A Multi-Parameter Optimization Approach for Complex Continuous Sparse Modelling

Emilie Chouzenoux, Jean-Christophe Pesquet
Université Paris-Est
LIGM, UMR CNRS 8049
Champs sur Marne, France
first.last@u-pem.fr

Anisia Florescu
Dunărea de Jos University
Electronics and Telecommunications Dept.
Galaţi, România
anisia.florescu@ugal.ro

***Abstract*—The main focus of this work is the estimation of a complex valued signal assumed to have a sparse representation in an uncountable dictionary of signals. The dictionary elements are parameterized by a real-valued vector and the available observations are corrupted with an additive noise. By applying a linearization technique, the original model is recast as a constrained sparse perturbed model. The problem of the computation of the involved multiple parameters is addressed from a nonconvex optimization viewpoint. A cost function is defined including an arbitrary Lipschitz differentiable data fidelity term accounting for the noise statistics, and an ℓ_0-like penalty. A proximal algorithm is then employed to solve the resulting nonconvex and nonsmooth minimization problem. Experimental results illustrate the good practical performance of the proposed approach when applied to 2D spectrum analysis.**

***Index Terms*—sparse modelling; continuous compressive sensing; proximity operator; hard thresholding; multivariate estimation; 2D spectrum estimation; forward-backward algorithm; nonconvex optimization.**

I. Introduction

A classical problem in sparse estimation [1] consists of recovering a signal $\overline{x}$ from a vector of observations:

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

where $w \in \mathbb{C}^Q$ is a realization of a random noise vector. Here, we assume that the signal $\overline{x} \in \mathbb{C}^Q$ admits a finite sparse representation on a family of vectors $\mathcal{E} = \{e_\nu \mid \nu \in \mathbb{R}^M\}$ of $\mathbb{C}^Q$ which are parameterized by a vector variable $\nu \in \mathbb{R}^M$. More precisely, we suppose that a finite subset $\{e_{\overline{\nu}_n} \mid 1 \leqslant n \leqslant N\}$ of distinct elements of $\mathcal{E}$ can be extracted such that

$$\overline{x} = \sum_{n=1}^{N} \overline{c}_n e_{\overline{\nu}_n} \tag{2}$$

where a large number of components of $\overline{c} = (\overline{c}_n)_{1\leqslant n\leqslant N} \in \mathbb{C}^N$ are equal to zero (overlined variables are used here to distinguish "true" vectors from generic variables).

In this paper, we consider the challenging context of estimation problems in which the parameters $(\overline{\nu}_n)_{1\leqslant n\leqslant N}$ are known in an imprecise manner, i.e. they are such that, for every $n \in \{1,\dots,N\}$,

$$\overline{\nu}_n = \theta_n + \overline{\delta}_n \tag{3}$$

where $\theta_n \in \mathbb{R}^M$ is some given value and $\overline{\delta}_n \in \mathbb{R}^M$ is an unknown error on the parameter to be estimated. Under suitable differentiability assumptions on function $\nu \mapsto e_\nu$, if we assume that the perturbations $(\overline{\delta}_n)_{1\leqslant n\leqslant N}$ are small, we can perform the following first-order Taylor expansion:

$$(\forall n \in \{1,\dots,N\}) \qquad e_{\overline{\nu}_n} \simeq e_{\theta_n} + e'_{\theta_n}\overline{\delta}_n \tag{4}$$

where $e'_{\theta_n} \in \mathbb{C}^{Q\times M}$ is the Jacobian matrix of $\nu \mapsto e_\nu = \big(e_\nu(q)\big)_{1\leqslant q\leqslant Q}$ at θ_n. For every $q \in \{1,\dots,Q\}$, the q-th line of e'_{θ_n} thus corresponds to the gradient of the q-th component function $\nu \mapsto e_\nu(q)$ at θ_n. With this approximation, Model (2) takes the following bilinear form

$$\overline{x} = \sum_{n=1}^{N} \big(\overline{c}_n e_{\theta_n} + \overline{c}_n e'_{\theta_n}\overline{\delta}_n\big). \tag{5}$$

A similar sparse approach for decomposing a signal in terms of translated versions of some features in a finite dictionary is addressed in [2] where the proposed convex ℓ_1 formulation is taylored for real-valued signals in the case when $M = 1$. Likewise, our work can be seen as bearing some similarities with the perturbed compressive sampling approach in [3] where a robust total least squares (TLS) approach based on an ℓ_1 regularization is developed. The difference is that, in this paper, we adopt a different formulation where an ℓ_0 cost is employed for the minimization process, instead of its ℓ_1 convex relaxation, and the perturbations $(\overline{\delta}_n)_{1\leqslant n\leqslant N}$ are constrained to satisfy the following inequalities:

$$(\forall n \in \{1,\dots,N\}) \qquad \|\overline{\delta}_n\| \leqslant \Delta_n, \tag{6}$$

the upper bounds $(\Delta_n)_{1\leqslant n\leqslant N} \in [0,+\infty)^N$ being defined by the user. Hereabove $\|\cdot\|$ denotes the Euclidean norm. Constraints (6) provide more flexibility than the TLS approach for controlling the perturbations $(\overline{\delta}_n)_{1\leqslant n\leqslant N}$. In addition, our method makes it possible to deal with a general Lipschitz differentiable data fidelity term. Note that, in our previous work [4], a similar linearization approach was adopted where the parameter ν was restricted to be a scalar ($M = 1$). The proposed forward-backward method was then shown to be an extension of an iterative hard thresholding algorithm [5]. In this work, we will be interested in the more intricate case when $M > 1$, which is useful for taking advantage of flexible dictionaries.

This paper is organized as follows: In Section II, the estimation of the sparse components and of the corresponding perturbation parameter vector are formulated as a nonconvex constrained optimization problem. An iterative algorithm for solving this problem is provided in Section III, which requires to derive the expression of the proximity operator of a function of several variables. A detailed discussion on this issue is provided. The proposed approach is illustrated on a 2D spectrum estimation example in Section IV. The conclusions of this work are given in Section V.

Notation: In the following, χ_S denotes the characteristic function of a set S which is equal to 0 on S and 1 elsewhere, and ι_S denotes the indicator function of a set S, which is equal to 0 on S and $+\infty$ elsewhere. The transpose and transconjugate operation for complex-valued vectors or matrices are denoted by $(\cdot)^\top$ and $(\cdot)^H$, respectively.

II. Variational formulation

We propose to estimate the parameters of the perturbed sparse model by solving the following optimization problem:

$$\underset{\substack{c=(c_n)_{1\leqslant n\leqslant N}\in\mathbb{C}^N\\ \delta=(\delta_n)_{1\leqslant n\leqslant N}\in B}}{\operatorname{minimize}} \quad \Phi\Big(\sum_{n=1}^{N}\big(c_n e_{\theta_n} + c_n e'_{\theta_n}\delta_n\big) - y\Big) + \lambda \ell_0(c) + \frac{\epsilon}{2}\|c\|^2 \tag{7}$$

where

$$B = \{(\delta_n)_{1\leqslant n\leqslant N} \in (\mathbb{R}^M)^N \mid (\forall n \in \{1,\ldots,N\})\ \|\delta_n\| \leqslant \Delta_n\}, \tag{8}$$

$\Phi\colon \mathbb{C}^Q \to \mathbb{R}$ is the data-fidelity term which is often chosen equal to the negative-log-likelihood of the noise corrupting the observations, $\lambda \in (0,+\infty)$ is a regularization constant serving to promote sparsity, and $\epsilon \in [0,+\infty)$. The last term plays a role similar to an elastic net regularization [6].

Let us now define the matrices $E = [e_{\theta_1} \ldots e_{\theta_N}] \in \mathbb{C}^{Q\times N}$, $E' = [e'_{\theta_1} \ldots e'_{\theta_N}] \in \mathbb{C}^{Q\times MN}$, and let us introduce the variable $d = [c_1\delta_1^\top \ldots c_N\delta_N^\top]^\top \in \mathbb{C}^{MN}$. In addition, let the function Ψ be defined as

$$(\forall c = (c_n)_{1\leqslant n\leqslant N} \in \mathbb{C}^N)(\forall d = [d_1^\top \ldots d_N^\top]^\top \in \mathbb{C}^{MN}) \quad \Psi(c,d) = \sum_{n=1}^{N} \psi_n(c_n, d_n) \tag{9}$$

where $(\forall n \in \{1,\ldots,N\})(\forall c_n \in \mathbb{C})(\forall d_n \in \mathbb{C}^M)$

$$\psi_n(c_n,d_n) = \lambda\, \chi_{\{0\}}(c_n) + \iota_{S_n}(c_n,d_n) + \frac{\epsilon}{2}|c_n|^2, \tag{10}$$

and S_n is the closed cone given by

$$S_n = \{(c_n,d_n) \in \mathbb{C}\times\mathbb{C}^M \mid \exists \delta_n \in \mathbb{R}^M \text{ with } \|\delta_n\| \leqslant \Delta_n, d_n = c_n\delta_n\}. \tag{11}$$

Then, Problem (7) is equivalent to minimizing function

$$(c,d) \mapsto \Phi\Big([E\ \ E']\begin{bmatrix} c\\ d\end{bmatrix} - y\Big) + \Psi(c,d). \tag{12}$$

We shall now see how this minimization can be performed numerically.

III. Proposed algorithm

A. Algorithm form

If we assume that Φ is a differentiable function, the previous split form of the objective function suggests the use of a forward-backward algorithm [7]:

$$\begin{array}{l}
c^{(0)} \in \mathbb{C}^N, d^{(0)} \in \mathbb{C}^{MN}\\
0 < \underline{\gamma} \leqslant \overline{\gamma} < \|EE^H + E'(E')^H\|^{-1}\\
\text{For } k = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\gamma^{(k)} \in [\underline{\gamma}, \overline{\gamma}]\\
D^{(k)} = \nabla\Phi\Big([E\ \ E']\begin{bmatrix} c^{(k)}\\ d^{(k)}\end{bmatrix} - y\Big)\\
(\widetilde{c}_n^{(k)})_{1\leqslant n\leqslant N} = c^{(k)} - \gamma^{(k)}E^H D^{(k)}\\
(\widetilde{d}_n^{(k)})_{1\leqslant n\leqslant N} = d^{(k)} - \gamma^{(k)}(E')^H D^{(k)}\\
(c_n^{(k+1)}, d_n^{(k+1)})_{1\leqslant n\leqslant N} = \big(\operatorname{prox}_{\gamma^{(k)}\psi_n}(\widetilde{c}_n^{(k)}, \widetilde{d}_n^{(k)})\big)_{1\leqslant n\leqslant N}.
\end{array}
\right.
\end{array} \tag{13}$$

We recall that the proximity operator of a proper, lower bounded, lower semi-continuous function $\varphi\colon \mathcal{H} \to (-\infty,+\infty]$ where $\mathcal{H}$ is a finite dimensional Hilbert space equipped with the norm $\|\cdot\|$ is defined as

$$(\forall u \in \mathcal{H}) \qquad \operatorname{prox}_\varphi(u) \in \underset{v\in\mathcal{H}}{\operatorname{Argmin}}\ \frac{1}{2}\|u-v\|^2 + \varphi(v). \tag{14}$$

Although the uniqueness of $\operatorname{prox}_\varphi(u)$ is guaranteed when φ is convex, this property is not necessarily satisfied in the nonconvex case.

Note that the convergence of the forward-backward is proven under some technical conditions, even in the nonconvex case [8], but a main difficulty here is to find a tractable manner for computing $\operatorname{prox}_{\gamma\psi_n}$ when $\gamma \in (0,+\infty)$ and $n \in \mathbb{N}$.

B. Derivation of the involved proximity operator

As shown by the next result whose proof is skipped due to the lack of space, the proximity operator can be computed in a relatively simple manner:

Proposition 1. *Let $\gamma \in (0,+\infty)$. For every $n \in \{1,\ldots,N\}$ and $(c_n,d_n) \in \mathbb{C}\times\mathbb{C}^M$, the proximity operator of $\gamma\psi_n$ is*

$$\operatorname{prox}_{\gamma\psi_n}(c_n,d_n) = \begin{cases} (0,0) & \text{if } \rho_n(\widetilde{\delta}_n) < 2\gamma\lambda\\ \dfrac{c_n + \widetilde{\delta}_n^\top d_n}{1+\gamma\epsilon+\|\widetilde{\delta}_n\|^2}(1,\widetilde{\delta}_n) & \text{otherwise,}\end{cases} \tag{15}$$

where ρ_n is the function defined as

$$(\forall \delta_n \in \mathbb{R}^M) \quad \rho_n(\delta_n) = \frac{|c_n + \delta_n^\top d_n|^2}{1+\gamma\epsilon+\|\delta_n\|^2} \tag{16}$$

and

$$\widetilde{\delta}_n \in \underset{\delta_n\in\mathbb{R}^M, \|\delta_n\|\leqslant\Delta_n}{\operatorname{Argmax}}\ \rho_n(\delta_n). \tag{17}$$

The determination of $\widetilde{\delta}_n$ in (17) consists of maximizing a ratio of quadratic functions over an Euclidean ball. This kind of problem has been investigated in the optimization literature [9]. We will see however that, due to the specific form of the problem under consideration, the optimization problem can be recast in a low-dimensional space and a simple characterization of $(\widetilde{\delta}_n)_{1\leqslant n\leqslant N}$ can be obtained.

Proposition 2. *Assume that $M \geqslant 2$. Let $\gamma \in (0,+\infty)$, let $n \in \{1,\ldots,N\}$, and let $(c_n,d_n) \in \mathbb{C}\times\mathbb{C}^M$. Let $U_n \in \mathbb{R}^{M\times 2}$ be a matrix whose columns form an orthonormal basis of a vector subspace including* $\operatorname{span}\{\operatorname{Re}\{d_n\},\operatorname{Im}\{d_n\}\}$, *let $\widetilde{d}_n = U_n^\top d_n \in \mathbb{C}^2$, and let*

$$A_n = \begin{bmatrix} \operatorname{Re}\{\widetilde{d}_n\widetilde{d}_n^H\} & \dfrac{\operatorname{Re}\{c_n^*\widetilde{d}_n\}}{(1+\gamma\epsilon)^{1/2}} \\ \dfrac{\operatorname{Re}\{c_n^*\widetilde{d}_n\}^\top}{(1+\gamma\epsilon)^{1/2}} & \dfrac{|c_n|^2}{1+\gamma\epsilon} \end{bmatrix} \in \mathbb{R}^{3\times 3}. \tag{18}$$

If there exists an eigenvector

$$\widetilde{\mu}_n = \begin{bmatrix} \widetilde{\mu}_n^+ \\ \widetilde{\mu}_n^- \end{bmatrix} \begin{matrix} \} \in \mathbb{R}^2 \\ \} \in \mathbb{R} \end{matrix}$$

of A_n associated with its largest eigenvalue, which is such that

$$(1+\gamma\epsilon)\|\widetilde{\mu}_n^+\|^2 < (\Delta_n\widetilde{\mu}_n^-)^2 \tag{19}$$

then an optimal solution to Problem (17) *is*

$$\widetilde{\delta}_n = \frac{(1+\gamma\epsilon)^{1/2}}{\widetilde{\mu}_n^-} U_n\,\widetilde{\mu}_n^+ . \tag{20}$$

Otherwise, we have $\widetilde{\delta}_n = \Delta_n U_n \overline{\alpha}_n$ where

$$\overline{\alpha}_n \in \underset{\substack{\alpha_n\in\mathbb{R}^2\\ \|\alpha_n\|=1}}{\operatorname{Argmin}} \;\big(2\Delta_n\alpha_n^\top\operatorname{Re}\{c_n^*\widetilde{d}_n\} + \Delta_n^2\alpha_n^\top\operatorname{Re}\{\widetilde{d}_n\widetilde{d}_n^H\}\alpha_n\big). \tag{21}$$

Some guidelines concerning the numerical implementation of the solution provided by the above proposition are given in the next subsection.

C. Implementation issues

(i) Let $n \in \mathbb{N}$. According to (16), if $d_n = 0$ and $c_n = 0$, then $\widetilde{\delta}_n$ can be chosen arbitrarily and $\rho_n(\widetilde{\delta}_n) = 0$.

(ii) The matrix A_n defined by (18) can be expressed as

$$A_n = \operatorname{Re}\{a_n\}\operatorname{Re}\{a_n\}^\top + \operatorname{Im}\{a_n\}\operatorname{Im}\{a_n\}^\top \tag{22}$$

where $a_n = [\widetilde{d}_n \quad (1+\gamma\epsilon)^{-1/2}c_n]^\top$. Let us assume that either d_n or c_n is nonzero, so that a_n is also nonzero. The rank of matrix A_n is thus either equal to 1 or 2. An eigenvector $\widetilde{\mu}_n$ of A_n associated with its largest eigenvalue $\widetilde{\vartheta}_n > 0$ can then be decomposed as

$$\widetilde{\mu}_n = \widetilde{\mu}_n^{\mathrm{R}}\operatorname{Re}\{a_n\} + \widetilde{\mu}_n^{\mathrm{I}}\operatorname{Im}\{a_n\} \tag{23}$$

where $[\widetilde{\mu}_n^{\mathrm{R}} \quad \widetilde{\mu}_n^{\mathrm{I}}]^\top \in \mathbb{R}^2$ is an eigenvector of the (nonzero) Gram matrix

$$\begin{bmatrix} \|\operatorname{Re}\{a_n\}\|^2 & \operatorname{Re}\{a_n\}^\top\operatorname{Im}\{a_n\} \\ \operatorname{Re}\{a_n\}^\top\operatorname{Im}\{a_n\} & \|\operatorname{Im}\{a_n\}\|^2 \end{bmatrix}$$

associated with its largest eigenvalue $\widetilde{\vartheta}_n$. (The result is even valid in the case when the Gram matrix is of rank 1). Without loss of generality, it can be assumed that $\|\operatorname{Re}\{a_n\}\| \neq \|\operatorname{Im}\{a_n\}\|$ or $\operatorname{Re}\{a_n\}^\top\operatorname{Im}\{a_n\} \neq 0$, since the vectors a_n for which this condition is not satisfied belong to a set of zero measure, which makes that this case is almost never met in practice. Simple calculations show then that

$$\widetilde{\vartheta}_n = \frac{\|\operatorname{Re}\{a_n\}\|^2 + \|\operatorname{Im}\{a_n\}\|^2 + \widetilde{\eta}_n}{2} \tag{24}$$

with

$$\widetilde{\eta}_n = \sqrt{(\|\operatorname{Re}\{a_n\}\|^2 - \|\operatorname{Im}\{a_n\}\|^2)^2 + 4(\operatorname{Re}\{a_n\}^\top\operatorname{Im}\{a_n\})^2} \tag{25}$$

and the multiplicity of this eigenvalue is equal to 1. A corresponding eigenvector is

$$\begin{bmatrix} \widetilde{\mu}_n^{\mathrm{R}} \\ \widetilde{\mu}_n^{\mathrm{I}} \end{bmatrix} = \begin{bmatrix} \operatorname{Re}\{a_n\}^\top\operatorname{Im}\{a_n\} \\ \widetilde{\vartheta}_n - \|\operatorname{Re}\{a_n\}\|^2 \end{bmatrix}. \tag{26}$$

To know whether an optimal solution to Problem (17) is given by (20), it is thus sufficient to test Condition (19) for $\widetilde{\mu}_n$ given by (23) and (26).

(iii) If we discard the trivial case when $\Delta_n = 0$, Problem (21) is equivalent to maximizing the quadratic function

$$\alpha_n \mapsto 2\alpha_n^\top\operatorname{Re}\{c_n^*\widetilde{d}_n\} + \Delta_n\alpha_n^\top\operatorname{Re}\{\widetilde{d}_n\widetilde{d}_n^H\}\alpha_n \tag{27}$$

on the unit circle. Such kind of quadratic optimization problems on the unit sphere has been well-investigated in the literature [10, Chap. 12]. It can be shown that the problem reduces to searching the roots of a quartic polynomial.

IV. Application to spectrum analysis

In this part, we provide an illustration of the validity of our approach in a 2D spectrum analysis problem. More precisely, we consider an irregularly sampled complex-valued image which can be modeled as a sum of noisy 2D cisoids corrupted with a white circular Gaussian noise with zero-mean. The observed data of size $Q = Q_1 \times Q_2$ result from a random sampling at distinct locations τ_q in $[0,Q_1]\times[0,Q_2]$ with $q \in \{1,\ldots,Q\}$. The employed dictionary consists of the following functions:

$$(\forall\nu\in\mathbb{R}^2)\quad e_\nu = \big(\exp\big(\jmath\nu^\top\tau_q\big)\big)_{1\leqslant q\leqslant Q} \tag{28}$$

and we have thus

$$(\forall\nu\in\mathbb{R}^2)\quad e'_\nu = \begin{bmatrix} \big(\jmath\tau_{1,q}\exp\big(\jmath\nu^\top\tau_q\big)\big)_{1\leqslant q\leqslant Q} \\ \big(\jmath\tau_{2,q}\exp\big(\jmath\nu^\top\tau_q\big)\big)_{1\leqslant q\leqslant Q} \end{bmatrix}^\top \tag{29}$$

where, for every $q \in \{1,\ldots,Q\}$, the components of vector τ_q are denoted by $(\tau_{1,q},\tau_{2,q})$. If no prior information is available about the frequency contents of the 2D field, a standard choice is to uniformly sample the 2D frequency domain, so that, for every $n_1 \in \{1,\dots,N_1\}$ and $n_2 \in \{1,\dots,N_2\}$,

$$\theta_{(n_1-1)N_2+n_2} = 2\pi \begin{bmatrix} (n_1-1)/N_1 \\ (n_2-1)/N_2 \end{bmatrix}. \tag{30}$$

With the notation used in the previous sections, this corresponds to a scenario where $M = 2$ and $N = N_1 N_2$. In our experiments, the angular frequencies of the sparse components are not on the search grid. The proposed perturbed sparse estimation technique can however be applied in order to estimate them by choosing, for every $n \in \{1,\dots,N\}$, $\Delta_n = \sqrt{2}\pi/\min\{N_1,N_2\}$.

The signal is estimated by using Algorithm (13), where Φ is the squared Euclidean norm. The resulting global normalized root mean square estimation errors are provided in Table I, for $Q_1 = Q_2 = 8$, $N_1 = N_2 = 32$, and four values of the signal-to-noise ratio (SNR). As expected, for the proposed method, the estimation error evolves proportionally to the noise standard-deviation. A comparison is drawn with a standard basis pursuit approach using an ℓ_1 norm or an ℓ_0 cost. The ℓ_1-based solution was chosen as an initial value for our algorithm, as well as for the iterative hard thresholding approach associated with the basic ℓ_0 penalty. Note that the regularization parameter λ was chosen in an automatic manner from the observed data by an additional search stage, assuming that the variance of the noise is known. The plots shown in Figure 1 allow us to evaluate the good quality of the estimates typically obtained when identifying the cisoid parameters.

SNR (dB)	ℓ_1	ℓ_0	Proposed method
14.72	0.4519	0.1292	**0.01930**
19.72	0.4496	0.1291	**0.01088**
24.72	0.4480	0.1291	**0.00614**
29.72	0.4471	0.1292	**0.00348**

TABLE I
NORMALIZED ROOT MEAN SQUARE RECONSTRUCTION ERROR.

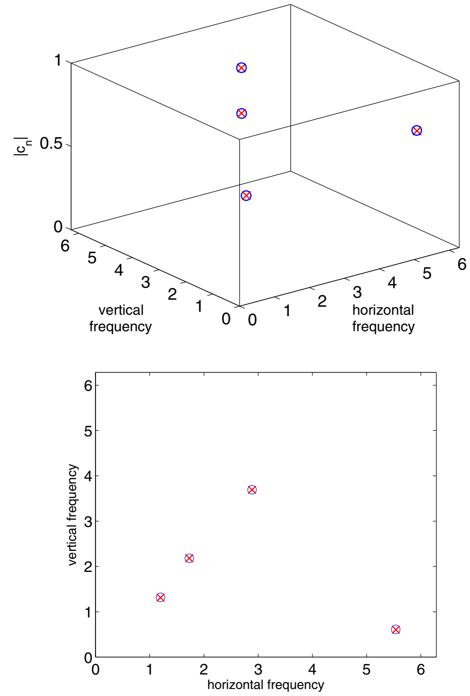

Fig. 1. Estimation results in the case of 4 cisoids (SNR = 24.72 dB): values of $(|c_n|)_{1\leqslant n\leqslant N}$ as a function of $(\theta_n)_{1\leqslant n\leqslant N}$ depicted with red crosses, and exact values $(|\bar{c}_n|)_{1\leqslant n\leqslant N}$ with blue circles (top); values of estimated angular frequency positions $(\nu_n)_{1\leqslant n\leqslant N}$ depicted with red crosses, and exact values $(\bar{\nu}_n)_{1\leqslant n\leqslant N}$ with blue circles (bottom).

V. CONCLUSION

In this paper, a new variational approach for the estimation of sparse signals has been proposed. The originality of our work resides in the use of a dictionary whose elements are parameterized by a real-valued vector of dimension greater than 1, while considering the context of estimation problems in which these parameters are known in an imprecise manner. A proximal iterative algorithm constitutes an efficient solution to the associated nonconvex optimization problem. Although few closed form expressions of the proximity operators of functions of several variables exist in the literature, it should be noticed that the proximity operator of the involved nonsmooth function has been fully characterized. Numerical experiments performed for a 2D spectrum analysis problem where the observations are irregularly randomly sampled allowed us to evaluate the good quality of the results generated with the proposed approach.

REFERENCES

[1] 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.

[2] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[3] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[4] A. Florescu, E. Chouzenoux, J.-C. Pesquet, and S. Ciochina, "A constrained optimization approach for complex sparse perturbed models," in *Proceedings of the Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013)*, Lausanne, Switzerland, 8-11 July 2013, pp. x–x+1.

[5] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, 2008.

[6] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Statist. Soc. A*, vol. 67, no. 2, pp. 301–320, 2005.

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

[8] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss-Seidel methods," *Math. Prog.*, vol. 137, no. 1-2, pp. 1–39, Feb. 2013.

[9] A. Beck, A. Ben-Tal and M. Teboulle, "Finding a global optimal solution for a quadratically constrained fractional quadratic problem with applications to the regularized total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 2, pp. 425–445, 2006.

[10] G. H. Golub and C. F. Van Loan, *Matrix computations*, John Hopkins University Press, Baltimore, 1996.