A Sparsity-Based Method for the Estimation of Spectral Lines From Irregularly Sampled Data

Sébastien Bourguignon, Hervé Carfantan, and Jérôme Idier

Abstract—We address the problem of estimating spectral lines from irregularly sampled data within the framework of sparse representations. Spectral analysis is formulated as a linear inverse problem, which is solved by minimizing an ℓ^1-norm penalized cost function. This approach can be viewed as a Basis Pursuit De-Noising (BPDN) problem using a dictionary of cisoids with high frequency resolution. In the studied case, however, usual BPDN characterizations of uniqueness and sparsity do not apply.

This paper deals with the ℓ^1-norm penalization of complex-valued variables, that brings satisfactory prior modeling for the estimation of spectral lines. An analytical characterization of the minimizer of the criterion is given and geometrical properties are derived about the uniqueness and the sparsity of the solution.

An efficient optimization strategy is proposed. Convergence properties of the Iterative Coordinate Descent (ICD) and Iterative Reweighted Least-Squares (IRLS) algorithms are first examined. Then, both strategies are merged in a convergent procedure, that takes advantage of the specificities of ICD and IRLS, considerably improving the convergence speed. The computation of the resulting spectrum estimator can be implemented efficiently for any sampling scheme.

Algorithm performance and estimation quality are illustrated throughout the paper using an artificial data set, typical of some astrophysical problems, where sampling irregularities are caused by day/night alternation. We show that accurate frequency location is achieved with high resolution. In particular, compared with sequential Matching Pursuit methods, the proposed approach is shown to achieve more robustness regarding sampling artifacts.

Index Terms—Algorithms, estimation, inverse problems, optimization methods, sparse representations, spectral analysis, time series.

I. INTRODUCTION

SPECTRAL analysis is a very important topic in signal processing, that is essential for many application fields of physics. The finite length of the time coverage, as well as the discrete nature of the data, considerably complicate the problem of estimating the Fourier spectrum from a given time series. More precisely, the *observed spectrum*—the Fourier spectrum of the data—is the convolution of the true spectrum

by the *spectral window*, whose lobes set the limits of the observed spectrum in terms of frequency resolution and amplitude precision.

In some applications, experimental conditions or instrument limitations may cause the data to be irregularly time-spaced. Much fewer attention was paid to the spectral analysis problem in this case. In particular, methods based on parametric representations [1], that efficiently achieve high resolution with regularly sampled data, cannot be applied directly. In astrophysics, moreover, specific observational conditions make the problem of spectrum estimation particularly difficult. For instance, the Earth's daily rotation and annual revolution may cause periodic gaps in the available data. As a consequence, the magnitude of the spectral window defined as

$$\mathcal{W}(f) = \frac{1}{N} \sum_{n=0}^{N-1} \exp\left(j2\pi f t_n\right) \tag{1}$$

for sampling times $t_n, n = 0 \ldots N - 1$, may show high secondary lobes, generating false peaks in the observed Fourier spectrum. An example of spectral window corresponding to the simulated data used in Section V is given in Fig. 5(b). It illustrates the difficulty of spectral analysis in this case.

With the aim of estimating spectral lines, fitting a multisine model to the available data is a hard task. Indeed, the likelihood of the data given a multisine model admits many local modes, even in the case of regular sampling [2]. *Prewhitening* techniques, widely used in astrophysics, perform the iterative deconvolution of the observed spectrum [3], [4]. At each iteration, the highest peak of the residual spectrum is removed, together with the associated contribution of the spectral window, generating a sequence of estimated spectral lines. Ad-hoc refinements such as the use of a *clean gain* for the CLEAN method [5] or additional local optimization steps for CLEANEST [6], aim to improve the efficiency of prewhitening techniques. Such methods can be viewed as implementations of a *Matching Pursuit* [7] estimation strategy. The resulting methodology, however, is still sensitive to sampling artifacts and may lead to false detections [8].

During the past decade, spectral analysis has been addressed as a linear inverse problem, and such an approach became a serious alternative to parametric methods to achieve high resolution [9]–[13]. The spectrum is discretized on a fixed frequency grid, and spectral lines are estimated by locating a few nonzero values in the corresponding amplitude vector, i.e., by reconstructing a *sparse* vector. The use of such a linear model gives more robustness to the estimation compared to parametric methods, especially regarding sampling artifacts [8]. On the other hand, it results in a considerable increase of the

number of unknowns, so efficient estimation algorithms are required. Specific algorithms were proposed in [9], [10], based on iterative weighted norm minimizations. For an adequately chosen norm, the estimator, implicitly defined as the fixed point of the algorithm, concentrates the energy on a few components. In [11]–[13], the spectral estimator was defined as the minimizer of a least-squares misfit criterion, penalized by a term that favors line spectra. Such an approach can be viewed as the Bayesian *Maximum A Posteriori* estimation for some prior distribution expressing sparsity.

Following the formulation in [11]–[13] of the spectral analysis problem, this paper studies its regularization by the ℓ^1-norm. Compared to other penalization functions that were proposed in this framework, the use of the ℓ^1-norm leads to exactly sparse estimators while preserving the convexity of the underlying optimization problem. Moreover, the specific structure of the functional to optimize allows a powerful algorithmic implementation for any sampling scheme, while the algorithms in [11], [12] are essentially designed for regularly sampled data.

In a more general context, the domain of *sparse representations* has become a very active field of research: a given data set is represented as a linear combination of a small number of elementary signals, called *atoms*, drawn from a large linearly dependant collection, called *dictionary* or *redundant basis*. The goal is to represent a wide class of signals, including nonstationarities, so the dictionary is usually composed of different families of functions with complementary structures—sinusoids, spikes, wavelets, shapelets, curvelets, etc. Given the important size of the dictionary, seeking for the smallest combination of atoms is usually unfeasible, and the problem is generally substituted by the minimization of the ℓ^1-norm, which is a convex problem: Chen *et al.* [14] named this methodology the *Basis Pursuit*. The approach adopted in this paper formulates a *Basis Pursuit De-Noising* (BPDN) problem [14]: the sparsest representation of the noisy data is searched in a dictionary made of an arbitrarily large number of cisoids. Some important differences arise with the classical *Basis Pursuit* literature, however:

- *Basis Pursuit* usually considers dictionaries made of structurally different atoms. When considering a high number of cisoids—high-resolution analysis—the corresponding dictionary has a high *mutual coherence*, in the sense defined in [15] for example. Thus, the conditions that guarantee the sparsity and the uniqueness of the solution, that were recently established in [16], [17], are not satisfied for this problem.
- The spectral analysis problem considers complex-valued variables, for which much fewer attention was paid compared to the real case. In particular, efficient algorithms such as Interior Point [14], [18] or Least Angle Regression [19], [20], are specific to real variables, for which optimization can be formulated as a quadratic program. With complex variables, however, optimization formulates a second-order cone (SOC) programming problem [21], [22], that can be solved by Interior Point algorithms, e.g., [23].

This paper brings some advances concerning sparse modeling via the penalization by the ℓ^1-norm. In particular, we show that the ℓ^1 penalization operating on the modulus of the complex variables gives more accurate estimates than its real-valued counterpart operating on real and imaginary parts, as proposed in [24]. Structural properties are investigated for the minimizer of the ℓ^1 penalized criterion, as an alternative to usual characterizations of BPDN solutions that are not applicable here, such as a low *mutual coherence* of the dictionary [16], [17] or positivity of the *Exact Recovery Coefficient* [17]. In application to the spectral analysis of regularly or irregularly sampled data, conditions are given for the uniqueness and the sparsity of the estimator. In terms of estimation quality, such a methodology is shown to bring more robustness regarding sampling artifacts, compared to *Matching Pursuit* methods. The paper also brings new elements concerning algorithmic issues. We consider two different optimization strategies that have been recently proposed for such problems, namely, Iterative Coordinate Descent (ICD) [18], [35] and Iterative Reweighted Least-Squares (IRLS) [26], and we propose a mixed strategy using *spacer steps* [27, Ch. 7]. The result is a convergent algorithm, that takes advantage of the specificities of the ICD and IRLS procedures, considerably improving the convergence speed. Simulations on artificial data show that the proposed procedure allows one to consider a very high frequency resolution, yielding a much lower computational cost than the SeDuMi package in the SOC programming framework [23].

The paper is structured as follows. In Section II, line spectra estimation is set as an underdetermined linear inverse problem. The penalization by the ℓ^1-norm of complex variables is studied and an analytical characterization of the minimum of the penalized criterion is given, that also yields a physical interpretation to the regularization parameter. Section III establishes conditions for the minimizer of any ℓ^1 penalized criterion to be unique. In application to high-resolution spectral analysis, properties are given for both regular and irregular sampling cases. Optimization algorithms are studied in Section IV, that ends with the design of a convergent and computationally efficient procedure. Finally, Section V is devoted to simulation results.

II. STATEMENT AND PRIOR MODELING

A. Linear Formulation of Spectral Analysis

The formulation of spectral analysis as a linear inverse problem [9]–[13] makes use of an inverse Fourier Transform model. The irregularly sampled time series $(t_n, y_n)_{n=0,...,N-1}$ is modeled as a noisy sum of a large number of cisoids with discretized frequencies $f_k = k f_{\max}/P$, $k \in \mathcal{P}$, $\mathcal{P} = \{-P, \ldots, P\}$. Note that for irregular sampling, parameter $f_{\max}$ is not limited to the Nyquist limit, allowing the reconstruction of the spectrum in a wider band than with regularly sampled data [28]. Formally, the reconstruction of the complex-valued spectral amplitudes $\boldsymbol{x} = [x_{-P}, \ldots, x_P]^{\mathrm{t}}$ from the data $\boldsymbol{y} = [y_0, \ldots, y_{N-1}]^{\mathrm{t}}$ is addressed

$$\boldsymbol{y} = \mathbf{W}\boldsymbol{x} + \boldsymbol{\epsilon} \qquad (2)$$

where

$$\mathbf{W} = \left\{ \frac{1}{\sqrt{N}} \exp(j 2\pi f_k t_n) \right\}_{n=0,...,N-1, k \in \mathcal{P}}$$

is a $N \times (2P + 1)$ matrix and $\boldsymbol{\epsilon}$ stands for perturbations such as model errors and observation noise. For irregularly sampled data, $\mathbf{W}$ does not identify with an inverse Discrete Fourier Transform operator, so it looses structural properties such as the orthonormal and circulant characters [29]. However, it can be shown that $\mathbf{W}^{\dagger}\mathbf{W}$ is still a Toeplitz matrix, with

$$\{\mathbf{W}^{\dagger}\mathbf{W}\}_{k,\ell} = \mathcal{W}\left(\frac{\ell - k}{P} f_{\max}\right). \tag{3}$$

To obtain high frequency resolution, which is our goal in this paper, one has to set $2P + 1 \gg N$, so the least square solution to problem (2) is under-determined.

B. Regularization Framework

We adopt a regularization approach in which the estimated spectrum is the minimizer of the penalized criterion

$$J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 + \lambda R(\boldsymbol{x}) \tag{4}$$

where the penalization function $R(\boldsymbol{x})$ is adequately designed for the estimation of spectral *lines*, i.e., for *sparse* vectors $\boldsymbol{x}$.

Quadratic penalization functions—ℓ^2-norms—yield *ridge* regression solutions [26]. They lead to a windowed DFT in the regular sampling case [30]. Although this property does not generalize to irregular sampling, empirical results established the inadequacy of ℓ^2-norms for line spectra estimation in this case [29]. Sacchi *et al.* [11] considered a log-Cauchy penalization, which is suited to sparse solutions but leads to a nonconvex criterion J, possibly admitting local minima. In [12], the $\ell^2\ell^1$ penalization function $R_{21}(\boldsymbol{x}) = \sum_k \sqrt{s^2 + |x_k|^2}$ was proposed, which gives *spiky* estimators for small positive values of parameter s, and leads to a strictly convex criterion. Note that the optimization strategies proposed in both papers are especially efficient for regularly sampled data: in this case, matrix $\mathbf{W}$ has a Fourier structure, which allows to compute high dimensional matrix products and system inversions at a reasonable cost using FFT.

In the last decade, the efficiency of the ℓ^1-norm to address sparsity has become a prolific field of research [10], [14]–[17], [31], [32]. This is the direction followed here for spectral analysis. In other words, the spectrum estimator is defined as $\hat{\boldsymbol{x}}_{\mathbb{C}} = \arg\min J_{\mathbb{C}}(\boldsymbol{x})$ with

$$J_{\mathbb{C}}(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|_2^2 + \lambda R_{\mathbb{C}}(\boldsymbol{x}) \tag{5}$$

where $R_{\mathbb{C}}$ is the ℓ^1-norm for complex-valued vectors

$$R_{\mathbb{C}}(\boldsymbol{x}) = \sum_k |x_k| = \sum_k \sqrt{\Re(x_k)^2 + \Im(x_k)^2}. \tag{6}$$

The reason motivating this choice is twofold.

- First, ℓ^1 penalization may be viewed as an intermediate choice leading to a strictly sparse solution, and still defining a convex (although not strictly) penalized criterion, with no local minima. In addition, the minimizer of $J_{\mathbb{C}}$ is almost always unique, provided it is sufficiently sparse; see Section III-C.
- Second, the specific structure of $J_{\mathbb{C}}$ brought by the ℓ^1-norm allows powerful algorithmic implementations.

Using $R_{\mathbb{C}}$ as a penalization function corresponds to the BPDN methodology [14], where estimator $\hat{\boldsymbol{x}}_{\mathbb{C}}$ is searched as the sparsest representation of the noisy data $\boldsymbol{y}$ in a dictionary of cisoids with arbitrarily thinly discretized frequencies. However, some substantial differences with the classical *Basis Pursuit* literature arise in our case, which are examined in Section II-D and Section III.

C. Interpretation of the Regularization Parameter

In a Bayesian interpretation, parameter λ can be viewed as an inverse signal-to-noise ratio (e.g., [8]). Here, a characterization of the minimum of criterion (5) is given, which provides a physical interpretation to parameter λ. It is a straightforward generalization of the condition derived in [33] in the real-valued case, that is also mentioned in [16].

Property 1: $\hat{\boldsymbol{x}}$ minimizes criterion (5) if and only if

$$\begin{cases} (i) & \forall k \text{ such that } |\hat{x}_k| = 0 : \quad |r_k| \leq \lambda, \\ (ii) & \forall k \text{ such that } |\hat{x}_k| \neq 0 : \quad r_k + \lambda \hat{x}_k / |\hat{x}_k| = 0 \end{cases}$$

with $\boldsymbol{r} = \mathbf{W}^{\dagger}(\mathbf{W}\hat{\boldsymbol{x}} - \boldsymbol{y})$.

The proof follows the same scheme as its real-variable counterpart, that can be found in [33]. Note that $\boldsymbol{r}$ is merely the discretized Fourier spectrum of the residual $\mathbf{W}\hat{\boldsymbol{x}} - \boldsymbol{y}$. Thus, λ is an upper bound of the Fourier spectrum of the residual, discretized on the frequency grid, and this bound is attained at all the locations of the nonzero components in $\hat{\boldsymbol{x}}$.

Let us remark that estimator $\hat{\boldsymbol{x}}_{\mathbb{C}}$ may suffer from some bias in the amplitude estimation due to the penalization term in (5). In practice, it is preferable to re-estimate the amplitudes of the detected frequencies by least-squares (see Section V). Thus, the maximum of the final residual spectrum is usually lower than λ. However, Property 1 can help to tune the regularization parameter. Let $\lambda_{\max}$ be the maximum of the Fourier spectrum of the data: $\lambda_{\max} = \max|\mathbf{W}^{\dagger}\boldsymbol{y}|$. For $\lambda \geq \lambda_{\max}$, estimator $\hat{\boldsymbol{x}}_{\mathbb{C}}$ is identically zero. Setting λ to some percentage, say 5%, of $\lambda_{\max}$ guarantees that the residual spectrum is lower than 5% of the maximum of the Fourier spectrum of the data, which heuristically gives satisfactory results. The use of automatic methods to choose the best value for parameter λ is still an open problem, which is out of the scope of this paper.

D. ℓ^1 Penalization for Complex-Valued Vectors

Much attention has been already paid to ℓ^1 penalization, but most of the studies have only addressed the case of real-valued vectors of unknowns [14], [16], [31], [32]. For instance, in the case of spectral analysis, BPDN was applied in [24] to a real-variable model with sine and cosine functions, instead of model (2) with cisoids. This allows to formulate the optimization as a quadratic programming problem, which can be tackled by powerful algorithms. However, the penalization function then reads $R_{\mathbb{R}}(\boldsymbol{x}) = \sum_k(|\Re(x_k)| + |\Im(x_k)|)$, which is obviously not equivalent to (6).

The example in Fig. 1 illustrates that the two alternatives are likely to provide quite different results. It corresponds to a simulated data set with five sinusoids presented in Section V. Spectral lines are estimated as the minimizers $\hat{\boldsymbol{x}}_{\mathbb{R}}$ and $\hat{\boldsymbol{x}}_{\mathbb{C}}$ of criterion (4) using penalizations $R_{\mathbb{R}}$ and $R_{\mathbb{C}}$, respectively. Since the
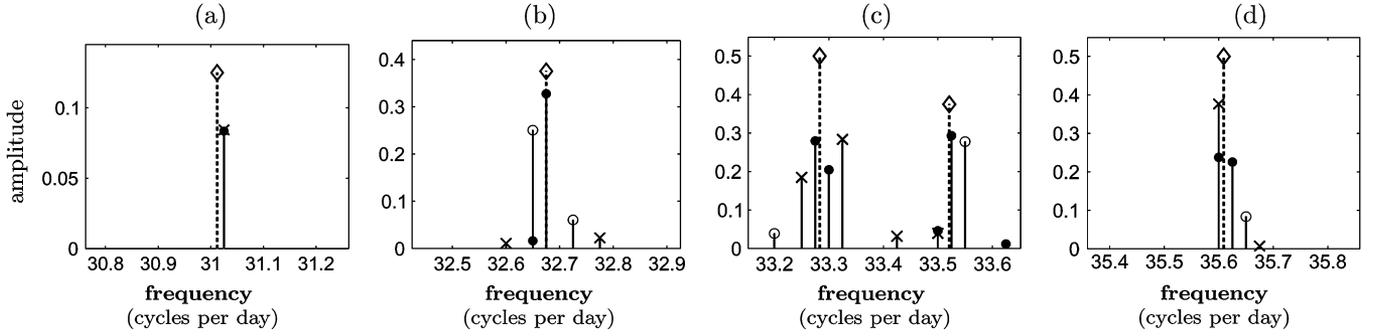
Fig. 1. Regularized estimation of the five spectral lines of the data in Section V using penalizations $R_\mathbb{C}$ and $R_\mathbb{R}$: zoom around the true frequencies. Parameter λ was set to $0.05\lambda_{\max}$. The black circles represents $|\hat{\boldsymbol{x}}_\mathbb{C}|$. The white circles and x-marks represent $|\Re(\hat{\boldsymbol{x}}_\mathbb{R})|$ and $|\Im(\hat{\boldsymbol{x}}_\mathbb{R})|$, respectively. The diamonds locate the true frequencies. (a) Frequency ν_1, (b) frequency ν_2 (c) frequencies ν_3 and ν_4, and (d) frequency ν_5.

true frequencies $\nu_{j,j=1,\ldots,5}$ do not belong to the reconstruction grid, $\hat{\boldsymbol{x}}_\mathbb{C}$ has two adjacent nonzero values around the true frequency for $\nu_{j,j=2,\ldots,5}$. On the other hand, both vectors $\Re(\hat{\boldsymbol{x}}_\mathbb{R})$ and $\Im(\hat{\boldsymbol{x}}_\mathbb{R})$ do possess a sparse structure, but nothing favors their nonzero values to be located at the same frequencies. Consequently, in Figs. 1(b)–(d), the real and imaginary parts of $\hat{\boldsymbol{x}}_\mathbb{R}$ have distinct supports, and none of them coincide with the correct one. The resulting frequency splitting in $\hat{\boldsymbol{x}}_\mathbb{R}$ is not satisfactory, since a true frequency is estimated by up to four nonzero components.

Note that no other value of the regularization parameter yields a better estimate $\hat{\boldsymbol{x}}_\mathbb{R}$. Fig. 2 represents the location of the nonzero components in $\hat{\boldsymbol{x}}_\mathbb{R}$ and $\hat{\boldsymbol{x}}_\mathbb{C}$ as a function of λ, zoomed around the two close frequencies (ν_3, ν_4). Acceptable estimates $\hat{\boldsymbol{x}}_\mathbb{R}$ are obtained for $\lambda < 0.15\lambda_{\max}$ only, since otherwise the line at frequency ν_1 is not detected. For frequency ν_3, $\hat{\boldsymbol{x}}_\mathbb{R}$ shows at least two nonzero components, separated by three grid steps. Moreover, the estimation of frequencies ν_3 and ν_4 obtained with $\hat{\boldsymbol{x}}_\mathbb{R}$ is always biased by more than one grid step, which is not the case with $\hat{\boldsymbol{x}}_\mathbb{C}$. Thus, sparsity information is more accurately represented by $R_\mathbb{C}$. That is, in terms of *Basis Pursuit*, a dictionary of cisoids is more relevant than a dictionary where elementary atoms are sine and cosine functions.

III. UNIQUENESS CONDITIONS OF THE SPARSE SOLUTION

Several recent works addressed the question of the uniqueness of the minimizer of any ℓ^1 penalized criterion. Indeed, [16], [17] propose sufficient conditions on the recovery of the sparsest solution that are also sufficient conditions for the minimizer to be unique. These conditions are based on the *mutual coherence* of the dictionary and the *Exact Recovery Coefficient* of a set of indexes. We first show that such conditions, however, are far too restrictive in our case. Then, alternate characterizations are given that can be applied to the spectral analysis problem.

A. Inefficiency of Mutual Coherence and ERC Based Conditions for Spectral Analysis

Let $\mathbf{A}$ be any redundant dictionary, i.e., a $N \times M$ matrix $(M > N)$ with normalized columns $\boldsymbol{a}_k$ (atoms). The mutual coherence μ of the dictionary is defined as [15]

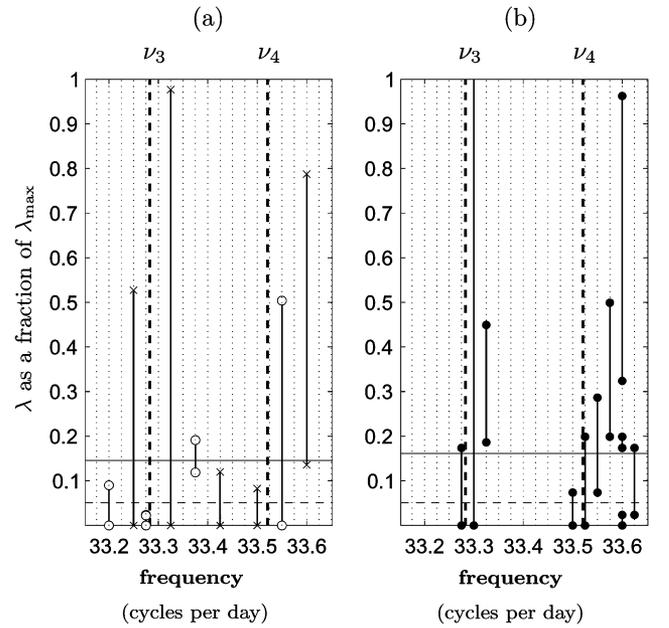$$\mu = \max_{k \neq \ell} |\boldsymbol{a}_k^\dagger \boldsymbol{a}_\ell|.$$

Fig. 2. (a) Locations of the nonzero components of $\hat{\boldsymbol{x}}_\mathbb{R}$ for different values of λ between 0 and $\lambda_{\max}$, in the neighborhood of frequencies (ν_3, ν_4). Segments with endpoints marked by a circle (resp., a x-mark) indicate the nonzero components of $\Re(\hat{\boldsymbol{x}}_\mathbb{R})$ (resp., $\Im(\hat{\boldsymbol{x}}_\mathbb{R})$). (b) Locations of the nonzero components of $\hat{\boldsymbol{x}}_\mathbb{C}$. The two vertical dashed lines locate ν_3 and ν_4, and vertical dotted lines represent the frequency reconstruction grid. The horizontal full line indicate the upper bound for acceptable solutions. The horizontal dashed line indicate the heuristic tuning $\lambda = 0.05\lambda_{\max}$.

Let $\mathcal{K} \subset \{1, \ldots, M\}$ be a set of indexes. The ERC of $\mathcal{K}$ is defined as [17]

$$\mathrm{ERC}(\mathcal{K}) = 1 - \max_{k \notin \mathcal{K}} \|\mathbf{A}^+ \boldsymbol{a}_k\|_1$$

where $\mathbf{A}^+$ is the pseudo inverse of matrix $\mathbf{A}$.

Results in [16, Th. 2 and 3] and [17, Corollary 9] show that the smaller μ, the weaker the conditions of recovery of the sparsest representation using the ℓ^1-norm. In the classical *Basis Pursuit* literature, $\mathbf{A}$ is often defined as the union of orthonormal bases, the latter being structurally different enough one from each other to ensure that μ takes a low value. In our case, the dictionary
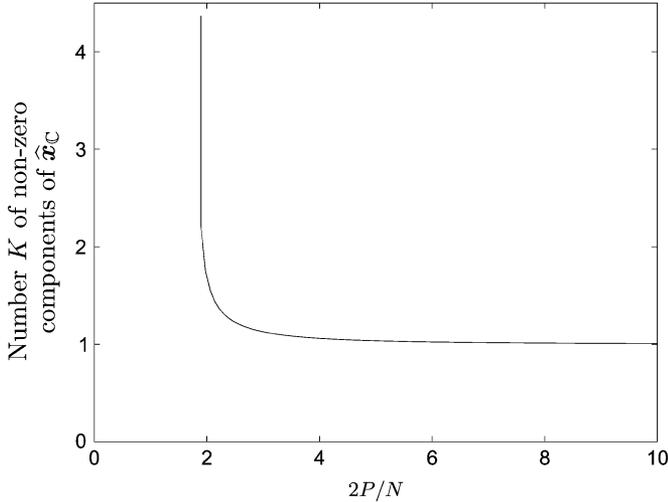
Fig. 3. Upper bound on the number K of nonzero components of $\hat{x}_{\mathbb{C}}$ such that uniqueness conditions in [16, Th. 3] apply.

used for high-resolution spectral analysis has a very coherent structure. According to (3), the coherence of $\mathbf{W}$ is given by

$$\mu = \max_{k \neq \ell} \left| \mathcal{W}\left(\frac{\ell - k}{P} f_{\max} \right) \right| \quad (7)$$

where $\mathcal{W}$ is the spectral window defined by (1). Let us remark that μ is not lower than the magnitude of the inner product of adjacent columns

$$\mu_0 = |\mathcal{W}(f_{\max}/P)|. \quad (8)$$

As high frequency resolution is required, i.e., $2P + 1 \gg N$, μ_0 corresponds to the magnitude of the spectral window function close to the origin, inside its central lobe. Therefore, μ is close to $|\mathcal{W}(0)| = 1$. See Appendix I for a quantitative analysis, where lower bounds are derived for μ.

The conditions in [16], [17] link uniqueness and sparsity of the minimizer of $J_{\mathbb{C}}$, for a given value of μ. Let K stand for the number of nonzero components of a solution $\hat{x}_{\mathbb{C}}$. The sharpest result is found in [16, Th. 3]. It says that a minimizer $\hat{x}_{\mathbb{C}}$ is the unique one if $K < (1 + 1/\mu)/2$. Hence, the lower bound derived for μ in Appendix I can be easily converted into an upper bound $K_{\max}$ for K. Fig. 3 displays such a bound as a function of the ratio $2P/N$, in the regular sampling case, with $f_{\max} = 1/(2T_s)$, where T_s is the sampling period. It clearly indicates that in the high-resolution framework, such a bound is very restrictive. Moreover, the situation is not more favorable in the irregular sampling case. Actually, Appendix I shows that the same bound $K_{\max}$ applies to all sampling schemes with the same empirical variance of the sampling instants. In other words, the uniqueness properties based on a condition of low coherence of the dictionary are not compatible with a high-resolution analysis.

Tropp [17, Th. 8] obtained a weaker, solution-dependent condition, that allows to verify a posteriori if the obtained solution is the sparsest one. If so, then the solution is unique. Although

no general result was obtained for such a condition in application to the considered spectral analysis problem, the *Exact Recovery Coefficient* (ERC) was always negative in practical cases, so that the condition in [17] was never verified. Simulations in Section V will reveal the inadequacy of both ERC and coherence-based conditions to establish the unicity of the minimizer of (5) in our case.

Nevertheless, it is possible to establish alternate uniqueness conditions that are more effective in the context of high-resolution spectral analysis, as shown below.

B. A Uniqueness Condition Based on the Unique Representation Property

Let $\mathbf{A}$ be a $N \times M$ matrix with $M > N$. Let us derive some uniqueness conditions for the minimizer of

$$H(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|^2 + \lambda R_{\mathbb{C}}(\boldsymbol{x}).$$

Let us assume that $\mathbf{A}$ satisfies the *Unique Representation Property* (URP), i.e., that any N columns of $\mathbf{A}$ are linearly independent [10]. In the following, the *support* $\mathrm{supp}(\boldsymbol{x})$ denotes the set of indexes where $\boldsymbol{x}$ is nonzero. Lemma 1 shows that sparse minimizers of H are isolated.

Lemma 1: Suppose that $\mathbf{A}$ satisfies the URP. Then, for all $\mathcal{K} \subset \{1, \ldots, M\}$ with Card $\mathcal{K} \leq N$, there is at most one minimizer of H supported by $\mathcal{K}$.

Proof: Suppose that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ minimize H and that they have the same support $\mathcal{K}$. Let $K = \mathrm{card}\,\mathcal{K}$, and $\mathbf{A}_{\mathcal{K}}$ be the $N \times K$ matrix formed by the columns of $\mathbf{A}$ with indexes in $\mathcal{K}$. Then, minimization of H over the set of vectors with support belonging to $\mathcal{K}$ amounts to minimize

$$H_{\mathcal{K}}(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}_{\mathcal{K}}\boldsymbol{u}\|_2^2 + \lambda R_{\mathbb{C}}(\boldsymbol{u}), \boldsymbol{u} \in \mathbb{C}^K.$$

Since the columns of $\mathbf{A}_{\mathcal{K}}$ are linearly independent, the quadratic term in $H_{\mathcal{K}}$ is strictly convex in $\boldsymbol{u}$, and so is $H_{\mathcal{K}}$. Therefore, $H_{\mathcal{K}}$ has a unique minimizer, say $\hat{\boldsymbol{u}}$. It is easy to show that $\hat{\boldsymbol{u}} = \hat{\boldsymbol{x}}_{\mathcal{K}}$, where $\hat{\boldsymbol{x}}_{\mathcal{K}} = [\hat{\boldsymbol{x}}_k]_{k \in \mathcal{K}}$. It is also true that $\hat{\boldsymbol{u}} = \hat{\boldsymbol{x}}'_{\mathcal{K}}$. Both $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ are zero outside $\mathcal{K}$, so $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}'$. ∎

The following property allows to link the uniqueness of the minimizer of H to its degree of sparsity.

Property 2: Let $\mathbf{A}$ satisfy the URP, $\hat{\boldsymbol{x}} \in \mathbb{C}^M$ minimize $H(\boldsymbol{x})$, and $K = \mathrm{card}\,\mathrm{supp}(\hat{\boldsymbol{x}})$. Then:

- If $K < N$, then every other minimizer has more than $N - K$ nonzero components.
- There exists at most one minimizer with at most $N/2$ nonzero components. Hence, if $K < N/2$, the sparsest solution to (5) is $\hat{\boldsymbol{x}}$ and is unique.

Proof: Firstly, suppose that $K < N$, and that $\hat{\boldsymbol{x}}'$ is another minimizer of H. Let $\mathcal{K} = \mathrm{supp}(\hat{\boldsymbol{x}}), \mathcal{K}' = \mathrm{supp}(\hat{\boldsymbol{x}}')$ and $K' = \mathrm{card}\,\mathcal{K}'$. Since H is convex, all vectors $\hat{\boldsymbol{x}}_\alpha = \alpha\hat{\boldsymbol{x}} + (1 - \alpha)\hat{\boldsymbol{x}}', 0 < \alpha < 1$, also minimize H. Therefore, there is an infinity of minimizers with a common support $\mathcal{K}_{\cup} = \mathcal{K} \cup \mathcal{K}'$. According to Lemma 1, it comes that $\mathrm{card}\,\mathcal{K}_{\cup} > N$. On the other hand, $\mathrm{card}\,\mathcal{K}_{\cup} \leq K + K'$, which allows to conclude that $K' > N - K$. In particular, if $K \leq N/2$, then $K' > N/2$. ∎

Note that Gorodnitsky and Rao [10] obtained similar results to characterize sparse solutions to the *noise-free* problem $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$. Results obtained by Fuchs [31, Corollary 1] are also closely related to Property 2, but they are limited to the real-valued case.

Property 2 gives a characterization of sparse minimizers of criterion H, that may be used when characterizations based on low mutual coherence or positive ERC cannot be applied. On the other hand, the URP may be irrelevant for classical dictionaries in Basis Pursuit applications. For example, if the dictionary contains spikes and cisoids [15], such that one cisoid is zero at instant t_m, then the family formed by all spikes at instants $t_{n, n \neq m}$ and this cisoid is linearly dependent.

C. Properties for High-Resolution Spectral Analysis

Let us now turn back to the spectral analysis problem, with a view to examine whether matrix $\mathbf{W}$ fulfills the URP.

1) Regular Sampling Case: In the regular sampling case, the following property holds.

Property 3: In the regular sampling case with $f_{\max}$ not lower than the Nyquist limit, i.e., $f_{\max} \leq 1/(2T_s)$, where T_s is the sampling period, $\mathbf{W}$ satisfies the URP, whatever the discretization step of the frequency grid.

Proof: See Appendix II. ∎

Such a property generalizes a result in [34] about sparse solutions to the exact problem $\boldsymbol{y} = \mathbf{W}\boldsymbol{x}$, obtained for a dictionary of cosine functions and restricted to the case of positive coefficients.

2) Irregular Sampling Case: In the irregular sampling case, a weaker property can be established. Basically, it claims that *for given sampling instants and a given frequency grid, matrix $\mathbf{W}$ is likely to satisfy the URP.*

Property 4: In the irregular sampling case, the values of $f_{\max}$ for which matrix $\mathbf{W}$ does not satisfy the URP are isolated. That is, if $\mathbf{W}$ does not satisfy the URP for a given $f_{\max}^o$, then $\mathbf{W}$ satisfies the URP for all $f_{\max}$ in the neighborhood of $f_{\max}^o$.

Proof: see Appendix III. ∎

Let us remark, yet, that it is easy to build a matrix $\mathbf{W}$ that does not satisfy the URP. For example, for given P and $f_{\max}$, select k_1 and k_2 in $\mathcal{P}$ and choose N instants t_n where the two sinusoids with frequencies $f_{k_1} = k_1 f_{\max}/P$ and $f_{k_2} = k_2 f_{\max}/P$ take the same values. Then, the four columns of $\mathbf{W}$ with indexes $(k_1, -k_1)$ and $(k_2, -k_2)$ are linearly dependent.

IV. OPTIMIZATION STRATEGIES FOR COMPLEX ℓ^1 PENALIZATION

In this section, optimization procedures are considered and discussed in terms of convergence speed. First, an Iterative Coordinate Descent (ICD) algorithm is described, based on the *Block Coordinate Relaxation* (BCR) procedure in [18]. Then, an Iterative Recursive Least Squares (IRLS) algorithm [11], [26] is studied. Finally, the two strategies are compared on simulations and merged to obtain a scheme that practically converges faster than both ICD and IRLS and is guaranteed to converge towards the minimum of (5). Note that the proposed algorithms

TABLE I
ICD ALGORITHM FOR THE MINIMIZATION OF $J_{\mathbb{C}}$

Iteration t: for $k \in \mathcal{P}$,
- update $e_k^{(t)} = \boldsymbol{y} - \sum_{\ell < k} \boldsymbol{w}_\ell x_\ell^{(t)} - \sum_{\ell > k} \boldsymbol{w}_\ell x_\ell^{(t-1)}$
- update $x_k^{(t)} = \phi_\lambda^{\mathrm{s}}(\boldsymbol{w}_k^\dagger e_k^{(t)})$

can be applied to the minimization of any ℓ^1 penalized criterion. In particular, for the spectral analysis problem, the same implementation is valid for both regular and irregular sampling, since no specific matrix properties are exploited.

Property 1 gives an explicit necessary and sufficient condition (NSC) satisfied at the minimum of criterion (5), that allows to test the convergence of any algorithm. In the following, this test was used with a numerical tolerance of 10^{-6} in the equality conditions.

A. Iterative Coordinate Descent (ICD)

The Iterative Coordinate Descent algorithm consists in performing successive 1-D minimizations with respect to each complex variable x_k. It can be viewed as a particular application of the BCR algorithm proposed in [18] for 1-D blocks. It is easy to show that

$$x_k^{\min} = \arg\min_{x_k} J_{\mathbb{C}}(\boldsymbol{x}) \Leftrightarrow x_k^{\min} = \phi_\lambda^{\mathrm{s}}\left(\boldsymbol{w}_k^\dagger e_k\right) \quad (9)$$

where $\boldsymbol{w}_k$ is the kth column of $\mathbf{W}, e_k = \boldsymbol{y} - \sum_{\ell \neq k} \boldsymbol{w}_\ell x_\ell$ and $\phi_\lambda^{\mathrm{s}}$ is the complex *soft shrinkage* function [17]

$$\forall u = \rho e^{j\phi}, \phi_\lambda^{\mathrm{s}}(u) = \begin{cases} (\rho - \lambda)e^{j\phi} & \text{if } \rho > \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, every scalar minimization can be performed at a very low computational cost. Table I summarizes the ICD algorithm. Convergence proofs towards the minimum of $J_{\mathbb{C}}$ can be found in [35]. As ICD operates by successively performing soft thresholding steps, it shows some similarities with the *Iterative Thresholding* procedure in [25]. However, let us note that the two algorithms are not equivalent, since the thresholding is not performed exactly the same way.

B. Iterative Reweighted Least-Squares (IRLS)

Figueiredo [26] proposed an Expectation-Maximization (EM) algorithm for ℓ^1 penalization, which can be extended to complex variables. It can be written as a two-step iterative procedure, where at iteration t
  i) build matrix $\mathbf{Q}^{(t)} = \mathrm{diag}\{|\boldsymbol{x}^{(t)}|\}$;
  ii) compute the new iterate by

$$\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}(\lambda \mathbf{I} + \mathbf{Q}^{(t)}\mathbf{W}^\dagger\mathbf{W}\mathbf{Q}^{(t)})^{-1}\mathbf{Q}^{(t)}\mathbf{W}^\dagger\boldsymbol{y}. \quad (10)$$

This EM algorithm can also be interpreted as an *Iterative Recursive Least-Squares* (IRLS) procedure [36], [37].

Such a structure benefits from the sparsity of the solution.
- The algorithm propagates the zero components. From step ii), it comes that:

$$\left(x_k^{(t)} = 0\right) \rightarrow \left(x_k^{(t+1)} = 0\right). \quad (11)$$

Consequently, the number of nonzero components in $\boldsymbol{x}^{(t)}$, say $K_{\mathcal{I}}^{(t)}$, never increases with the iteration number t.

- As $\mathbf{Q}^{(t)}$ has only $K_{\mathcal{I}}^{(t)}$ nonzero lines and columns, the computation of step ii) amounts to solving a $K_{\mathcal{I}}^{(t)} \times K_{\mathcal{I}}^{(t)}$ system. Then, the computational load either remains constant or decreases with the iteration number t.

Relation (11) clearly shows that such an algorithm will not converge towards the minimum of $J_{\mathbb{C}}$ from any initial point. An extreme example of non convergence is given by the null vector, which is a fixed point according to (11), be it a minimizer or not. This does not contradict the known convergence properties of EM and IRLS algorithms, since the latter are restricted to continuously differentiable criteria [38], [39].

Indeed, the IRLS algorithm is a generalized version of the Weiszfeld algorithm, initially proposed for minimizing the sum of Euclidean distances to a given set of points in $\mathbb{R}^P$, called *vertices* [40]. Kuhn [41] established the convergence of the Weiszfeld algorithm to the optimal solution if no point in the iterates is a vertex. If the iterated sequence reaches a vertex, then the algorithm remains at this vertex. A similar limitation is imposed by (11), so that convergence does not hold without restrictions. Nonetheless, the IRLS algorithm ensures that the criterion never increases [26]: $J(\boldsymbol{x}^{(t+1)}) \leq J(\boldsymbol{x}^{(t)})$ for any current point $\boldsymbol{x}^{(t)}$.

Note that (10) also reads

$$\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}\mathbf{W}^{\dagger}(\lambda\mathbf{I} + \mathbf{W}\mathbf{Q}^{(t)}\mathbf{W}^{\dagger})^{-1}\boldsymbol{y}. \qquad (12)$$

The computation of $\boldsymbol{x}^{(t+1)}$ using expression (12) amounts to solving a $N \times N$ system, and should be preferred to (10) as long as $K_{\mathcal{I}}^{(t)} \geq N$. When $K_{\mathcal{I}}^{(t)} < N$, a more efficient way to update $\boldsymbol{x}^{(t+1)}$ is

$$\boldsymbol{x}_{\mathcal{P}/\mathcal{I}}^{(t+1)} = \mathbf{0} \qquad (13a)$$

$$\boldsymbol{x}_{\mathcal{I}}^{(t+1)} = \left(\mathbf{W}_{\mathcal{I}}^{\dagger}\mathbf{W}_{\mathcal{I}} + \lambda\mathbf{Q}_{\mathcal{I}}^{-1}\right)^{-1}\mathbf{W}_{\mathcal{I}}^{\dagger}\boldsymbol{y} \qquad (13b)$$

where $\mathcal{I} = \mathrm{supp}\,(\boldsymbol{x}^{(t)}), \mathbf{Q}_{\mathcal{I}} = \mathrm{diag}\{|\boldsymbol{x}_{\mathcal{I}}^{(t)}|\}$, and $\mathbf{W}_{\mathcal{I}}$ corresponds to the matrix formed by the columns of $\mathbf{W}$ with indexes in $\mathcal{I}$.

An efficient implementation of the IRLS algorithm is given in Table II. Note that the dimension reduction, caused by the emergence of zero values in $\boldsymbol{x}^{(t)}$, requires some numerical thresholding step at each iteration. For the aforementioned reasons, this still weakens the conditions of convergence of the procedure. In practice, a valid strategy is achieved by testing the convergence with the characterization of the minimum given by Property 1. Despite the lack of theoretical guarantees, convergence towards the minimum of criterion $J_{\mathbb{C}}$ was always reached in our simulations.

C. Mixed Strategy

In this Subsection, the behavior of the ICD and IRLS algorithms are first compared in terms of convergence speed. Then, the two procedures are merged to design a hybrid and convergent algorithm. The data used in this section are those used in Section V, dedicated to simulation results, and the size of the problem is $N = 514, P = 2000$.
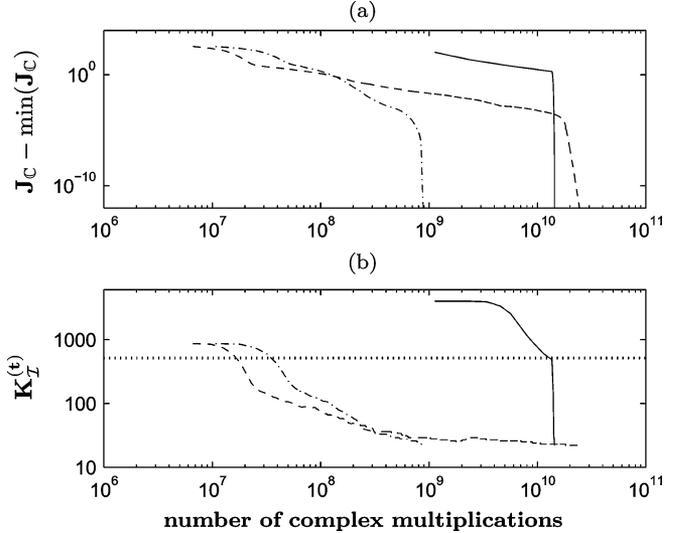

Fig. 4. Behavior of ICD (dashed), IRLS (solid), and hybrid (dash-dotted) algorithms. (a) Evolution of the criterion as a function of the CPU cost. (b) Corresponding evolution of the number of nonzero components. The horizontal dotted line represents the number N of data points.

1) Behavior of the Two Algorithms for Sparse Solutions: Fig. 4(a) plots the evolution of criterion $J_{\mathbb{C}}$ as a function of the computational cost, evaluated as the total number of complex multiplications, for the ICD and IRLS algorithms. At the beginning, the decrease of $J_{\mathbb{C}}$ is faster with ICD than with IRLS (when the IRLS iterations amount to solving $N \times N$ linear systems). On the contrary, when the number of nonzero components in the current iterate, $K_{\mathcal{I}}^{(t)}$, is small enough, $J_{\mathbb{C}}$ decreases more rapidly with the IRLS algorithm than with ICD. Fig. 4(b) reveals the link between the decrease rates of the two algorithms and $K_{\mathcal{I}}^{(t)}$. For the ICD algorithm, the evolution of $K_{\mathcal{I}}^{(t)}$ shows small steps at the beginning and longer steps at the end. This behavior is inverted for the IRLS algorithm. In other words, ICD is good at decreasing the number of nonzero components and bad at estimating the corresponding amplitudes. On the contrary, IRLS lacks efficiency in locating the nonzero components, but once $K_{\mathcal{I}}^{(t)}$ is small, the corresponding amplitudes are estimated very quickly. Note that ICD can be initialized with $\boldsymbol{x}^{(0)} = \mathbf{0}$, which is closer to the sparse solution, whereas IRLS has to be initialized without any zero value.

2) A Hybrid and Convergent Algorithm: The complementary properties of ICD and IRLS suggest the definition of a hybrid algorithm merging the two strategies. Let $M = 2P + 1$ be the size of the unknown vector $\boldsymbol{x}$. Iteration t of ICD (one sweep of the frequency axis) requires $\mathrm{N}M + 2NK_{\mathcal{I}}^{(t)}$ complex multiplications. For $K_{\mathcal{I}}^{(t)} < N$, performing q iterations of IRLS (using step 2 in Table II) requires $qC_{\mathrm{inv}}(K_{\mathcal{I}}^{(t)}) + NK_{\mathcal{I}}^{(t)}$ complex multiplications, where $C_{\mathrm{inv}}(K)$ is the number of multiplications required for solving a $K \times K$ linear system. We used L-U factorization, with

$$C_{\mathrm{inv}}(K) = \frac{1}{2}K^3 + \frac{8}{3}K^2 + \frac{41}{6}K.$$

We propose to insert q iterations of the IRLS algorithm after each ICD sweep, as soon as the two parts have the same compu-

TABLE II
IRLS ALGORITHM FOR THE MINIMIZATION OF J_{C}

Initialize $\boldsymbol{x}^{(0)}$ without any zero component and at iteration t do:

1) if $K_{\mathcal{I}}^{(t)} > N$, perform IRLS steps by solving the $N \times N$ system (12), with $\mathbf{Q}^{(t)} = \mathrm{diag}(\{|\boldsymbol{x}^{(t)}|\})$;

2) as soon as $K_{\mathcal{I}}^{(t)} \leq N$, perform IRLS steps by updating only the non-zero components of $\boldsymbol{x}^{(t)}$: solve the $K_{\mathcal{I}}^{(t)} \times K_{\mathcal{I}}^{(t)}$ system (13b) with $\mathbf{Q}_{\mathcal{I}} = \mathrm{diag}(\{|\boldsymbol{x}_{\mathcal{I}}^{(t)}|\})$, where $\mathcal{I}$ indexes the non-zero components in $\boldsymbol{x}^{(t)}$.

TABLE III
HYBRID PROCEDURE FOR THE MINIMIZATION OF J_{C}

Iteration t:

- perform one ICD sweep (Table I);
- if $q^{(t)} = \frac{NM + NK_{\mathcal{I}}^{(t)}}{C_{\mathrm{inv}}(K_{\mathcal{I}}^{(t)})} \geq 1$, then perform $\left\lfloor q^{(t)} \right\rfloor$ IRLS iterations (Table II).

tational cost. As the ICD algorithm converges towards the minimum of (5) and each IRLS iteration does not increase the value of the criterion, this scheme defines a convergent procedure with ICD *spacer steps* [27, Ch. 7]. Table III gives the resulting algorithm, and the corresponding saving in CPU cost is shown in Fig. 4. At the beginning, the decrease rate is similar to the one obtained with ICD[1] and switches to the IRLS-like decrease rate as soon as the number of nonzero components in the current iterate is small enough. In this example, the number of complex multiplications required by the three algorithms was approximately $1.43\,10^{10}$ for IRLS, $2.46\,10^{10}$ for ICD, and $1.02\,10^{9}$ for the hybrid algorithm. Hence, a gain of 10 in computational cost is obtained by merging ICD and IRLS.

The minimization of (5) formulates a second-order cone (SOC) programming problem [22], so that the optimization can also be performed by efficient interior point algorithms, e.g., [23]. However, in a different application framework with real variables, Sardy *et al.* [18] have found empirically that an ICD-like algorithm converges faster that the interior point algorithm proposed in [14]. Such a behavior has been confirmed in our experiments. The CPU time required for optimization by the hybrid algorithm in Table III (using a C and Matlab implementation for the ICD and IRLS parts, respectively) is about 11 s with a 3.4 GHz Intel P4 processor with 4 GB RAM. Optimizations by ICD and IRLS require approximately 330 s and 100 s, respectively. As a comparison, optimization using a SOC programming formulation with the SeDuMi package[2] [23], which also mixes Matlab and C, requires more than 1300 s to solve the same problem. However, in the considered

---

[1]Fig. 4 shows an extra cost at the beginning of the hybrid algorithm compared to ICD. It is due to the computation of the Toeplitz matrix $\mathbf{W}^{\dagger}\mathbf{W}$ and of $\mathbf{W}^{\dagger}\boldsymbol{y}$, which requires 2 NM complex multiplications.

[2]Implementation was taken from D. Malioutov's web page: http://ssg.mit.edu/group/dmm/dmm.shtml

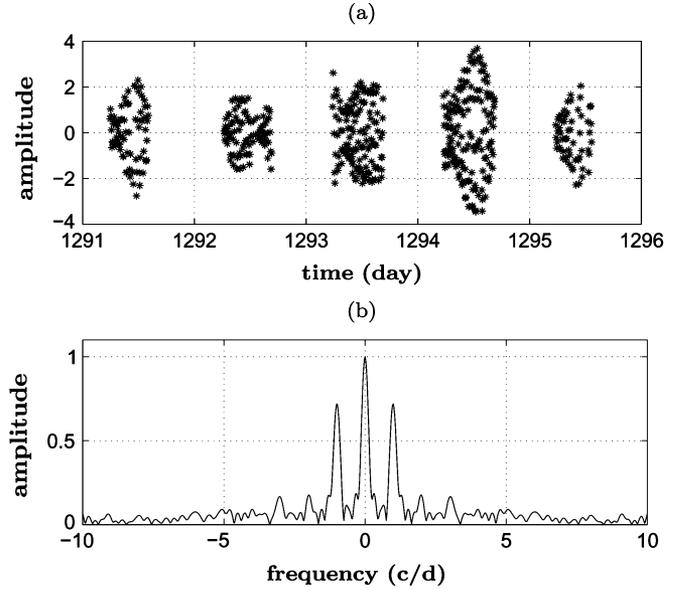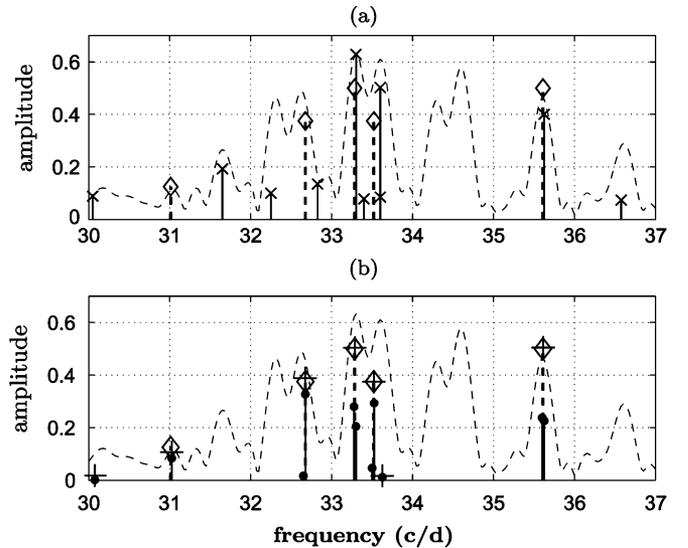Fig. 5. Artificial data. (a) Irregularly sampled time series. (b) Corresponding spectral window.

Fig. 6. (a) Zoom in the frequency range $[30, 37]$c/d of the results of a *Matching Pursuit* algorithm (x-marks). (b) Results obtained by minimizing $J_{\mathrm{C}} : |\hat{\boldsymbol{x}}_{\mathrm{C}}|$ (black circles) and posterior amplitude re-estimation (+ signs). The diamonds locate the true spectral lines and the dotted line represents the Fourier spectrum of the data.

case where the data are real-valued, an equivalent real variable optimization problem can be considered (with sines and cosines instead of cisoids), with only $2P + 1$ real unknowns. This transformed problem can still be written in terms of SOC programming. It is solved by SeDuMi in approximately 170 s. Note that the minimizer obtained by SOC satisfies the NSC of Property 1 up to 10^{-5}, whereas the tolerance used in ICD, IRLS and our hybrid approach was set to 10^{-6}.

V. SIMULATION RESULTS

The artificial data in Fig. 5(a) is the sum of five sinusoids with frequencies ranging from 31 cycles per day (c/d) to 35.6 c/d. Gaussian noise was added to the data with $\mathrm{SNR} = 17$ dB.

The $N = 514$ sampling instants correspond to true observational data [8], covering a period of 5 days with gaps caused by day/night alternation. The resulting spectral window in Fig. 5(b) shows high sidelobes at 1 c/d. For such a data set, the Fourier spectrum shows many false peaks, so that a *Matching Pursuit* method is unable to retrieve the five lines correctly, as shown in Fig. 6(a).

Spectral estimation was performed by minimizing criterion $J_\mathbb{C}$ with $f_{\max} = 50$ c/d and $P = 2000$. Parameter λ was set to 5% of the maximum of $|\mathbf{W}^\dagger \boldsymbol{y}|$. In this example, the mutual coherence of the cisoid dictionary is $\mu \simeq 0.982$, so that uniqueness conditions in [16] would impose the solution $\hat{\boldsymbol{x}}_\mathbb{C}$ to have at most one nonzero component (see Section III-A). Tropp's ERC [17] is negative ($\mathrm{ERC} = -155$). Yet, provided that matrix $\mathbf{W}$ satisfies the Unique Representation Property, which is highly likely according to Property 4, Property 2 allows to conclude that $\hat{\boldsymbol{x}}_\mathbb{C}$ is unique.

Fig. 6(b) shows that estimator $\hat{\boldsymbol{x}}_\mathbb{C}$ is sparse and correctly locates the five frequencies: the nonzero values of $\hat{\boldsymbol{x}}_\mathbb{C}$ locate the closest approximation of the true frequencies on the reconstruction grid—a zoom on the estimated frequencies was shown in Fig. 1. Let us remark that amplitudes are systematically underestimated by $\hat{\boldsymbol{x}}_\mathbb{C}$ because of the ℓ^1-norm penalization term in $J_\mathbb{C}$. Once the frequencies are correctly located, however, a reestimation of amplitudes in the least-squares sense provides accurate results, as shown in Fig. 6(b).

Note that the results in Fig. 6(b) are plotted for positive frequencies only. If the data are real-valued, the estimated spectrum is expected to have the Hermitian symmetry, i.e., $\boldsymbol{x}^H = \boldsymbol{x}$, where $x_k^H = x_{-k}^*$. Such a symmetry is not guaranteed here. However, if $\boldsymbol{y} \in \mathbb{R}^N$, it can be shown [12] that $J_\mathbb{C}(\hat{\boldsymbol{x}}_\mathbb{C}) = J_\mathbb{C}(\hat{\boldsymbol{x}}_\mathbb{C}^H)$. Thus, if the conditions of uniqueness of $\hat{\boldsymbol{x}}_\mathbb{C}$ are satisfied, which is almost always true (see Properties 2 to 4), then $\hat{\boldsymbol{x}}_\mathbb{C}^H = \hat{\boldsymbol{x}}_\mathbb{C}$. Anyhow, let us remark that $(\hat{\boldsymbol{x}}_\mathbb{C} + \hat{\boldsymbol{x}}_\mathbb{C}^H)/2$ is a Hermitian minimizer of criterion $J_\mathbb{C}$ since the latter is convex.

VI. CONCLUSION

We addressed the estimation of spectral lines from irregularly sampled data as a linear and underdetermined inverse problem. Regularization was proposed by minimizing a least-squares cost function, penalized by the ℓ^1-norm. Satisfactory prior sparse modeling was achieved by considering ℓ^1 penalization with *complex* variables, whereas ℓ^1 penalization on real variables was shown to lead to frequency splitting.

Alternate characterizations of sparse solutions were established for the spectral analysis problem. In particular, with regularly sampled data, the minimizer of the penalized criterion is unique if it is sparse enough. This property remains almost always true with irregular sampling.

The computation of the estimator can be performed very efficiently by an hybrid optimization algorithm, which merges two well-known optimization methods with complementary properties. In particular, the sparsity of the solution is advantageously exploited to reduce the computational time. The resulting procedure is ensured to converge towards the minimum of the criterion. Furthermore, this minimum can be characterized explicitly, which allows to test convergence with a strong condition.

Such a methodology was shown to provide high-resolution and robust estimates, especially insensitive to sampling artifacts that frequently appear in astrophysical applications.

APPENDIX I
LOWER BOUND ON THE MUTUAL COHERENCE PARAMETER FOR SPECTRAL ANALYSIS

Property I: The mutual coherence μ of a dictionary $\mathbf{W}$ of cisoids of frequencies $\{k f_{\max}/P\}_{k=-P\ldots P}$, sampled at distinct times $\{t_n\}_{n=0\ldots N-1}$, is strictly lower bounded by

$$\mu_{\min} = \left(\max\left\{ 0, 1 - \frac{2\pi^2 f_{\max}^2}{N^2 P^2} \sum_{m,n=0}^{N-1} (t_n - t_m)^2 \right\} \right)^{1/2} \tag{14}$$

$$= \left(\max\left\{ 0, 1 - \frac{4\pi^2 f_{\max}^2}{P^2} \widetilde{\sigma}^2 \right\} \right)^{1/2} \tag{15}$$

where $\widetilde{\sigma}^2$ is the empirical variance of the sampling instants

$$\widetilde{\sigma}^2 = \frac{1}{N} \sum_n t_n^2 - \frac{1}{N^2} \left(\sum_n t_n \right)^2.$$

In the regular sampling case, $t_n = nT_s$, and at the Nyquist rate $f_{\max} = 1/2T_s$, we have

$$\mu_{\min} = \left(\max\left\{ 0, 1 - \frac{\pi^2}{12} \frac{N^2 - 1}{P^2} \right\} \right)^{1/2}. \tag{16}$$

Proof: Given (7) and (8), we have $\mu \geq \mu_0$, where $\mu_0 = |\mathcal{W}(f_{\max}/P)|$ and $\mathcal{W}$ is the spectral window function (1). It is easy to show that

$$|\mathcal{W}(f)|^2 = \frac{1}{N^2} \sum_{m,n=0}^{N-1} \exp(j2\pi f(t_n - t_m))$$

$$= \frac{1}{N^2} \sum_{m,n=0}^{N-1} \cos 2\pi f(t_n - t_m).$$

Now let us remark that $\cos x > 1 - x^2/2$ for all $x \neq 0$. Therefore, we have

$$|\mathcal{W}(f)|^2 > 1 - \frac{2\pi^2 f^2}{N^2} \sum_{m,n=0}^{N-1} (t_n - t_m)^2 \tag{17}$$

for all $f \neq 0$. Letting $f = f_{\max}/P$, the inequality $\mu_0 > \mu_{\min}$ is readily obtained where $\mu_{\min}$ is given by (14), and hence $\mu > \mu_{\min}$. Finally, (15) is easy to deduce from (14). The proof of (16) is immediate, using $\widetilde{\sigma}^2 = (N^2 - 1)T_s^2/12$ when $t_n = nT_s$ and $f_{\max} = 1/2T_s$. ∎

APPENDIX II
PROOF OF PROPERTY 3

Consider regular sampling: $t_n = nT_s$ (where T_s is the sampling period) and $f_{\max} = \alpha/(2T_s)$, where α

is some under- or over-sampling rate (for $\alpha = 1$, f_{max} corresponds to the Nyquist frequency). Matrix $\mathbf{W}$ reads $\{1/(\sqrt{N}) \exp(j\alpha\pi kn/P)\}_{n=0,\dots,N-1, k=-P,\dots,P}$.

Let $\{k_m\}_{m=0,\dots,N-1}$ be an increasing sequence from $\mathcal{P} = \{-P, \dots, P\}$ indexing N arbitrary columns from $\mathbf{W}$, and consider the $N \times N$ matrix

$$\mathbf{U}_N = \left\{ \frac{1}{\sqrt{N}} \exp\left(j\alpha\pi\frac{k_m}{P}n\right) \right\}_{n,m=0,\dots,N-1}.$$

$\mathbf{U}_N$ is a full rank matrix if and only if the N constraints

$$\forall m = 0,\dots,N-1, \sum_{n=0}^{N-1} \beta_n \exp\left(j\alpha\pi\frac{k_m}{P}n\right) = 0$$

involve $\beta_n = 0, \forall n = 0,\dots,N-1$. The polynomial $Q(X) = \sum_{n=0}^{N-1} \beta_n X^n$ is of degree $d \leq N-1$ and is zero for $z_m = \exp(j\alpha\pi k_m/P), m = 0,\dots,N-1$. If the zeros z_m are all different, then the polynomial $Q(X)$ has N zeros and is thus identically zero: $\beta_n = 0, \forall n = 0,\dots,N-1$. It holds

$$\text{for } m \neq q, z_m \neq z_q \iff \frac{\alpha\pi k_m}{P} \not\equiv \frac{\alpha\pi k_q}{P}[2\pi]. \qquad (18)$$

A sufficient condition for $z_m \neq z_q$ is $\alpha\pi|k_m - k_q|/P < 2\pi$, which must be true for every sequence $\{k_m\}_{m=0,\dots,N-1}$ from $\mathcal{P}$. The maximum value of $|k_m - k_q|$ among all sequences is $2P$, so $z_m \neq z_q$ is ensured if $\alpha < 1$.

The case $\alpha = 1$ has to be considered separately. Indeed, in this case, the first and last columns of $\mathbf{W}$ are identical, so one has to consider the matrix formed by the $2P$ last columns of $\mathbf{W}$, that is, $\mathbf{W}_{reg} = \{1/\sqrt{N} \exp(j\pi kn/P)\}_{n=0,\dots,N-1, k=1-P,\dots,P}$. The maximum value of $|k_m - k_q|$ among all sequences $\{k_m\}_{m=0,\dots,N-1}$ from $\{1-P, \dots, P\}$ is $2P-1$, so $z_m \neq z_q$ is ensured if $\alpha < 2P/(2P-1)$, which is true for $\alpha = 1$.

As a conclusion, matrix $\mathbf{W}$ satisfies the URP condition if $\alpha \leq 1$.

APPENDIX III
PROOF OF PROPERTY 4

Let $\mathbf{U}_N$ be a matrix formed by N arbitrary columns of matrix $\mathbf{W}$

$$\mathbf{U}_N = \left\{ \exp\left(j2\pi\frac{k_m}{P}f_{max}t_n\right) \right\}_{n,m=0,\dots,N-1}$$

for some frequency indexes $k_{m,m=0,\dots,N-1} \in \mathcal{P}$ arranged in increasing order. The determinant of $\mathbf{U}_N$ is

$$\det\mathbf{U}_N = \sum_{\sigma\in S_N} \text{sgn}(\sigma)u_{0,\sigma(0)}u_{1,\sigma(1)},\dots,u_{N-1,\sigma(N-1)}$$

where S_N is the set of all permutations in dimension N and $\text{sgn}(\sigma) \in \{-1; 1\}$ is the signature of permutation σ. That is

$$\det\mathbf{U}_N = \sum_{\sigma\in S_n} \text{sgn}(\sigma) \exp\left(j2\pi \sum_{n=0}^{N-1} \frac{k_n}{P}f_{max}t_{\sigma_n}\right).$$

Consider $\det\mathbf{U}_N$ as a function of parameter f_{max}, $\det\mathbf{U}_N = G_N(f_{max})$, with

$$G_N(x) = \sum_{\sigma\in S_n} \text{sgn}(\sigma) \exp(j2\pi\alpha_\sigma x),$$

where $\alpha_\sigma = 1/P \sum_{n=0}^{N-1} k_n t_{\sigma_n}$.

Function G_N is the finite sum of trigonometric functions with frequencies α_σ, weighted by ± 1. Let us show that at least one frequency among all α_σ appears exactly once in the definition of G_N: as $\{k_m\}_m$ and $\{t_n\}_n$ are strictly increasing sequences, the frequencies α_σ range from $\alpha_{min} = 1/P \sum_{n=0}^{N-1} k_n t_{N-1-n}$ to $\alpha_{max} = 1/P \sum_{n=0}^{N-1} k_n t_n$ and, e.g., value α_{max} is reached only once among all α_σ, for $\sigma = [0,\dots,N-1]$. Thus, the function G_N cannot be zero on any interval.

Let $\mathcal{Z}$ denote the set of values of f_{max} for which $\mathbf{W}$ does not satisfy the URP, that is, for which there exists a submatrix $\mathbf{U}_N$ made of N columns of $\mathbf{W}$, such that $\det\mathbf{U}_N = 0$. $\mathcal{Z}$ is the finite union of sets of isolated points. Thus, $\mathcal{Z}$ only contains isolated points.

REFERENCES

[1] S. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, Mar. 1989.

[3] F. J. M. Barning, "The numerical analysis of the light-curve of 12 Lacertae," *Bull. Astronom. Inst. Netherlands*, vol. 17, pp. 22–28, Aug. 1963.

[4] D. F. Gray and K. Desikachary, "A new approach to periodogram analyses," *Astrophys. J.*, vol. 181, pp. 523–530, Apr. 1973.

[5] D. H. Roberts, J. Lehar, and J. W. Dreher, "Time series analysis with CLEAN. I. Derivation of a spectrum," *Astronom. J.*, vol. 93, no. 4, pp. 968–989, Apr. 1987.

[6] G. Foster, "The cleanest Fourier spectrum," *Astronom. J.*, vol. 109, no. 4, pp. 1889–1902, Apr. 1995.

[7] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[8] S. Bourguignon, H. Carfantan, and T. Böhm, "SparSpec: A new method for fitting multiple sinusoids with irregularly sampled data," *Astron. Astrophys.*, vol. 462, pp. 379–387, Jan. 2007.

[9] S. Cabrera and T. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 842–851, Apr. 1991.

[10] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[11] M. Sacchi, T. Ulrych, and C. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 31–38, Jan. 1998.

[12] P. Ciuciu, J. Idier, and J.-F. Giovannelli, "Regularized estimation of mixed spectra using a circular Gibbs-Markov model," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2201–2213, Oct. 2001.

[13] F. Dublanchet, "Contribution de la Méthodologie Bayésienne à l'analyse Spectrale de Raies Pures et à la Goniométrie Haute Résolution," Ph.D. dissertation, Univ. de Paris-Sud, Paris, France, Oct. 1996.

[14] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[15] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[16] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.

[17] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[18] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *J. Comput. Graph. Statist.*, vol. 9, pp. 361–379, 2000.

[19] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–403, 2000.

[20] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[21] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[22] S. Winter, H. Sawada, and S. Makino, "On real and complex valued ℓ^1-norm minimization for overcomplete blind source separation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2005, pp. 86–89.

[23] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11, no. 12, pp. 625–653.

[24] S. Chen and D. Donoho, "Application of basis pursuit in spectrum estimation," in *Proc. IEEE ICASSP*, Seattle, WA, May 1998, vol. 3, pp. 1865–1868.

[25] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[26] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 1150–1159, Sep. 2003.

[27] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1989.

[28] L. Eyer and P. Bartholdi, "Variable stars: Which Nyquist frequency," *Astron. Astrophys. Suppl. Series*, vol. 135, pp. 1–3, Feb. 1999.

[29] S. Bourguignon, H. Carfantan, and L. Jahan, "Regularized estimation of line spectra from irregularly sampled astrophysical data," in *Proc. Physics in Signal and Image Processing*, Toulouse, France, Jan. 2005.

[30] J.-F. Giovannelli and J. Idier, "Bayesian interpretation of periodograms," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1988–1996, 2001.

[31] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[32] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[33] S. Alliney and S. A. Ruzinsky, "An algorithm for the minimization of mixed ℓ_1 and ℓ_2 norms with application to Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 618–627, Mar. 1994.

[34] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Proc. IEEE ICASSP*, Philadelphia, PA, Mar. 2005, vol. 5, pp. 729–732.

[35] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim Theory Appl.*, vol. 109, pp. 475–494, 2001.

[36] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[37] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, Sept. 2004.

[38] C. F. J. Wu, "On the convergence of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

[39] R. Byrd and D. Payne, Convergence of the Iterative Reweighted Least Squares Algorithm for Robust Regression, John Hopkins Univ., Baltimore, MD, Tech. Rep. 313, Jun. 1979.

[40] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM J. Numer. Anal.*, vol. 36, no. 2, pp. 354–367, 1999.

[41] H. Kuhn, "A note on Fermat's problem," *Math. Progr.*, vol. 4, pp. 98–107, 1973.

Sébastien Bourguignon was born in Dijon, France in 1977. He received the diploma degree in electrical engineering from École Supérieure d'Électricité, Gif-sur-Yvette, France, and the engineer degree in telecommunications from ETSIT, Universidad Politécnica de Madrid, Spain, both in 2001, and the Ph.D. degree in signal processing from University Paul Sabatier, Toulouse, France, in 2005.

From 2002 to 2007, he was with the Astrophysics laboratory of Toulouse-Tarbes, France. He recently joined the Fisheries Technology laboratory of Ifremer, the French research institute for exploitation of the sea. His research interests include Bayesian inference and estimation, time series analysis, optimization and MCMC algorithms and, recently, acoustic imaging, detection and classification methods for fisheries applications.

Hervé Carfantan was born in France in 1968. He graduated from the École Supérieure d'Infomatique Électronique Automatique, Paris, France, post-graduated in Control and Signal Processing from University of Paris-Sud, Orsay, France, both in 1992 and received the Ph.D. degree in physics from University of Paris-Sud in 1996.

He is presently an Assistant Professor in the University Paul Sabatier-Toulouse 3 and a researcher in the Signal, Image & Instrumentation team of the Astrophysics laboratory of Toulouse-Tarbes, France. His main interests are in inverse problems, estimation, spectral analysis and optimization. His application fields essentially concern astronomical and geophysical data analysis.

Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from École Supérieure d'Électricité, Gif-sur-Yvette, France, in 1988 and the Ph.D. degree in physics from University of Paris-Sud, Orsay, France, in 1991.

Since 1991, he joined the Centre National de la Recherche Scientifique. He is currently Senior Researcher at the Institut de Recherche en Communications et Cybernétique in Nantes. His major scientific interests are in probabilistic approaches to inverse problems for signal and image processing.