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.
Galați, 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 accounting for the noise statistics, and an arbitrary Lipschitz differentiable data fidelity term including 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 \( (\delta_n)_{1 \leq n \leq N} \) are small, we can perform the following first-order Taylor expansion:

\[
(\forall n \in \{1, \ldots, N\}) \quad e_{\nu_n} \approx e_{\nu_n} + e_{\nu_n}' \delta_n
\]

where \( e_{\nu_n}' \) is the Jacobian matrix of \( \nu \mapsto e_\nu = (c_n(q))_{1 \leq q \leq Q} \) at \( \nu_n \). For every \( q \in \{1, \ldots, Q\} \), the \( q \)-th line of \( e_{\nu_n}' \) thus corresponds to the gradient of the \( q \)-th component function \( \nu \mapsto e_{\nu_n}(q) \) at \( \nu_n \). With this approximation, Model (2) takes the following bilinear form

\[
\pi = \sum_{n=1}^{N} \left( c_n e_{\nu_n} + c_n e_{\nu_n}' \delta_n \right).
\]

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 \( \ell_1 \) formulation is tailored 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 \( \ell_1 \) regularization is developed. The difference is that, in this paper, we adopt a different formulation where an \( \ell_0 \) cost is employed for the minimization process, instead of its \( \ell_1 \) convex relaxation, and the perturbations \( (\delta_n)_{1 \leq n \leq N} \) are constrained to satisfy the following inequalities:

\[
(\forall n \in \{1, \ldots, N\}) \quad \|\delta_n\| \leq \Delta_n,
\]

where \( \theta_n \in \mathbb{R}^M \) is some given value and \( \delta_n \in \mathbb{R}^M \) 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 \( (\delta_n)_{1 \leq n \leq N} \) are small, we can perform the following first-order Taylor expansion:

\[
(\forall n \in \{1, \ldots, N\}) \quad e_{\nu_n} \approx e_{\nu_n} + e_{\nu_n}' \delta_n
\]

where \( e_{\nu_n}' \) is the Jacobian matrix of \( \nu \mapsto e_\nu = (c_n(q))_{1 \leq q \leq Q} \) at \( \nu_n \). For every \( q \in \{1, \ldots, Q\} \), the \( q \)-th line of \( e_{\nu_n}' \) thus corresponds to the gradient of the \( q \)-th component function \( \nu \mapsto e_{\nu_n}(q) \) at \( \nu_n \). With this approximation, Model (2) takes the following bilinear form

\[
\pi = \sum_{n=1}^{N} \left( c_n e_{\nu_n} + c_n e_{\nu_n}' \delta_n \right).
\]
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, \( \chi_S \) denotes the characteristic function of a set \( S \) which is equal to 0 on \( S \) and 1 elsewhere, and \( \iota_S \) denotes the indicator function of a set \( S \), which is equal to 0 on \( S \) and \( +\infty \) elsewhere. The transpose and conjugate operation for complex-valued vectors or matrices are denoted by \((\cdot)^T\) and \((\cdot)^H\), respectively.

## II. VARIATIONAL FORMULATION

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

\[
\begin{align*}
\text{minimize} & \quad \Phi \left( \sum_{n=1}^{N} (c_n e_{\theta_n} + c_n e_{\theta_n}^{\prime} \delta_n) - y \right) \\
& \quad + \lambda e_0(c) + \frac{\varepsilon}{2} \|c\|^2 \quad (7)
\end{align*}
\]

where

\[
B = \{(\delta_n)_{1 \leq n \leq N} \in (\mathbb{R}^M)^N \mid (\forall n \in \{1, \ldots, N\}) \|\delta_n\| \leq \Delta_n\}, \quad \Phi: \mathbb{C}^Q \rightarrow \mathbb{R} \text{ is the data-fidelity term which is often chosen equal to the negative-log-likelihood of the noise corrupting the observations, } \lambda \in (0, +\infty) \text{ is a regularization constant serving to promote sparsity, and } \varepsilon \in [0, +\infty). \text{ The last term plays a role similar to an elastic net regularization [6].}
\]

Let us now define the matrices \( E = [e_{\theta_1}, \ldots, e_{\theta_N}] \in \mathbb{C}^{Q \times N} \), \( E' = [e_{\theta_1}^{\prime}, \ldots, e_{\theta_N}^{\prime}] \in \mathbb{C}^{Q \times MN} \), and let us introduce the variable \( d = [d_1^{\top}, 1, \ldots, c_N e_{\delta_n}^{\prime}]^{\top} \in \mathbb{C}^{MN} \). In addition, let the function \( \Psi \) be defined as

\[
\Psi(c, d) = \sum_{n=1}^{N} \psi_n(c_n, d_n) \quad (9)
\]

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

\[
\psi_n(c_n, d_n) = \lambda \chi_{\{0\}}(c_n) + \iota_{S_n}(c_n, d_n) + \frac{\varepsilon}{2} |c_n|^2,
\]

and \( S_n \) is the closed cone given by

\[
S_n = \{(c_n, d_n) \in \mathbb{C} \times \mathbb{C}^M \mid \exists \delta_n \in \mathbb{R}^M \text{ with } \|\delta_n\| \leq \Delta_n, d_n = c_n \delta_n \}. \quad (11)
\]

Then, Problem (7) is equivalent to minimizing function

\[
(c, d) \mapsto \Phi \left( [E E'] \begin{bmatrix} c & d \end{bmatrix} - y \right) + \Psi(c, d). \quad (12)
\]

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

## III. PROPOSED ALGORITHM

### A. Algorithm form

If we assume that \( \Phi \) is a differentiable function, the previous split form of the objective function suggests the use of a forward-backward algorithm [7].

\[
\begin{align*}
\varepsilon(0) & \in \mathbb{C}^N, d(0) \in \mathbb{C}^{MN} \\
0 < \gamma \leq \bar{\gamma} < \|EE' + E'(E'')^H\|^{-1} \\
\text{For } k = 0, 1, \ldots \\
\gamma(k) & \in [\gamma, \bar{\gamma}] \\
D(k) & = \nabla \Phi \left( [E E'] \begin{bmatrix} c(k) & d(k) \end{bmatrix} - y \right) \\
(c_n(k+1), d_n(k+1))_{1 \leq n \leq N} & = (\text{prox}_{\gamma(k) \psi_n}(c_n(k), d_n(k)))_{1 \leq n \leq N}. \quad (13)
\end{align*}
\]

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

\[
(\forall u \in \mathcal{H}) \quad \text{prox}_{\varphi}(u) = \text{Argmin}_{v \in \mathcal{H}} \left\{ \frac{1}{2} \|u - v\|^2 + \varphi(v) \right\}. \quad (14)
\]

Although the uniqueness of \( \text{prox}_{\varphi}(u) \) is guaranteed when \( \varphi \) 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 \( \text{prox}_{\gamma \psi_n} \) 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

\[
(\forall \delta_n \in \mathbb{R}^M) \quad \rho_n(\delta_n) = \frac{|c_n + \bar{\delta}_n|_{d_n}^2}{1 + \gamma \epsilon + \|\delta_n\|^2} \quad (15)
\]

where \( \rho_n \) is the function defined as

\[
(\forall \delta_n \in \mathbb{R}^M) \quad \rho_n(\delta_n) = \frac{|c_n + \bar{\delta}_n|_{d_n}^2}{1 + 2\gamma \epsilon + \|\delta_n\|^2} \quad (16)
\]

and

\[
\bar{\delta}_n \in \text{Argmax}_{\delta_n \in \mathbb{R}^M, \|\delta_n\| \leq \Delta_n} \rho_n(\delta_n). \quad (17)
\]
The determination of $\bar{\delta}_n$ 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 $(\bar{\delta}_n)_{1 \leq n \leq N}$ can be obtained.

**Proposition 2.** Assume that $M \geq 2$. 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 $\text{span}\{\text{Re}(d_n), \text{Im}(d_n)\}$, let $\bar{d}_n = U_n^T d_n \in \mathbb{C}^2$, and let

$$A_n = \begin{bmatrix} \text{Re}(\bar{c}_n^* \bar{d}_n^H) & \text{Re}(\bar{c}_n d_n^H) \\ \text{Re}(\bar{c}_n^* d_n^H)^T & (1 + \gamma\epsilon)^{1/2} \end{bmatrix} \in \mathbb{R}^{3 \times 3}. \quad (18)$$

If there exists an eigenvector 

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

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

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

then an optimal solution to Problem (17) is

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

Otherwise, we have $\bar{\delta}_n = \Delta_n U_n \bar{\pi}_n$, where

$$\bar{\pi}_n \in \text{Argmin}_{\alpha_n \in \mathbb{R}^2} \frac{2\Delta_n \alpha_n^T \text{Re}(\bar{c}_n^* \bar{d}_n^H) + \Delta_n^2 \alpha_n^T \text{Re}(\bar{d}_n^H \bar{d}_n^H)^T \alpha_n}{\|\alpha_n\|_2^2} \quad (21)$$

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 $\delta_n$ can be chosen arbitrarily and $\rho_n(\delta_n) = 0$.

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

$$A_n = \text{Re}(a_n) \text{Re}(a_n)^T + \text{Im}(a_n) \text{Im}(a_n)^T \quad (22)$$

where $a_n = [\bar{d}_n^H (1 + \gamma\epsilon)^{-1/2} c_n]^T$. 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 $\bar{\mu}_n$ of $A_n$ associated with its largest eigenvalue $\bar{\delta}_n > 0$ can then be decomposed as

$$\bar{\mu}_n = \bar{\mu}_n^R \text{Re}(a_n) + \bar{\mu}_n^I \text{Im}(a_n) \quad (23)$$

where $\bar{\mu}_n^R$ and $\bar{\mu}_n^I$ are an eigenvector of the (nonzero) Gram matrix

$$\begin{bmatrix} \|\text{Re}(a_n)\|^2 & \text{Re}(a_n)^T \text{Im}(a_n) \\ \text{Re}(a_n)^T \text{Im}(a_n) & \|\text{Im}(a_n)\|^2 \end{bmatrix} \quad (24)$$

associated with its largest eigenvalue $\bar{\delta}_n$. (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 $\|\text{Re}(a_n)\| \neq \|\text{Im}(a_n)\|$ or $\text{Re}(a_n)^T \text{Im}(a_n) \neq 0$, 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

$$\bar{\delta}_n = \frac{\|\text{Re}(a_n)\|^2 + \|\text{Im}(a_n)\|^2 + \bar{\eta}_n}{2} \quad (25)$$

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

$$\begin{bmatrix} \bar{\nu}_n^R \\ \bar{\nu}_n^I \end{bmatrix} = \begin{bmatrix} \text{Re}(a_n)^T \text{Im}(a_n) \\ -\|\text{Re}(a_n)\|^2 \|\text{Im}(a_n)\|^2 \end{bmatrix} \quad (26)$$

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

$$\alpha_n \mapsto 2\alpha_n^T \text{Re}(\bar{c}_n^* \bar{d}_n^H) + \Delta_n \alpha_n^T \text{Re}(\bar{d}_n^H \bar{d}_n^H) \alpha_n \quad (27)$$

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 $\tau_q$ in $[0, Q_1] \times [0, Q_2]$ with $q \in \{1, \ldots, Q\}$. The employed dictionary consists of the following functions:

$$(\forall \nu \in \mathbb{R}^2) \quad e_{\nu} = (\exp(\mu^T \tau_q))_{1 \leq q \leq Q} \quad (28)$$

and we have thus

$$(\forall \nu \in \mathbb{R}^2) \quad e'_{\nu} = \begin{bmatrix} \tau_{1,q} \exp(\mu^T \tau_q) & \cdots & \tau_{Q,q} \exp(\mu^T \tau_q) \end{bmatrix}_{1 \leq q \leq Q}^T \quad (29)$$

where, for every $q \in \{1, \ldots, Q\}$, the components of vector $\tau_q$ 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 $n_1 \in \{1, \ldots, N_1\}$ and $n_2 \in \{1, \ldots, N_2\}$,

$$
\theta_{(n_1-1)N_2+n_2} = 2\pi \left[ \frac{(n_1-1)/N_1}{(n_2-1)/N_2} \right].
$$

(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 $n \in \{1, \ldots, N\}$, $\Delta_n = \sqrt{2\pi} / \min\{N_1, N_2\}$.

The signal is estimated by using Algorithm (13), where $\Phi$ 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 $\ell_1$ norm or an $\ell_0$ cost. The $\ell_1$-based solution was chosen as an initial value for our algorithm, as well as for the iterative hard thresholding approach associated with the basic $\ell_0$ penalty. Note that the regularization parameter $\lambda$ 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.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>$\ell_1$</th>
<th>$\ell_0$</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.72</td>
<td>0.4519</td>
<td>0.1292</td>
<td>0.01930</td>
</tr>
<tr>
<td>19.72</td>
<td>0.4496</td>
<td>0.1291</td>
<td>0.01088</td>
</tr>
<tr>
<td>24.72</td>
<td>0.4480</td>
<td>0.1291</td>
<td>0.00614</td>
</tr>
<tr>
<td>29.72</td>
<td>0.4471</td>
<td>0.1292</td>
<td>0.00348</td>
</tr>
</tbody>
</table>

TABLE I

NORMALIZED ROOT MEAN SQUARE RECONSTRUCTION ERROR.

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