,d_n)\in\mathbb{C}^2) \qquad \psi_n(c_n,d_n) = \lambda\,\ell_0(c_n) + \iota_{S_n}(c_n,d_n), \tag{25}$$

where ι_{S_n} is the indicator function of the closed nonconvex cone:

$$S_n = \{(c_n,d_n)\in\mathbb{C}^2 \;\mid\; \exists\delta_n\in[-\Delta_n,\Delta_n], d_n = \delta_n c_n\}. \tag{26}$$

The form of the objective function in (24) suggests the use of a forward-backward algorithm [11]:

$$\begin{array}{l}
(c^{(0)},d^{(0)}) \in (\mathbb{C}^N)^2\\
0<\underline{\gamma}\leq\overline{\gamma}<\|\Gamma\|\|\tilde{E}\tilde{E}^H + E'(E')^H\|^{-1}\\
\text{For } k=0,1,\ldots\\
\left\lfloor\begin{array}{l}
D^{(k)} = \Gamma^{-1}\Big([\tilde{E}\;\; E']\begin{bmatrix}c^{(k)}\\ d^{(k)}\end{bmatrix} - y\Big)\\
(\widetilde{c}_n^{(k)})_{1\leq n\leq N} = c^{(k)} - \gamma^{(k)}\tilde{E}^H D^{(k)},\;\; \gamma^{(k)}\in(\underline{\gamma},\overline{\gamma})\\
(\widetilde{d}_n^{(k)})_{1\leq n\leq N} = d^{(k)} - \gamma^{(k)}(E')^H D^{(k)}\\
(c_n^{(k+1)},d_n^{(k+1)})_{1\leq n\leq N} = \big(\operatorname{prox}_{\gamma^{(k)}\psi_n}(\widetilde{c}_n^{(k)},\widetilde{d}_n^{(k)})\big)_{1\leq n\leq N}.
\end{array}\right.
\end{array} \tag{27}$$

In order to implement the above algorithm, the expressions of the involved proximity operators $(\operatorname{prox}_{\gamma^{(k)}\psi_n})_{1\leq n\leq N}$ are needed at each iteration $k\in\mathbb{N}$ (see [12] for more technical details). Note that this algorithm can be viewed as an extension of an iterative hard thresholding algorithm [13]. Its convergence to a critical point of (24) can be established.

IV. Numerical illustration

To illustrate the validity of our approach, we consider $Q = 50$ observations of a complex-valued signal corresponding to the sum of $M = 6$ cisoids which have been irregularly sampled in a random manner over $[0,Q]$ (see (17)), the associated phases being uniformly distributed over $[0,2\pi]$. The discrete-time observations are corrupted with a white circular Gaussian noise with zero-mean and variance σ^2 ($\Gamma = \sigma^2 I_Q$). The employed dictionary consists of $N = 500$ cisoids, the angular frequencies of which are $(\theta_n)_{1\leq n\leq N} = \big(2\pi(n-1)/N\big)_{1\leq n\leq N}$. The frequencies of the sparse components do not belong to the search grid. It must be emphasized that, for many spectrum estimation methods, the choice of an appropriate search frequency grid constitutes the main difficulty [14] . We have tested various values of the signal-to-noise ratio (expressed in dB), which is defined as

$$\mathrm{SNR} = 20\log_{10}\Big(\frac{c_{\min}}{\sigma}\Big) \tag{28}$$

where $c_{\min}$ is the minimum modulus value of the nonzero complex amplitudes of the original cisoids.

In a first experiment, 500 noise realizations have been generated by changing the sampling times and phase values randomly for each run. Fig. 1 allows us to assess the good performance of the proposed approach. In a second experiment, 400 noise realizations have been generated for each SNR value, while keeping the same sampling time and phase values, so as to compare the variance of the estimation errors with CRBs. The CRBs have been computed from the diagonal terms of the inverse FIM as expressed by Proposition 1. In Fig. 2, we show the results obtained for the worst estimated cisoid. There is however a good fit with the derived lower bounds over a wide range of SNR values. The fact that the CRB is actually valid for perfectly unbiased estimators may account for some of the observed differences, especially at high SNR.

V. Conclusion

In this paper, we have provided the expression of the Fisher Information Matrix for the estimation of a sparse complex-valued vector. An important feature of our approach is that the

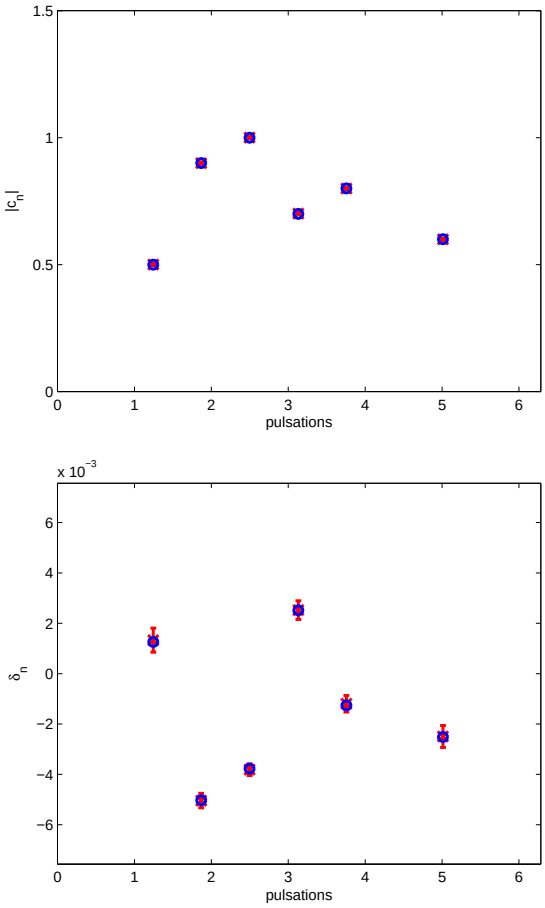

Figure 1. Estimation results in the case of 6 cisoids (SNR = 23.27 dB): values of $(|c_n|)_{1\leq n\leq N}$ (top); frequency perturbations $(\delta_n)_{1\leq n\leq N}$ as a function of $(\theta_n)_{1\leq n\leq N}$ (bottom). The exact values are depicted with blue circles and the confidence intervals on the estimates in red (the mean is indicated by a cross).

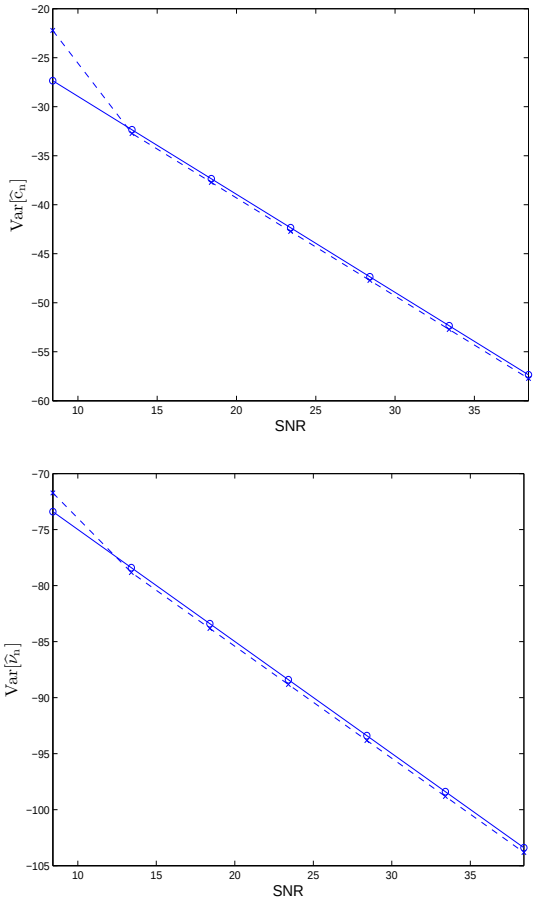

Figure 2. Variance of the estimation error on the complex amplitude (top) and the frequency (bottom) for one of the cisoids ($M = 6$) vs SNR. The estimates obtained with Algorithm (27) are indicated with x symbols and dashed lines and CRBs are indicated with o symbols and continuous lines.

dictionary elements are parametrized by a scalar real variable the value of which is allowed to vary continuously. The estimation problem has been formulated as a nonconvex and nonsmooth minimization problem which can be efficiently solved with a proximal forward-backward algorithm. Simulation examples have illustrated the good practical performance of the proposed method and its ability to generate estimation errors with variances close to the predicted lower bounds. Finally, it is important to note that our approach is applicable to dictionaries including other functions than cisoids.

REFERENCES

[1] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 14, pp. 342–351, 2005.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 2010.

[3] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York, Springer-Verlag, Second Edition, 1998.

[4] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramer–Rao bound with a singular Fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, no 6, pp. 453–456, June 2009.

[5] Z. Ben-Haim and Y. C. Eldar, "The Cramer–Rao bound for sparse estimation," *IEEE Trans. Signal Process.*, vol. 58, no 6, pp. 3384–3389, June 2010.

[6] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no 2, pp. 21–30, Mar. 2008.

[7] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: the proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, pp. 5101–5125, 2011.

[8] J. A. Legg and D. A. Gray, "Performance bounds for polynomial phase parameter estimation with nonuniform and random sampling schemes," *IEEE Trans. Signal Process.*, vol. 48, no 2, pp. 331–337, Aug. 2002.

[9] J.M. Francos and B. Friedlander, "Bounds for estimation of complex exponentials in unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no 9, pp. 2176–2185, Sep. 1995.

[10] C. Ekanadham, D. Tranchina, and E. P. Simoncelli,"Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no10, pp. 4735–4744, Oct. 2011.

[11] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," In H. H. Bauschke, R. S.Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, Springer-Verlag, New York, 2011.

[12] A. Florescu, E. Chouzenoux, J.-C. Pesquet, S. Ciochină, "A constrained optimization approach for complex sparse perturbed models," technical report, http://hal.archives-ouvertes.fr/hal-00783298, Jul. 2013

[13] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no 5, pp. 629–654, 2008.

[14] P. Stoica and P. Babu,"Sparse estimation of spectral lines: Grid selection problems and their solutions," *IEEE Trans. Signal Process.*, vol. 60, no 2, pp. 962–967, Feb. 2012.